

Philosopher Elliot Temple Discusses Epistemology with Less Wrong

<http://lesswrong.com/>

<https://elliottemple.com/>

et@elliottemple.com

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Questions about AGI's Importance

0 curi 31 October 2017 08:50PM

Why expect AGIs to be better at thinking than human beings? Is there some argument that human thinking problems are primarily due to hardware constraints? Has anyone here put much thought into parenting/educating AGIs?

Article Navigation

[Comments \(117\)](#)

Tags: [ai](#)

Comments (117)

Sort By: **Best**

Comment

Show help

korin43 31 October 2017 10:15:25PM 1 point

I suspect this has been answered on here before in a lot more detail, but:

- Evolution isn't necessarily trying to make us smart; it's just trying to make us survive and reproduce
- Evolution tends to find local optima (see: obviously stupid designs like how the optical nerve works)

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

- We seem to be pretty good at making things that are better than what evolution comes up with (see: no birds on the moon, no predators with natural machine guns, etc.)

Also, specifically in AI, there is some precedent for there to be only a few years between "researchers get AI to do something at all" and "this AI is better at its task than any human who has ever lived". Chess did it a while. It just happened with Go. I suspect we're crossing that point with image recognition now.

curi 31 October 2017 10:29:20PM * 0 points



Do you expect AGI to be qualitatively or quantitatively better at thinking than humans?

Do you think there are different types of intelligence? If so, what types? And would AGI be the same type as humans?

EDIT: By "intelligence" I mean general intelligence.

curi 10 November 2017 06:47:37PM 0 points



I'm getting an error trying to load Lumifer's comment in the highly nested discussion, but I can see it in my inbox, so I'll try replying here without the nesting. For this comment, I will quote everything I reply to so it stands alone better.

Isn't it convenient that I don't have to care about these infinitely many theories?

why not?

Why not what?

Why don't you have to care about the infinity of theories?

you can criticize categories, e.g. all ideas with feature X

How can you know that every single theory in that infinity has feature X? or belongs to the same category?

It depends which infinity we're talking about. Suppose the problem is *persuading LW ppl about Paths Forward* and you say "Use a shovel". That refers to infinitely many different potential solutions. However, they can be criticized as a group by pointing out that a shovel won't help solve the problem. What does a shovel have to do with it? Irrelevant!

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

This criticism only applies to the infinite category of ideas about shovels, not everything. I'm able to criticize that whole infinite group as a unit because it was brought up as a unit, and defined according to having a particular feature for all the theories in the group (that they involve trying to solve the problem specifically with a shovel.)

The criticism is also contextual. It relates to using shovels for this particular problem. But shovels still help with some other problems. The context the criticism works in is broader than the single problem about paths forward persuasion of LW ppl – e.g. it also applies to anti-induction persuasion of Objectivists. This is typical – the point has some applicability to multiple contexts, but not universal applicability.

If you instead said "Do something" then you'd be bringing up a different infinity with more stuff in it, and I'd have a different reply: "Do what? That isn't helpful because you're pointing me to a large number of non-solutions without pointing out any solution. I agree there is a solution contained in there, somewhere, but I don't know what it is, and you don't seem to either, so I can't use it currently. So I'm stuck with the regular options like doing a solution I do know of or spending more time looking for solutions."

I will admit that there may be a solution with a shovel that actually would work (one way to get this is to take some great solution and then tack on a shovel, which is not optimal but may still be way better than anything we currently know of). So my criticism doesn't 100% rule shovels out. However, it rules shovels out *for the time being, as far as is known, pending a new idea about how to make a shovel work*. We can only act on solutions we know of, and I have a criticism of the shovel category of ideas as we currently understand it. Our current understanding is that shovels help us dig, and can be used as weapons, and can be salvaged for resources like wood and metal, and can be sold, but that just vaguely saying "use a shovel somehow" does not help me solve a problem of intellectually persuading people.

you can't observe entities

My nervous system makes perfectly good entities out of my sensory stream. Moreover, a rat's nervous system also makes perfectly good entities out of its sensory stream regardless of the fact that the rat has never heard of epistemology and is not very philosophically literate.

I don't think humans think like rats, and I propose we don't debate animal "intelligence" at this time. I'll try to speak to the issue in a different way.

I think humans have adequate control over their observing that they don't get *stuck and unable to make progress* due to built-in biases and errors. For example, people can consciously think "that looked like a dog at first glance, but actually isn't a painting of a dog". So you can put thought into what the entities are. To the extent you have a default, you can partly change what that default is, and partly reinterpret it after doing the observation. And you're capable of observing in a sufficiently non-lossy way to get whatever information you need (at least with tools like microscopes for some cases). You aren't just inherently, permanently blind to some ways of dividing up the world into entities, or some observable things.

[Link] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians

by curi | 1v (0c)

[Link] Intercellular competition and the inevitability of multicellular aging

by Gunnar_Zarncke | 1v (3c)

Announcing the AI Alignment Prize

by cousin_it | 6v (2c)

Problems as dragons and papercuts

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

And whatever default your genes gave you about entities is not super reliable. It may be pretty good, but it's very much capable of errors. So I'll make a weaker claim: you can't infallibly observe entities. You need to put some actual thought into what the entities are and aren't, and the inductivist perspective doesn't address this well. (As to rats, they actually *start making gross errors* in some situations, due to their inability to think like a human to deal with situations they weren't evolved for.)

you have to interpret what entities there are (or not – as you advocated by saying only prediction matters)

or not

Or not? Prediction matters, but entities are an awfully convenient way to make predictions.

but when two ways of thinking about entities (or, a third option, not thinking about entities at all) give identical predictions, then you said it doesn't matter which you do? one entity (or none) is as good as another as long as the predictions come out the same?

but i don't think all ways of looking at the world in terms of entities are equally convenient for aiding us in making predictions (or for some other important things like coming up with new hypotheses!)

Lumifer 10 November 2017 07:52:33PM 0 points



Huh, that shaft ended in loud screech and a clang... Let's drop another shaft!

Why don't you have to care about the infinity of theories?

I don't have to care about the infinity of theories because if they all make exactly the same predictions, I don't care that they are different.

This is highly convenient because I am, to quote an Agent, "only human" and humans are not well set up to deal with infinities.

they can be criticized as a group by pointing out that a shovel won't help solve the problem

How do you know that without examining the specific theories?

We can only act on solutions we know of, and I have a criticism of the shovel category of ideas as we currently understand it.

Right, but the point is that you do not have solution at the moment and there is an infinity of theories which propose potential shovel-ready solutions. You have no basis for rejecting them because "I don't know of a

IlyaShpitser (16)
 curi (14)
 Dagon (11)
 NancyLebovitz (6)
 satt (6)
 cousin_it (6)
 Habryka (6)
 ChristianKI (6)
 SquirrelInHell (5)
 root (5)

RECENT KARMA AWARDS

solution with a shovel" -- they are **new to you** solutions, that's the whole point.

To the extent you have a default, you can partly change what that default is, and partly reinterpret it after doing the observation.

Yes, of course, but you were claiming there are no such things as observations at all, merely some photons and such flying around. Being prone to errors is an entirely different question.

One entity (or none) is as good as another as long as the predictions come out the same?

Predictions do not come out of nowhere. They are made by models (= imperfect representations of reality) and "entity" is just a different word for a "model". If you don't have any entities, what exactly generates your predictions?

curi 10 November 2017 08:33:02PM * 0 points 

I don't find these replies very responsive. Are you trying to understand what I'm getting at, or just writing local replies to a selection of my points? This is not the first time I've tried to write some substantial explanation and gotten not much engagement from you (IMO).

Lumifer 10 November 2017 09:08:07PM 0 points 

Oh, I understand what you are getting at. I just think that you're wrong.

I'm writing local replies because fisking walls of text gets tedious very very quickly. There is no point in debating secondary effects when it's pretty clear that the source disagreement is deeper.

curi 10 November 2017 09:14:43PM * 0 points 

I'm going to end the discussion now, unless you object. I'm willing to consider objections.

I'm stopping for a variety of reasons, some of which I talked about previously like your discussion limitations like about references. I think you don't understand and aren't willing to do what it takes to understand.

If we stop and you later want to get these issues addressed, you would be welcome to post to the FI forum: <http://fallibleideas.com/discussion-info>

Lumifer 10 November 2017 09:20:59PM 0 points [-]

I think you don't understand and aren't willing to do what it takes to understand.

s/understand/be convinced/g and I'll agree :-)

Was a fun ride!

ImmortalRationalist 02 November 2017 06:00:08PM 0 points [-]

Here is a somewhat relevant video.

whpearson 01 November 2017 12:21:21PM 0 points [-]

Has anyone here put much thought into parenting/educating AGIs?

I'm interested in General Intelligence Augmentation, what it would be like try and build/train an artificial brain lobe and try and make it part of a normal human intelligence.

I wrote a bit on my current thoughts on how I expect to align it using training/education here but watching this presentation is necessary for context.

silver 01 November 2017 09:32:32AM 0 points [-]

Because

"[the brain] is sending signals at a millionth the speed of light, firing at 100 Hz, and even in heat dissipation [...] 50000 times the thermodynamic minimum energy expenditure per binary switch operation"

<https://www.youtube.com/watch?v=EUjc1WuyPT8&t=3320s>

AI will be quantitatively smarter because it'll be able to think over 10000 times faster (arbitrary conservative lower bound) and it will be qualitatively smarter because its software will be built by an algorithm far better than evolution

Lumifer 01 November 2017 03:57:45PM 0 points



AI will be quantitatively smarter because it'll be able to think over 10000 times faster

My calculator can add large numbers much, much faster than I. That doesn't make it "quantitatively smarter".

an algorithm far better than evolution

Given that no one has any idea about what that algorithm might look like, statements like this seem a bit premature.

Tehuti 05 November 2017 08:26:07PM * 0 points



My calculator can add large numbers much, much faster than I. That doesn't make it "quantitatively smarter."

Your brain actually performs much more analysis each second than any computer we have:

At the time of this writing, the fastest supercomputer in the world is the Tianhe-2 in Guangzhou, China, and has a maximum processing speed of 54.902 petaFLOPS. A petaFLOP is a quadrillion (one thousand trillion) floating point calculations per second. That's a huge amount of calculations, and yet, that doesn't even come close to the processing speed of the human brain. In contrast, our miraculous brains operate on the next order higher. Although it is impossible to precisely calculate, it is postulated that the human brain operates at 1 exaFLOP, which is equivalent to a billion billion calculations per second.

<https://www.scienceabc.com/humans/the-human-brain-vs-supercomputers-which-one-wins.html>

Of course this is structurally very different from a CPU or a GPU etc, but the whole power of the brain is still way bigger.

curi 01 November 2017 09:22:22PM * 0 points



I think AGIs will be built by evolution, and use evolution for their own thinking, because I think human thinking uses evolution (replication with variation and selection of ideas). I don't think any other method of knowledge creation is known, other than evolution.

Lumifer 02 November 2017 12:37:23AM 1 point [-]

I don't think any other method of knowledge creation is known, other than evolution.

The scientific method doesn't look much like evolution to me. At a simpler level, things like observation and experimentation don't look like it, either.

username2 10 November 2017 01:23:40PM 0 points [-]

I went down the rabbit hole of your ensuing discussion and it seems to have broken LW, but didn't look like you were very convinced yet. Thanks for taking one for the team.

Lumifer 10 November 2017 03:39:26PM 0 points [-]

Too deep we delved there, and woke the nameless fear...

I suspect there is an implicit max thread depth and once it's reached, LW's gears and cranks (if only!) screech to a halt.

curi 02 November 2017 12:48:54AM 0 points [-]

The scientific method involves guesses (called "hypotheses") and criticism (including by experimental tests). That follows the pattern of evolution (exactly, not by analogy): replication with variation (guessing), and selection (criticism).

Lumifer 02 November 2017 04:10:00PM 1 point [-]

That follows the pattern of evolution (exactly, not by analogy)

Not at all. Hypothesis generation doesn't look like taking the current view and randomly changing one element in it. More importantly, science is mostly teleological and evolution is not.

But let's take a trivial example. Let's say I'm walking by a food place and I notice a new to me dish. I order it, eat it, and decide that it's tasty. I have acquired knowledge. How's that like evolution?

curi 02 November 2017 05:59:39PM * 0 points [-]

the way you decide it's tasty is by guessing it's tasty, and guessing some other things, and criticizing those guesses, and "it's tasty" survives criticism while its rivals don't.

lots of this is done at an unconscious level.

it has to be this way b/c it's the only known way of creating knowledge that could actually work. if you find it awkward or burdensome, that doesn't make it impossible – which puts it ahead of its rivals.

Lumifer 02 November 2017 06:23:20PM 0 points [-]

The word you're looking for is "testing". I *test* whether that thing is tasty.

Testing is not the same thing as evolution.

it has to be this way b/c it's the only known way of creating knowledge that could actually work

That's an entirely circular argument.

curi 02 November 2017 06:30:30PM 0 points [-]

Evolution is an abstract pattern which makes progress via the correction of errors using selection. If something fits the pattern, then it's evolution.

Would you agree with something like: if induction doesn't work, and CR does, then it's a good idea to accept CR? Even if you find it counter-intuitive and awkward from your current perspective?

Lumifer 02 November 2017 08:01:39PM * 0 points [-]

Evolution is an abstract pattern which makes progress via the correction of errors using selection

I think we might be having terminology problems -- in particular I feel that you stick the "evolution" label on vastly broader things.

First, the notion of progress. Evolution doesn't do progress not being teleological. Evolution does **adapation** to the current environment. A decrease in complexity is not an uncommon event in evolution, for example. A mass die-off is not an uncommon event, either.

Second, evolution doesn't correct "errors". Those are not errors, those are random exploratory steps. A random walk. And evolution does not correct them, it just kills off those who misstep (which is 99.99%+ of steps).

| if induction doesn't work, and CR does, then it's a good idea to accept CR?

Sure. Please provide empirical evidence.

And I still don't understand what's wrong with plain-vanilla *observation* as a way to acquire knowledge.

[continue this thread »](#)

Elo 02 November 2017 04:44:47AM 0 points

[-]

| The scientific method

You read the same book as me! "Theory And Reality - Peter Godfrey Smith". I am surprised you say this.

What you describe is the hypothetico-deductive method (https://en.wikipedia.org/wiki/Scientific_Method pictured here is the hypothetico-deductive method, wikipedia is wrong and disagrees with it's own sources). The hypothetico-deductive method involves guesses but the scientific method according to that book is about:

1. observation
2. measurement (and building models that can be predictive of that measurement)
3. standing on the shoulders of the existing body of knowledge.
4. ???
5. Profit!

Edit: that wiki page has changed a lot over the last few months and now I am less sure about what it says.

curi 02 November 2017 07:04:51AM 0 points [-]

I don't understand what reading a book has to do with it, or what you wish me to take from the wikipedia link. In my comment I stated the CR position on scientific method, which is my position. Do you have a criticism of it?

curi 01 November 2017 09:38:23AM 0 points [-]

i think humans don't use their full computational capacity. why expect an AGI to?

in what way do you think AGI will have a better algorithm than humans? what sort of differences do you have in mind?

silver 01 November 2017 10:30:43AM * 0 points [-]

It doesn't really matter whether the AI uses their full computational capacity. If the AI has a 100000 times larger capacity (which is again a conservative lower bound) and it only uses 1% of it, it will still be 1000 as smart as the human's full capacity.

AGI's algorithm will be better, because it has instant access to more facts than any human has time to memorize, and it will not have all of the biases that humans have. The entire point of the sequences is to list dozens of ways that the human brain reliably fails.

curi 01 November 2017 08:23:28PM 0 points [-]

If the advantage is speed, then in one year an AI that thinks 10,000x faster could be as productive as a person who lives for 10,000 years. Something like that. Or as productive as one year each from 10,000 people. But a person could live to 10,000 and not be very productive, ever. That's easy, right? Because they get stuck, unhappy, bored, superstitious ... all kinds of things can go wrong with their thinking. If AGI only has a speed advantage, that won't make it immune to dishonesty, wishful thinking, etc. Right?

Humans have fast access to facts via google, databases, and other tools, so memorizing isn't crucial.

The entire point of the sequences is to list dozens of ways that the human brain reliably fails.

I thought they talked about things like biases. Couldn't an AGI be biased, too?

Lumifer 01 November 2017 08:26:18PM * 0 points 

For fun ways in which NN classifiers reliably fail, google up adversarial inputs :-)

Example

Elo 01 November 2017 08:38:50PM 0 points 

Rubbish in, rubbish out - right?

Lumifer 02 November 2017 12:33:31AM * 0 points 

No, not quite. It's more like "let us poke around this NN and we'll be able to craft inputs which look like one thing to a human and a completely different thing to the NN, and the NN is very sure of it".

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Lumifer comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: [ai](#)

Comments (117)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above](#).

Lumifer 02 November 2017 08:01:39PM * 0 points

Evolution is an abstract pattern which makes progress via the correction of errors using selection

I think we might be having terminology problems -- in particular I feel that you stick the "evolution" label on vastly broader things.

First, the notion of progress. Evolution doesn't do progress not being teleological. Evolution does **adapation** to the current environment. A decrease in complexity is not an uncommon event in evolution, for example. A mass die-off is not an uncommon event, either.

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

- NEAREST MEETUPS**
- Brussels - We meet every month: 12 January 2019 02:00PM
 - LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM
Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS
Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

Second, evolution doesn't correct "errors". Those are not errors, those are random exploratory steps. A random walk. And evolution does not correct them, it just kills off those who misstep (which is 99.99%+ of steps).

| if induction doesn't work, and CR does, then it's a good idea to accept CR?

Sure. Please provide empirical evidence.

And I still don't understand what's wrong with plain-vanilla *observation* as a way to acquire knowledge.

curi 02 November 2017 08:27:18PM 0 points



killing off a misstep is a way of getting rid of that error. the stuff that doesn't work is probabilistically removed from later generations – so the effect there is error correction. (experimenting itself isn't a mistake, but some of the experiments work badly – error).

Evolution adapts, yes. Adapting something to solve a particular problem = creating knowledge of how to solve that problem. Biological evolution is limited in what problems it solves but still powerful enough to create human intelligence b/c of the ability for a single piece of knowledge to solve multiple problems.

abstractly, guesses and criticism fits the pattern of evolution: there are generations of ideas. the ideas in the next generation aren't purely random, they retain some things that worked in the previous generation (to some extent we're seeing variation instead of something totally separate), and then criticism is selection. if you keep applying the same criticism over and over, you'll get ideas adapted to not being refuted by that criticism.

| Please provide empirical evidence.

Our disagreement is about philosophy.

| And I still don't understand what's wrong with plain-vanilla observation as a way to acquire knowledge.

What do you observe (observation is lossy and there are many choices about where to focus your attention), and then what do you learn from it? Any set of observations fits infinitely many patterns.

Lumifer 02 November 2017 08:41:01PM * 1 point



| abstractly, guesses and criticism fits the pattern of evolution

I still don't think so, but as I mentioned it's merely a terminology problem: you are using the word "evolution" in an unexpected way.

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

Please provide empirical evidence.

Our disagreement is about philosophy.

Ah, well then. In this case I probably should inform you that your mistakes are due to the invisible dragon in my garage. When he gets gas, his dreams are troubled and seep into the minds of humans, corrupting their epistemology. See, he is a Philosophical Dragon.

What do you observe ... and then what do you learn from it?

I observe a rock and learn that there is a rock in front of me.

curi 02 November 2017 09:43:25PM 0 points [-]

why did you learn there was a rock in front of you, instead of an alien that looks like a rock?

do you, perhaps, have a *criticism* of the alien suggestion?

Lumifer 03 November 2017 12:50:40AM 0 points [-]

I cannot guarantee that it's not an alien that looks like a rock, but my priors insist that it's highly improbable.

do you, perhaps, have a criticism of the alien suggestion?

Me, no, but you might want to talk to that chap over there, William of Ockham...

curi 03 November 2017 04:09:12AM 0 points [-]

so you prefer a dogmatic prior over criticisms which are themselves exposed to criticism?

Lumifer 03 November 2017 02:22:40PM 0 points [-]

a dogmatic prior

How is it dogmatic when a prior's sole purpose in life is literally to be updated, to change?

over criticisms

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)

by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)

by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)

by cousin_it | 6v (2c)

[Problems as dragons and papercuts](#)

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

Which criticisms? Where do they come from? Who makes them and for what reason?

curi 03 November 2017 05:34:22PM 0 points [-]

How is it dogmatic when a prior's sole purpose in life is literally to be updated, to change?

not by critical arguments.

Which criticisms? Where do they come from? Who makes them and for what reason?

humans make critical arguments, like the ones in this discussion.

Lumifer 03 November 2017 07:38:34PM 0 points [-]

So we started here:

I observe a rock and learn that there is a rock in front of me.

There are just two of us here, me and the rock. If there are no humans around to make criticisms, I cannot acquire knowledge?

not by critical arguments

If these critical arguments get to count as **evidence**, yes, by them, too. If they don't, well, that raises interesting questions.

curi 03 November 2017 08:39:46PM 0 points [-]

There are just two of us here, me and the rock. If there are no humans around to make criticisms, I cannot acquire knowledge?

You are a human who is present and can criticize.

If these critical arguments get to count as evidence, yes, by them, too. If they don't, well, that raises interesting questions.

You're defining "evidence" differently than I am. I think evidence refers to what you might call *empirical* evidence. How do you incorporate critical arguments into probability updating?

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

[continue this thread »](#)

[ABOUT LESS WRONG](#) | [REPORT ISSUES](#) | [PRIVACY & TERMS](#)

Powered by **Reddit** 

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

curi comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: ai

Comments (117)

Sort By: Best

You are viewing a single comment's thread. [Show more comments above](#).

curi 03 November 2017 08:39:46PM 0 points

There are just two of us here, me and the rock. If there are no humans around to make criticisms, I cannot acquire knowledge?

You are a human who is present and can criticize.

If these critical arguments get to count as evidence, yes, by them, too. If they don't, well, that raises interesting questions.

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

You're defining "evidence" differently than I am. I think evidence refers to what you might call *empirical* evidence. How do you incorporate critical arguments into probability updating?

Lumifer 04 November 2017 12:24:11AM 0 points 

You are a human who is present and can criticize.

But I don't do that. My eyes send some information to my brain, my brain does, basically, pattern-matching and says "looks like a rock". Another part of the brain runs a sanity check ("Would seeing a rock be reasonable here? Yes.") and I'm done.

In particular, I do NOT generate a large number of hypotheses about what that thing might be and internally criticize them.

How do you incorporate critical arguments into probability updating?

Easily enough. **Valid** critical arguments tend to point to empirical evidence which contradicts the hypothesis. Other than that, the only valid arguments that come to mind are those which demonstrate incoherency or internal contradictions.

curi 04 November 2017 01:36:38AM 0 points 

We have massive philosophical differences. I think you're wrong in important ways and that your school of thought has been refuted by literature it hasn't answered (by e.g. Popper and Deutsch).

Are you interested in resolving this in a serious, thorough way to a conclusion? I understand this would take a large effort by each of us.

Lumifer 05 November 2017 12:12:30AM 0 points 

Are you interested in resolving this in a serious, thorough way to a conclusion?

Depends on what "resolving" means. If you have in mind pinpointing the precise issues from which our disagreement stems, sure. But I don't think it would take a large effort.

On the other hand, if what you have in mind is teaching me the proper way to do philosophy, that's much more problematic...

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

curi 05 November 2017 12:20:13AM 0 points [-]

i mean figuring out the disagreements *and discussing them and resolving the disagreements*. actually figuring out which positions are correct and why. not just agreeing to disagree.

Lumifer 05 November 2017 02:46:58AM 0 points [-]

I suspect we have disagreements about what does "correct" mean and what criteria of correctness we can use to establish it :-)

But we can start by figuring out the precise questions to which we answer differently. Do you have any guesses?

curi 05 November 2017 06:23:33AM 0 points [-]

You believe Bayesian Epistemology and I believe Critical Rationalism. They disagree about e.g. induction, empiricism, instrumentalism.

Lumifer 05 November 2017 06:57:46AM 0 points [-]

Labels aren't terribly useful.

Let's start with the basics. We'll probably agree that external/objective reality exists. That we can gain some knowledge of that reality, and that this knowledge cannot be perfect. So far so good?

Thus we have reality and we have imperfect models of this reality in our heads. What happens when we have multiple models for same piece of reality?

curi 05 November 2017 08:02:13AM 0 points [-]

Labels aren't terribly useful.

Why not? They have meanings which people familiar with the field have substantial convergence about.

[Link] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians

by curi | 1v (0c)

[Link] Intercellular competition and the inevitability of multicellular aging

by Gunnar_Zarncke | 1v (3c)

Announcing the AI Alignment Prize

by cousin_it | 6v (2c)

Problems as dragons and papercuts

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

Let's start with the basics. We'll probably agree that external/objective reality exists. That we can gain some knowledge of that reality, and that this knowledge cannot be perfect. So far so good?

Yes.

Thus we have reality and we have imperfect models of this reality in our heads.

Yes, and these models are not merely about prediction.

What happens when we have multiple models for same piece of reality?

Critical arguments.

Lumifer 05 November 2017 10:16:52PM 0 points [-]

Why not?

Because most discussions suffer from the problem of different people understanding the same word differently. This is especially pronounced for labels (aka shortcuts to complicated concepts).

Critical arguments.

Hold on. First, is it acceptable to have multiple models at the same time? Do you *have* to declare one of them the best? It's not uncommon to have many models none of which you can falsify at the moment, how do you sort them out?

[continue this thread »](#)

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Lumifer comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: ai

Comments (117)

Sort By: Best

You are viewing a single comment's thread. [Show more comments above](#).

Lumifer 05 November 2017 10:16:52PM 0 points

Why not?

Because most discussions suffer from the problem of different people understanding the same word differently. This is especially pronounced for labels (aka shortcuts to complicated concepts).

Critical arguments.

Hold on. First, is it acceptable to have multiple models at the same time? Do you *have* to declare one of them the best? It's not uncommon to have many models none of which you can falsify at the moment, how do you sort

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

them out?

curi 05 November 2017 11:06:21PM 0 points [-]

You can "have" multiple models in the sense of knowing about them and being able to use them if you wanted to.

But if your N models all contradict, then at least N-1 of them are wrong. So you shouldn't simultaneously believe 2+ of them are true.

You can always have a non-refuted idea about how to proceed in life (with low enough resource cost, not with e.g. infinite time). This stuff is covered at length but is complicated to learn. Are you interested in doing things like reading a bunch and discussing it as you go along so you can learn it?

Lumifer 06 November 2017 04:07:57AM 0 points [-]

you shouldn't simultaneously believe 2+ of them are true

Is "true" a binary value or you can have fractions? Is it possible for a model to be X% true?

Also, you have N models which contradict somewhere (otherwise they would be identical). You can't falsify any of them at the moment. How do you go about selecting between them?

Are you interested in doing things like reading a bunch and discussing it as you go along so you can learn it?

No. As I pointed out before, I am not interested in being taught.

curi 06 November 2017 04:58:29AM * 0 points [-]

Is "true" a binary value or you can have fractions? Is it possible for a model to be X% true?

Binary.

Also, you have N models which contradict somewhere (otherwise they would be identical). You can't falsify any of them at the moment. How do you go about selecting between them?

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as a
Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

I said we have answers to this but they are complicated, and you said you don't want to read enough to understand them. I don't know why you're repeating the question. Do you disbelieve me that understanding it in a few paragraphs of forum discussion is unrealistic?

I'm totally open to discussing approaches to resolving disagreement, but I'm not open to you simply ignoring my suggestions about how to proceed and then trying to proceed in a way I don't think will work without saying why I'm mistaken about it being a bad approach. I'm also open to discussing where it's worth spending time and why, and how to decide that, and addressing skepticism. One approach is to start reading things and stop at the first thing you think is a mistake or have a question about, then comment. If you think you find a mistake, you only read more if it's fixed or you then discover you were mistaken about the mistake.

| No. As I pointed out before, I am not interested in being taught.

What if you're mistaken? Is the plan to stay mistaken, or is there a way to become less wrong? Do you have some kind of alternative you think is better which lets you learn all the important things while e.g. avoiding reading?

Lumifer 06 November 2017 06:23:36AM 0 points [-]

| Binary.

That's interesting. We've agreed that all models are imperfect representations of reality. Why are some imperfect models true? From a certain point of view all of them are false.

| Do you disbelieve me that understanding it in a few paragraphs of forum discussion is unrealistic?

Why, yes, I do. If you can't concisely draw at least the outlines of your position, I might even disbelieve that you understand your own views.

| simply ignoring my suggestions

You have reading comprehension problems. I didn't ignore them -- I explicitly said I don't agree to them. Let me repeat: I am not interested in being assigned a reading list.

| What if you're mistaken?

That's quite possible. But as I noted, reading comprehension is useful: I am quite interested in learning, I am quite uninterested in being taught, at least in this context.

[Link] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians

by curi | 1v (0c)

[Link] Intercellular competition and the inevitability of multicellular aging

by Gunnar_Zarncke | 1v (3c)

Announcing the AI Alignment Prize

by cousin_it | 6v (2c)

Problems as dragons and papercuts

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

curi 06 November 2017 06:32:54AM * 0 points [-]

Why are some imperfect models true?

I didn't say that they are.

You didn't ask for an outline, you asked for an answer. Those are different things.

It's hard to give outlines of solutions to people who are unfamiliar with the framework your solution is designed within, and who don't want to put in effort. Do you understand that problem? I also don't see the point of trying to write custom material for you (under the extra constraint of keeping it very short, and while having very limited info about you to customize with), when you don't want to read the canonical stuff, and I don't expect you to still be speaking to me in a few days (because of repeated indicators of hostility and disinterest).

Under what circumstances would you learn or answer David Deutsch on epistemology?

Lumifer 06 November 2017 03:37:35PM 0 points [-]

I didn't say that they are.

So what is it that you are you saying?

All models are imperfect. If some are true, my question stands. If none are true, I don't see any use for the concept of "true".

Do you understand that problem?

No, I don't. If you are unable to explain your position other than by saying "Go read the book, it explains everything", I have an inclination to think you yourself don't understand what you are trying to say.

I don't expect you to "write custom material". I expect you to be able to hold a *conversation* where you are can put forward your views in a clear and concise manner.

answer David Deutsch on epistemology

The thing is, if I want to go read Popper or Deutsch, I can go read Popper or Deutsch. I don't see what you will be able to add to my reading -- I can do it myself.

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

curi 06 November 2017 06:25:02PM 0 points



No, I don't. If you are unable to explain your position other than by saying "Go read the book, it explains everything", I have an inclination to think you yourself don't understand what you are trying to say.

Would that be your belief if I wrote the book? If I was in the acknowledgements of the book? If my best friend wrote it and I'd discussed the material at length with him? If the author was a fan of mine? You seem to be trying to judge credentials in some way without saying where the lines are, and without asking any questions about mine.

And I didn't say go read the book it explains everything, I said I don't want to rewrite the book for you, so start reading it and reply when you have your first criticism or question – just as if it was a forum post.

The thing is, if I want to go read Popper or Deutsch, I can go read Popper or Deutsch. I don't see what you will be able to add to my reading -- I can do it myself.

You will have comments when you read – questions and criticisms – which we can discuss as they come up. That's different than reading it alone.

Why do you want me to rewrite canonical material? Are you going to refuse to read any links or references of any kind, ever? What are the rules for when you do read those?

Why *don't* you go read DD/Popper, with or without me? Have you answered them? Have you any reference answering them? If not, why leave outstanding criticisms unanswered? Isn't that problematic?

The point is, you seem to draw some important distinction between 1) material I take responsibility for but didn't personally write (maybe you're assuming I won't take responsibility for things I reference the same as if I wrote them? I will.) 2) material I wrote in the past. 3) material I wrote specifically for this conversation. We have a massive archive of writing, some of it very polished. It refutes various claims LW makes that you seem to believe. You haven't answered it. You don't seem to know of anything that answers it. Yet you aren't interested. Under what circumstances would you be interested?

If you want material to be customized for you in some way, i don't know what way it is. If you read a little canonical stuff and said "This isn't working for me b/c it targets audience X and I

have trait Y" then I could help bridge that gap for you. But you haven't expressed any objection of that type.

Lumifer 06 November 2017 06:58:26PM * 0 points



I looked you up. That clarifies a lot:

I'm a philosopher. ... I sell educational philosophy material and philosophy consulting.

All problems can be solved by knowing how. I tell you how. I figure out your problems and their solutions. I help you learn anything you'd like to learn. ... Your life could be better. I can help.

Oh, and LOL at

I have expert knowledge in parenting/education, physics, economics, evolution, psychiatry, social dynamics, relationships, business, politics, and some parts of history.

So, Mr. Expert On Everything (and "world class at several computer games"), I am sorry but I'm looking neither for a philosophy tutor nor for a life mentor. You want to *teach* me: I do not desire to *be taught*.

We can talk about interesting problems, e.g. in epistemology, but if your position is that reading your favourite book has to be the beginning of all discussions, we're not going to get anywhere.

Would that be your belief if I wrote the book?

Yes. Inability to clearly and succinctly formulate the main tenets points to a lack of understanding. I don't care if you're friends with Deutsch and spent a lot of time chatting with him.

Have you answered them? Have you any reference answering them?

"Answering" Popperianism requires a book-length effort at least. I don't see any reason for me to spend that effort. As to references, Popper published many decades ago. Since the entire world hasn't converted to his views, I would expect to find a lot of references which disagree with him and Deutsch. Surely you're not arguing that there are none?

you seem to draw some important distinction

No, I do not. I'm attempting to hold a small, local, mostly self-contained conversation about epistemology where we can build certain structures out of certain well-defined words and see if they fail under stress. You want to turn it into an educational "read the textbook" session.

| refutes various claims LW makes that you seem to believe

So, be specific. Which claims do you think I believe? Please list and refute.

As to "a massive archive of writing", yes, indeed we do. Much of it **disagrees** with Popper and Deutsch. So what?

| If you want material to be customized for you

Sigh. Let me repeat once again, in caps and bold:

I DO NOT WANT TO BE TAUGHT BY YOU.

curi 06 November 2017 07:10:39PM * 0 points [-]

When I rewrite canonical material for this discussion, what should I change from the original? How should it differ from copy/pasting passages? Should I just paste stuff from sources and not tell you it's pasted, and then you'll engage with it?

| As to "a massive archive of writing", yes, indeed we do. Much of it disagrees with
| Popper and Deutsch. So what?

Where is the argument that CR is mistaken? CR provides arguments that Bayesian Epistemology is mistaken.

| "Answering" Popperianism requires a book-length effort at least. I don't see any
| reason for me to spend that effort.

Has anyone done it? If not, do you see a problem there? If so, can you give a reference that you will take responsibility for?

| Surely you're not arguing that there are none?

I claim none of the existing criticism of CR is correct. I take it you don't know of any that's correct, but wish to ignore the matter anyway. Why? Under what circumstances do you think arguments should be answered instead of ignored? Only when popular?

Laughing at me is rude and a non-argument. Yelling is also rude.

You seem to be hostile to the idea of discussing methodology before discussing a particular topic, even though we disagree about methodology. Do you think discussion methodology is unimportant and boring?

[continue this thread »](#)

LessWrong DISCUSSION

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Custom Search

curi comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: [ai](#)

Comments (117)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above](#).

curi 06 November 2017 07:10:39PM * 0 points

When I rewrite canonical material for this discussion, what should I change from the original? How should it differ from copy/pasting passages? Should I just paste stuff from sources and not tell you it's pasted, and then you'll engage with it?

As to "a massive archive of writing", yes, indeed we do. Much of it disagrees with Popper and Deutsch. So what?

Where is the argument that CR is mistaken? CR provides arguments that Bayesian Epistemology is mistaken.

alanforr

Create new article

Karma Score

Submit a new link

17

Add new meetup

25

Preferences



Log out

Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

"Answering" Popperianism requires a book-length effort at least. I don't see any reason for me to spend that effort.

Has anyone done it? If not, do you see a problem there? If so, can you give a reference that you will take responsibility for?

Surely you're not arguing that there are none?

I claim none of the existing criticism of CR is correct. I take it you don't know of any that's correct, but wish to ignore the matter anyway. Why? Under what circumstances do you think arguments should be answered instead of ignored? Only when popular?

Laughing at me is rude and a non-argument. Yelling is also rude.

You seem to be hostile to the idea of discussing methodology before discussing a particular topic, even though we disagree about methodology. Do you think discussion methodology is unimportant and boring?

Lumifer 06 November 2017 07:35:39PM 0 points

[-]

When I rewrite canonical material for this discussion, what should I change from the original?

I don't expect you to be a copy-and-paste bot, I expect you to be able **to hold a conversation**. I don't particularly care whether you quote, modify, or invent from scratch. You have been remarkably resistant to a concise formulation of your views relevant to our discussion -- if you feel that nothing less than Deutsch's whole book can do, well, we have a problem.

I claim none of the existing criticism of CR is correct.

Of course you do :-)

Under what circumstances do you think arguments should be answered instead of ignored?

Notice how you have NOT presented any arguments to be answered. You merely pointed in the general direction of a philosophical theory which claims (don't they all?) to have the answers.

Laughing at me is rude and a non-argument. Yelling is also rude.

I laugh a lot. Laughing is good.

As to yelling, well, you ignored that sentence only what, three or four times? Do you hear me now? :-)

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

You seem to be hostile to the idea of discussing methodology before discussing a particular topic

Au contraire. I would like to discuss methodology, but "go read a book" is not discussion.

curi 06 November 2017 07:39:55PM * 0 points [-]

I'm trying to discuss the methodology of reading text until your first comment/question/criticism and then replying. You have been ignoring this.

I did not ignore you about teaching, I heard you and I'm trying to have a peer discussion. But you keep interpreting things differently than I do. OK, understandable, but you need to be tolerant of different perspectives instead of yell. I am not yelling about being ignored about the methodology point about looking for the first mistake.

I can give you specific sources with details but first I asked if you'd be willing to look at them and you said *no*, so that's why I didn't actually give you a specific reference. I'm also bringing up that there is literature criticizing your school of thought, which your school of thought seems to have no answer to – isn't that a problem? Or what is your methodology such that that is ignorable? Or do you deny this is the case?

We disagree about e.g. induction. So you want me to rewrite one of the arguments about induction I've written in the past, because you don't want a reference. Right? I don't understand the purpose of this. Explain? It sounds like duplication of work to me.

Lumifer 06 November 2017 08:13:29PM 0 points [-]

I'm trying to discuss the methodology of reading text until your first comment/question/criticism and then replying

So you want to do exegesis. That makes the subject of the inquiry the text itself and the meaning contained in it.

The issue is that I'm not particularly interested in the text and CR. I'm interested in basic epistemological approaches of which CR is merely one. It's basically the difference between dissecting frogs and reading a book about the proper ways to dissect frogs and what you would find if you cut one open. In this case I want to dissect frogs and not read books.

I am not yelling about being ignored about the methodology point

I'm not *ignoring* it -- I'm explicitly telling you I *don't want* it.

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)

by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)

by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)

by cousin_it | 6v (2c)

[Problems as dragons and papercuts](#)

by Elo | 1v (1c)

LATEST OPEN THREAD

[By the way, is there an explanation](#)

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

[I learned something new. As a](#)

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

there is literature criticizing your school of thought

Oh boy. From the fact that you found me on LW you immediately deduced what my school of thought is? That might have been... hasty :-) And remember, I told you that labels are not terribly useful?

We disagree about e.g. induction.

We do? Did I say anything about induction? I'm sure there is a strawman waiting in the wings to be conveniently demolished, but what does it have to do with me?

curi 06 November 2017 08:21:28PM 0 points [-]

We disagree about e.g. induction.

We do? Did I say anything about induction? I'm sure there is a strawman waiting in the wings to be conveniently demolished, but what does it have to do with me?

I have been paying attention to what you wrote, e.g.:

And I still don't understand what's wrong with plain-vanilla observation as a way to acquire knowledge.

This statement indicates to me that we disagree about induction.

The issue is that I'm not particularly interested in the text and CR. I'm interested in basic epistemological approaches of which CR is merely one. It's basically the difference between dissecting frogs and reading a book about the proper ways to dissect frogs and what you would find if you cut one open. In this case I want to dissect frogs and not read books.

What exactly do you think is different btwn a text by DD, a text by me, and new text typed by me into this forum? To me they are all text, but you treat them totally different. Plz explain the methodology.

You express your disinterest in CR. Since I'm writing CR ideas, I take that as disinterest in what I'm saying. What would it take for you to become interested and try to address all known criticisms of your positions? Also do you have a website where you've written down your views to expose them to criticism, or do you have a reference which does this for you and which you'll take responsibility for?

Lumifer 06 November 2017 08:35:30PM 0 points [-]

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

| This statement indicates to me that we disagree about induction.

Induction isn't about acquiring knowledge from observations, induction is about generalizing from some limited set of observations to universal rules/laws.

| To me they are all text, but you treat them totally different

The key is a couple of comments up:

| I'm attempting to hold a small, local, mostly self-contained conversation about epistemology where we can build certain structures out of certain well-defined words and see if they fail under stress.

Note: **local**. Note: **self-contained**.

As I said, I don't care if you quote or write original text. What I'm looking for is small, specific, limited in scope.

| Since I'm writing CR ideas, I take that as disinterest in what I'm saying.

To the extent you're promoting/popularizing CR, yes, I'm uninterested in being swayed to its side.

| What would it take for you to become interested and try to address all known criticisms of your positions?

Time. Loads and loads of free time :-D

| do you have a website where you've written down your views to expose them to criticism, or do you have a reference which does this for you and which you'll take responsibility for?

Nope and nope. Sorry.

curi 06 November 2017 08:42:05PM * 0 points



| Induction isn't about acquiring knowledge from observations, induction is about generalizing from some limited set of observations to universal rules/laws.

I don't understand how arguing with me about induction is going to prove your point that we *don't disagree*.

Note: local. Note: self-contained.

Why do you want it to be local and self-contained? I don't want to exclude important ideas based on their source. I want to judge ideas by their substance, regardless of their source. But you started objecting to that, so here we are and I've tried many times to get you to clarify your methodology. I'm now trying again, despite the yelling and ridiculing.

I also don't know what your rules are – if I wrote something a month ago, can I link that? Yesterday, but it was originally for some other conversation? So I've been trying to find out what your methodology rules are, because I literally don't know what you consider allowable in the conversation or not, plus also I think I disagree iwth your methodology (but I'm still trying to clarify it).

To the extent you're promoting/popularizing CR, yes, I'm uninterested in being swayed to its side.

What if it's correct and you're mistaken? This isn't a matter of sides, but truth. I read you as saying you don't care about the truth if CR is true, but I guess you mean something else – what?

Time. Loads and loads of free time :-D

What would convince you to reallocate time? If you don't have time to think much, we could just stop now... I organized my life to have time to deal with ideas.

Nope and nope. Sorry.

Why not? Are you very interested in ideas? Are you young and new to trying to trying to understand things? Old and new? Don't see the value of a website or any kind of canonical statements of your views?

Lumifer 06 November 2017 09:12:11PM * 0 points [-]

I don't understand how arguing with me about induction is going to prove your point that we don't disagree.

Oh, but it's *meta* arguing! :-)

In any case, the point is that you assume I hold some positions without any... support for these assumptions.

| Why do you want it to be local and self-contained?

("Local" doesn't mean you can't bring it quotes from a book. It means none of your arguments are incorporated by reference but instead have to be fully included in the text of the thread)

Basically to prevent the conversation from losing shape and clarity. Most philosophical discussions tend to sink into the quicksand of subtly (or not so subtly) different definitions for words used and degenerate into mutually-incomprehensible stand-offs or splotchy messes.

Also -- a fun observation -- a lot of people adept at quoting from sources turn out to have a very shaky understanding of what these sources actually mean and what the implications are (this is a general observation, not aimed at you in particular).

The rules are the rules of a *conversation*: you talk/type in easily digestible chunks, you can quote anything you want but don't use "pointers" (points over yonder: "that thing over there proves my point, go check it out if you doubt it"), pass your variables by value. It would help if you give hard definitions for the terms you use.

| What if it's correct and you're mistaken?

We haven't figured out what does "correct" mean :-)

| I read you as saying you don't care about the truth if CR is true, but I guess you mean something else – what?

My time and attention are limited. I don't feel establishing the validity of CR should be at the top of my to-do list.

| What would convince you to reallocate time?

Changes in relative importance of things. There is a local saying coming from Eliezer that beliefs should pay rent. If the validity of CR starts to affect my life in major ways, I would reallocate time to thinking about it.

And you realize, of course, that there are a great many more ideas than CR, so even you decided to dedicate your life to "deal with ideas", CR is *still* not the obvious choice.

| Why not?

There is a variety of reasons. One is that I'm not particularly interested in converting everyone to my worldview. Another is that it changes on occasion. Yet another is that putting up a vanity website would do pretty much nothing useful for me.

curi 06 November 2017 09:21:21PM * 0 points [-]

If the validity of CR starts to affect my life in major ways, I would reallocate time to thinking about it.

I bet it does. What do you do and what are some of your main philosophical beliefs which you would think it's important if they're mistaken? (I'll be happy to answer the same question though not with any use of pointers to my websites banned.)

And you realize, of course, that there are a great many more ideas than CR, so even you decided to dedicate your life to "deal with ideas", CR is still not the obvious choice.

I reviewed all the well known options (and some but not all obscure ones – and I don't mind reviewing more obscure ones when someone interested in conversation brings one up) and made a judgement about which is correct and non-refuted, and that all the others are refuted by arguments I know. In epistemology, that one is CR.

I would expect other people to attempt something like this, but I find they normally haven't – and don't want to begin. Does this sort of project interest you? If not, what sort of truth-seeking does interest you?

And if you want me to put in extra work to use fewer references than I normally would – do you have any value to offer to motivate me to do this? For example, do you think you'll continue the conversation to a conclusion? Most people don't, and I currently don't expect you to, and I'd rather not jump through a bunch of hoops for you and then you just stop responding.

Lumifer 06 November 2017 09:40:12PM 0 points [-]

I bet it does.

What exactly is the falsifiable claim that you're making and how would you expect it to be falsified? :-)

some of your main philosophical beliefs which you would think it's important if they're mistaken?

Oh, there are lot. Existence of afterlife, for example. The nature of morality. Things like that.

I reviewed all the well known options (and some but not all obscure ones ...) and made a judgement about which is correct

How confident are you of your judgement?

Does this sort of project interest you?

Not particularly because of lack of relevancy (see above about paying rent). I don't feel the need to pass a judgement on a set of options if that choice will lead to zero change.

do you think you'll continue the conversation to a conclusion?

I don't expect this conversation to have a conclusion in the sense of general agreement that A is wrong and B is correct. I view it more as a -- to use a Culture name -- *A Frank Exchange Of Views* which might lead to new information being exchanged, new angles of view opened, maybe even new perspectives -- but nothing as decisive as a sharp-edged black-and-white conclusion.

[continue this thread »](#)

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Lumifer comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: ai

Comments (117)

Sort By: Best

You are viewing a single comment's thread. [Show more comments above](#).

Lumifer 06 November 2017 09:40:12PM 0 points

I bet it does.

What exactly is the falsifiable claim that you're making and how would you expect it to be falsified? :-)

some of your main philosophical beliefs which you would think it's important if they're mistaken?

Oh, there are lot. Existence of afterlife, for example. The nature of morality. Things like that.

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

I reviewed all the well known options (and some but not all obscure ones ...) and made a judgement about which is correct

How confident are you of your judgement?

Does this sort of project interest you?

Not particularly because of lack of relevancy (see above about paying rent). I don't feel the need to pass a judgement on a set of options if that choice will lead to zero change.

do you think you'll continue the conversation to a conclusion?

I don't expect this conversation to have a conclusion in the sense of general agreement that A is wrong and B is correct. I view it more as a -- to use a Culture name -- *A Frank Exchange Of Views* which might lead to new information being exchanged, new angles of view opened, maybe even new perspectives -- but nothing as decisive as a sharp-edged black-and-white conclusion.

curi 06 November 2017 09:49:21PM 0 points

[-]

Oh, there are lot. Existence of afterlife, for example. The nature of morality. Things like that.

Will you briefly indicate some specifics, especially things you think CR might disagree about?

How confident are you of your judgement?

Very, because I've put a great deal of effort (as have some others) into doing this investigation, finding people who believe I'm mistaken and are willing to discuss, etc. There are no major outstanding leads left that need checking but haven't been checked. I genuinely don't know what more I could do that would make a big difference. I can do some lesser things like double check more things that have been singled checked, or make more websites and optimize them more and get more traffic to them so that there's more potential criticism (both raw traffic quantity and also getting specific smart ppl).

Not particularly because of lack of relevancy (see above about paying rent). I don't feel the need to pass a judgement on a set of options if that choice will lead to zero change.

Why do you think knowing what way of thinking is correct would lead to zero change? It led to tons of change for me. For you, I'd expect it to mean re-evaluating more or less your entire life and making huge changes. Areas of change-implication include parenting, relationships/marriage, how to discuss, induction, views on science and ways of judging scientific claims, approach to AGI, etc.

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

but nothing as decisive as a sharp-edged black-and-white conclusion.

Do you think that sort of conclusion is a valuable thing to reach in general? About some issues? I do.

Lumifer 06 November 2017 10:03:58PM 0 points [-]

things you think CR might disagree about?

These things are orthogonal to CR, CR standing or falling does not affect them.

That's precisely the reason I'm not terribly interested in heavily engaging with CR.

Very

From my point of view it's a bad sign.

I'd expect it to mean re-evaluating more or less your entire life and making huge changes. Areas of change-implication include parenting, relationships/marriage, how to discuss, induction, views on science and ways of judging scientific claims, approach to AGI, etc.

How so? I don't see why changing views on epistemology would lead a different approach to, say, marriage or parenting.

Do you think that sort of conclusion is a valuable thing to reach in general?

Valuable, but rarely available for issues of importance.

curi 06 November 2017 10:08:19PM * 0 points [-]

Epistemology is the field which says how knowledge is created.

Solutions to problems are a type of knowledge.

How to solve problems in a marriage is therefore determined substantially by epistemology.

Education of children is primarily an issue of helping them create knowledge. How to do this depends on how knowledge is created.

You're mistaken about what is orthogonal to CR. You mentioned afterlife – what to believe about that is a matter of judging arguments (or put another way: creating knowledge of whether there is or isn't an

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)

by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)

by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)

by cousin_it | 6v (2c)

[Problems as dragons and papercuts](#)

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

afterlife), and for that you need epistemology which is the field that tells you the methods of discussing and evaluating ideas. You also mentioned morality. Moral argument is governed by epistemology, and also lots of morality is basically derived from epistemology because morality is about how to live and some of the key things about how to live are to live in an rational, error-correcting and problem-solving way.

| Valuable, but rarely available for issues of importance.

What if it was routinely available, if you knew how? That's what my epistemology says. So there's impact-on-life there!

| From my point of view it's a bad sign.

If you can suggest a way I should change my methods for judging this, please share it. (If you have preliminary questions first, feel free to ask them!)

Lumifer 06 November 2017 10:20:30PM 0 points [-]

| How to solve problems in a marriage is therefore determined substantially by epistemology.

Cute play with words, but bears no relationship to the real world. Ditto for parenting. Ditto for afterlife.

You're offering a version of the argument that since physics deals with the lowest (most basic) levels of matter, all other sciences are (or should be) physics: chemistry, biology, sociology, etc. So solving problems in marriage is physics because you are both made out of atoms.

| What if it was routinely available

We have a basic disagreement: you think that models are either true or not, and I think, to quote George Box, that "All models are wrong but some are useful".

| change my methods for judging this

Rely less on whether someone can successfully argue something and more on empirical reality.

curi 06 November 2017 10:24:08PM * 0 points [-]

I'm not playing with words, I'm expressing the CR perspective. You apparently disagree, but if CR is correct then what I said is correct. So CR's correctness has consequences for your life.

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

I am not offering reductionism. Married people *literally* do things like discuss disagreements and try to solve problems – exactly the kind of thing CR governs. That doesn't mean CR is the only thing you need to know – you *also* need to know relationship-specific stuff (which you btw need to *learn* – and so CR is relevant there).

| We have a basic disagreement: you think that models are either true or not,

I think many ideas aren't models. This is a CR belief which would have impacts on your thinking if you understood it and decided it was correct.

| Rely less on whether someone can successfully argue something and more on empirical reality.

Can you be more specific? How does anything I'm doing or saying clash with reality? Arguments about reality are totally welcome, and I've both sought them out and created them myself.

BTW CR philosopher David Deutsch is literally a founder of a parenting/educational movement. Here is one of my essays about CR and parenting: <http://fallibleideas.com/taking-children-seriously>

Lumifer 07 November 2017 04:10:31PM 0 points



| I'm expressing the CR perspective

So what is the domain that CR claims? I thought it was merely epistemology, but apparently it includes marital counseling and parenting advice?

By the way, your style pattern-matches to religious proselytizing very well.

| I think many ideas aren't models.

So far we had the underlying reality and imperfect representations thereof which we called "models". What is an "idea"?

| Can you be more specific?

You said

| I've put a great deal of effort (as have some others) into doing this investigation, finding people who believe I'm mistaken and are willing to discuss, etc. ... make more websites and optimize them more and get more traffic to them so that there's more potential criticism

You're looking for criticism from people, not from reality.

Think about it this way: let's say you have an idea about how to make a killing in financial markets. Your understanding of how to figure out whether it works is to ask all your friends and interested strangers (IRL and on the 'net) to criticize it. If they can't convince you it's bad, you declare it good.

But there is another way -- you don't ask anyone's opinion, but instead actually attempt to trade it and see if it works.

I prefer the second type of testing claims to the first one.

curi 07 November 2017 05:54:03PM * 0 points [-]

| So what is the domain that CR claims?

CR is an epistemology. It has *implications*, not domain claims.

Methods of thinking are used in *every* field!

| By the way, your style pattern-matches to religious proselytizing very well.

Can you link an example? I'm skeptical but I'd like to read something similar to my writing.

| You're looking for criticism from people, not from reality.

I've done both. But the primary issue here is critical argument, not testing, b/c it's about philosophy, not science. My tests are anecdotal and don't really matter to the discussion.

If there's a particular test you think is important for me to do, what is it?

EDIT forgot link about ideas <http://fallibleideas.com/ideas>

Lumifer 07 November 2017 06:23:29PM * 0 points [-]

| It has implications, not domain claims.

You were much more gung ho about it just a little bit earlier:

|| Epistemology is the field which says how knowledge is created. Solutions to problems are a type of knowledge. How to solve problems ... is therefore

determined substantially by epistemology.

...Moral argument is governed by epistemology, and also lots of morality is basically derived from epistemology

and on your website you're quite explicit that your approach can solve ALL problems.

Can you link an example?

Not so much writing style, but argumentative style. Basically, you comments try to set in a number of hooks (like "*This stuff is covered at length but is complicated to learn. Are you interested in doing things like reading a bunch and discussing it as you go along so you can learn it?*" or "*What do you do and what are some of your main philosophical beliefs which you would think it's important if they're mistaken?*"), these hooks have a line and all lines lead back to "*start reading this book and let's discuss it*" which is where you really want to end up. And there is the promise that this philosophy will significantly influence my entire life.

I see this as having a lot of parallels with classic proselytizing, say, Christian, where you set your hooks ("Are you unhappy? Does life make no sense to you?"), all lines lead to reading the Good News and inviting Jesus into your heart and, of course, once you accept Him into your life, that life is supposed to change dramatically.

But the primary issue here is critical argument, not testing, b/c it's about philosophy, not science.

Note another disagreement point: about the relative value of critical arguments vs empirical testing :-)

If there's a particular test you think is important for me to do, what is it?

The standard one: does it work?

For example, you are offering parenting advice. Does it work? How do you know? Ditto for all the other kind of life advice that you offer and want to charge for.

curi 07 November 2017 06:33:31PM 0 points



Yes my philosophy works great. I have a great life, lots of success, etc, etc.

This is anecdotal and open to debate about how to interpret the test results. I don't wish to switch from debating ideas to sharing tons of personal info and debating my life choices (some of which are successful at non-standard values, and so will appear unsuccessful, and the right values have to be debated to judge it, and etc etc).

Even if my personal life was a mess, that still wouldn't refute my philosophy. That wouldn't be an argument which refutes any particular epistemology claim.

You seem to object to the concept of critical argument, and its role as the method of dealing with many issues.

▮ You were much more gung ho about it just a little bit earlier:

I don't see the difference. Implications are a big deal.

[continue this thread »](#)

LessWrong DISCUSSION

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Custom Search

curi comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: [ai](#)

Comments (117)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above](#).

curi 07 November 2017 06:33:31PM 0 points



Yes my philosophy works great. I have a great life, lots of success, etc, etc.

This is anecdotal and open to debate about how to interpret the test results. I don't wish to switch from debating ideas to sharing tons of personal info and debating my life choices (some of which are successful at non-standard values, and so will appear unsuccessful, and the right values have to be debated to judge it, and etc etc).

Even if my personal life was a mess, that still wouldn't refute my philosophy. That wouldn't be an argument which refutes any particular epistemology claim.

alanforr

Karma Score
17
25

Create new article

Submit a new link

Add new meetup

Preferences

Log out

Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067->
by curi on Less Wrong Lacks
Representatives and Paths Forward |

You seem to object to the concept of critical argument, and its role as the method of dealing with many issues.

You were much more gung ho about it just a little bit earlier:

I don't see the difference. Implications are a big deal.

Lumifer 07 November 2017 07:14:37PM 0 points

This is anecdotal

I don't mean your personal life.

You offer advice professionally. How do you know that your advice leads to desired outcomes? Does it? In which percentage of cases? Did you measure anything?

You seem to object to the concept of critical argument

I don't object to the concept. I object to it being sufficient to determine whether something is "true" (using your terminology) and to the idea that enough critical arguments can replace real-life testing.

I don't see the difference.

When people say "X has implications for this" and "This is determined substantially by X", these sentences usually have different meanings.

curi 07 November 2017 07:28:35PM 0 points

I have no interest in violating the privacy of my clients, or claiming my philosophy is good b/c of my consulting results. I'm not claiming that, so you don't need to challenge it.

Such methods could not settle the philosophical issues, anyway. I might communicate badly, My clients might be a non-random sample of people with very ambitious goals. My clients might not do what I advised. etc, etc, etc. Any empirical results would be logically compatible with my philosophy being true.

"This is determined substantially by X"

please don't paraphrase me incorrectly, in quote marks, while omitting any actual quote.

Lumifer 07 November 2017 07:34:28PM * 0 points

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

What does this have to do with the privacy of your clients? I am not asking you to tell me stories, I'm asking whether **you** have any metrics of the **performance** of the product that you're selling.

| Any empirical results would be logically compatible with my philosophy being true.

I thought you were Popperian. Is your philosophy empirically *falsifiable*, then?

| please don't paraphrase me incorrectly

Direct quote:

| | How to solve problems in a marriage is therefore determined substantially by epistemology.

curi 07 November 2017 07:50:34PM 0 points



Thanks for the quote; I was mistaken to say your paraphrase was incorrect. They're big implications. I don't see the point of this part of the discussion.

Popperians say *scientific* ideas should be (empirically) falsifiable. Philosophy isn't empirically falsifiable, it's addressed by critical arguments.

I do not use consulting metrics in marketing or other public statements; they relate to private matters; I'm not going to discuss them. However I thought of a better way to approach this:

I've given lots of advice, for free, in public, with permalinks. So, unlike my private consulting, I'll talk about that. Broadly here are the results:

Some people love my advice. Super fans! A larger number of people don't want to talk with me. Haters! (I'm intentionally saying the results are pretty polarized.)

How is that to settle anything? Are we to go by popular opinion? You brought this topic up to try to get away from people. But I regard this as being about people! And btw I don't know what metrics you would consider appropriate for this.

What I wanted to look at isn't people but critical arguments, and my claim is that FI is non-refuted – meaning not just that no refutation is known to me, but also that no one else knows one who is willing to share it. I think it's wise to survey the literature, take public comments, seek out discussions at a variety of forums, etc, in addition to thinking about it personally. That's a worthwhile extra step to help find refutations.

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)

by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)

by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)

by cousin_it | 6v (2c)

[Problems as dragons and papercuts](#)

by Elo | 1v (1c)

LATEST OPEN THREAD

[By the way, is there an explanation](#)

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

[I learned something new. As a](#)

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

So the thing I was talking about, as I see it, was fundamentally about ideas (particularly critical arguments), not people; and the thing you're bringing up is about what people do, how they react to advice, etc – about people rather than arguments/ideas.

I was trying to talk about the current objective state of the intellectual debate; you're bringing up the issue of how people react to me and what happens in their lives.

Lumifer 08 November 2017 02:08:21AM 0 points

[-]

| Philosophy isn't empirically falsifiable

Hold on, hold on. Your philosophy isn't abstract ruminations about the numbers of angels on the head of a pin. Your philosophy has *implications*. BIG implications. In fact, you're saying it changes people's lives!

And these are phenomena of the empirical realm. We can look at them. We can evaluate them. We can see if the "implications" actually lead to consequences that your philosophy predicts and expects. Unless your philosophy just shrugs and says "Beats me, I have no idea what these interventions will do", it makes **predictions** about these implications.

And the good thing about all these is that they are verifiable and falsifiable.

So.. how about testing these implications? If they fail, would you insist it has no bearing on the philosophy?

| I thought of a better way to approach this ... How is that to settle anything?

I agree, the public reaction to ideas doesn't tell you much. But how is this "a better way", then?

| What I wanted to look at isn't people but critical arguments

I was talking mostly about the whole of reality, not just people, and my point is that critical arguments by themselves are insufficient.

| the current objective state of the intellectual debate

What is the word "objective" doing in there?

| you're bringing up the issue of how people react to me

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

No, I don't. You just did. I'm talking about testing your ideas in reality, in particular, by the simplest test of whether they work.

curi 08 November 2017 04:08:54AM * 0 points [-]

As before, you don't know how CR works, we have *massive* philosophical differences, and your questions are based on assuming aspects of your philosophy are true. Are you interested in understanding a *different* perspective, or do you just want to challenge my ideas to meet the criteria your framework says matter?

Lumifer 08 November 2017 04:04:40PM 0 points [-]

| your questions are based on assuming aspects of your philosophy are true

I don't think so. At the moment we are operating in a very simple, almost crude, framework: there's reality, there are models, we can detect some mismatches between the reality and the models. Isn't falsification one of the favourite Popperian ideas?

| Are you interested in understanding a different perspective

I am asking you questions, am I not? And offering you -- what do you call them? ah -- critical arguments.

curi 08 November 2017 06:39:46PM 0 points [-]

| Popperians say scientific ideas should be (empirically) falsifiable. Philosophy isn't empirically falsifiable, it's addressed by critical arguments.

I let you take substantial control over conversation flow. You took it here – you overestimated your knowledge of Popper and were totally wrong. You do not seem to have learned from this error.

You didn't answer my question about your interest, and you seem totally lost as to what we disagree about. You're still, in response to "your questions are based on assuming aspects of your philosophy are true", making the same assumptions while denying it. You don't have anything like a sense of what we disagree about, but you're trying to lead the conversation

anyway. Your questions are in service of lines of argument, not finding out what I think – and the lines of argument don't make sense because you don't know what to target.

Lumifer 08 November 2017 06:51:03PM * 0 points [-]

| and were totally wrong

What exactly did I say that was totally wrong? Quote, please.

| making the same assumptions

These assumptions take half a sentence. There are exactly three of them:

| | there's reality, there are models, we can detect some mismatches between the
| reality and the models

Which one do you think is unjustified?

| the lines of argument don't make sense because you don't know what to target

Supply me with targets, then :-D

[continue this thread »](#)

ChristianKI 08 November 2017 05:46:11PM 0 points [-]

| Isn't falsification one of the favourite Popperian ideas?

I don't think you are supposed to use it for the important models.

Lumifer 08 November 2017 06:36:56PM 0 points [-]

The ones too important to be falsified? :-D

LessWrong DISCUSSION

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Custom Search

Lumifer comments on Questions about AGI's Importance - Less Wrong Discussion

 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: [ai](#)

Comments (117)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above](#).

Lumifer 08 November 2017 06:51:03PM * 0 points 

| and were totally wrong

What exactly did I say that was totally wrong? Quote, please.

| making the same assumptions

These assumptions take half a sentence. There are exactly three of them:

| | there's reality, there are models, we can detect some mismatches between the reality and the models

alanforr

Karma Score
17
25



Create new article

Submit a new link

Add new meetup

Preferences

Log out

 Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067->
by curi on Less Wrong Lacks
Representatives and Paths Forward |

Which one do you think is unjustified?

the lines of argument don't make sense because you don't know what to target

Supply me with targets, then :-D

curi 08 November 2017 07:09:49PM * 0 points [-]

Quoting:

Any empirical results would be logically compatible with my philosophy being true.

I thought you were Popperian. Is your philosophy empirically *falsifiable*, then?

Popperians say *scientific* ideas should be (empirically) falsifiable. Philosophy isn't empirically falsifiable, it's addressed by critical arguments.

I regard this as indicating you misunderstand CR.

Then later:

Isn't falsification one of the favourite Popperian ideas?

In science, yes, testing is a favored idea, though even in science most ideas are rejected without being tested:

<http://curi.us/1504-the-most-important-improvement-to-popperian-philosophy-of-science>

But you don't want references, and I don't want to rewrite or copy/paste my blog post which is itself summarizing some information from books that would be better to look at directly.

I have a lot of targets on my websites, like <http://fallibleideas.com> and <https://reasonandmorality.com>, but you've said you don't want to look at them.

Do you have a website with information I could skim to find disagreements? Earlier, IIRC, I tried to ask about some of your important beliefs but you didn't put forward some positions to debate.

Is there any written philosophy material you think is correct, and would be super interested to learn contains mistakes? Or do you just think the ideas in your head are correct but they aren't written down, and you'd like to learn about mistakes in those? Or do you think your own ideas have some flaws, but are pretty good, so if I pointed out a couple mistakes it might not make much difference to you?

0 points

Thanks, it was not clear to me that it
by turchin on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by J_Thomas_Moros on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by entirelyuseless on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by turchin on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by turchin | 0v (4c)

**Fables grow around missed natural
experiments**
by MaryCh | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by curi | 1v (41c)

[Link] The Little Dragon is Dead
by SquirrelInHell | 0v (1c)

[Link] AGI
by curi | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by mirefek | 2v (8c)

What do you want to get out of this discussion? Coming to agree about some major philosophy issues would be a *big* effort. Under what sort of circumstances do you expect you would stop discussing? Do you have a discussion methodology which is written down anywhere? I do. <http://curi.us/1898-paths-forward-short-summary>

I have a philosophy I think is non-refuted. I don't know of any mistakes and would be happy to find out. It's also written down in public to expose it to scrutiny.

Lumifer 08 November 2017 07:53:32PM * 0 points [-]

I regard this as indicating you misunderstand CR

Your philosophy is advertised as "All problems can be solved by knowing how. I tell you how."

This looks to me as crossing the demarcation threshold. Would you insist that there are no possible empirical observations which can invalidate you advice?

Do you have a website with information I could skim to find disagreements? ... Is there any written philosophy material you think is correct, and would be super interested to learn contains mistakes?

You asked before. Still nope and nope.

Under what sort of circumstances do you expect you would stop discussing?

When you stop being interesting.

I don't know of any mistakes and would be happy to find out.

Define "mistake".

curi 08 November 2017 08:01:09PM * 0 points [-]

You can bring up observations in a discussion of a piece of advice, but as always the role of the evidence is governed by arguments stating its role. And the primary issue here is argument.

All problems can be solved by knowing how.

This is a theory claim.

I tell you how.

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)

by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)

by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)

by cousin_it | 6v (2c)

[Problems as dragons and papercuts](#)

by Elo | 1v (1c)

LATEST OPEN THREAD

[By the way, is there an explanation](#)

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

[I learned something new. As a](#)

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

This is a claim that I have substantial problem solving knowledge for sale, but is not intended to indicate I already know full solutions to all problems. It's sufficiently non-specific that I don't think it's a very good target for discussion.

| When you stop being interesting.

Why are you interested now?

| Define "mistake".

<http://fallibleideas.com/definitions>

And are you really unfamiliar with this common English word? Do you know what being wrong is? Less wrong? Error? Flaw?

Are you trying to raise some sort of philosophical issue? If so, please state it directly.

| | You asked before. Still nope and nope.

What about the rest?

| Or do you just think the ideas in your head are correct but they aren't written down, and you'd like to learn about mistakes in those? Or do you think your own ideas have some flaws, but are pretty good, so if I pointed out a couple mistakes it might not make much difference to you?

Lumifer 08 November 2017 08:29:48PM * 0 points

[-]

| Why are you interested now?

I'm interested in smart weird people :-P

| And are you really unfamiliar with this common English word?

Oh, boy. We are having fundamental philosophical disagreements and you think dictionary definitions of things like "wrong" are adequate?

You say that philosophy is not falsifiable. OK, let's assume that for the time being. So can we apply the term "wrong" to some philosophies and "right" to others? On which basis? You will say "critical arguments". What is a critical argument? Within which framework are you going to evaluate them? You want "mistakes" pointed out to you. What kind of things will you accept as a "mistake" and what kind of things will you accept as indicating that it's valid?

IlyaShpitser (16)
 curi (14)
 Dagon (11)
 NancyLebovitz (6)
 satt (6)
 cousin_it (6)
 Habryka (6)
 ChristianKI (6)
 SquirrelInHell (5)
 root (5)

RECENT KARMA AWARDS

I disagree that definitions are not all that important.

| do you just think the ideas in your head are correct

Well, obviously I think they are correct to some degree (remember, for me "truth" is not a binary category).

| and you'd like to learn about mistakes in those?

See above: what is a "mistake", given that we're deliberately ignoring empirical testing?

Things I'd like to learn are more like new to me frameworks, angles of view, reinterpretations of known facts. To use [Scott Alexander's terminology](#), I want to notice concept-shaped holes.

curi 08 November 2017 08:52:58PM 0 points

[-]

Criteria of mistakes are themselves open to discussion. Some typical important ways to point out mistakes are:

- 1) internal contradictions, logical errors
- 2) non sequiturs
- 3) a reason X wouldn't solve problem Y, even though X is being offered as a solution to Y
- 4) an idea assumes/uses and also contradicts some context (e.g. background knowledge)
- 5) pointing out a contradiction with evidence
- 6) pointing out ambiguity, vagueness

there are many other types of critical arguments. for example, sometimes an argument, X, claims to refute Y, but X, if correct, refutes *everything* (or everything in a relevant category). it's a generic argument that could equally well be used on everything, and is being selectively applied to Y. that's a criticism of X's capacity to criticize Y.

Ideas solve problems (put another way, they have purposes), with "problem" understood very broadly (including answering questions, explaining an issue, accomplishing a goal). A mistake is something which prevents an idea from solving a problem it's intended to solve (it fails to work for its purpose).

By correcting mistakes we get better ideas. We fix issues preventing our problems from being solved and our purposes achieved (including the purpose of correctly intellectually understanding philosophy, science, etc). We should prefer non-refuted ideas (no known mistakes) to refuted ideas (known mistakes).

Lumifer 08 November 2017 09:07:46PM 0 points



| Some typical important ways to point out mistakes

Ways to *point out* mistakes? Then the question remains: what is a "mistake"? A finger pointing at the moon is not the moon.

Your (4) is the same thing as (1) -- or (5), take your pick. Your (5) is forbidden here -- remember, we are deliberately keeping to one side of the demarcation threshold -- no empirical evidence or empirical testing allowed. (6) is quite curious -- is being vague a "mistake"?

| Ideas solve problems

In the real world? Then they are falsifiable and we can bring empirical evidence to bear. You were very anxious to avoid that.

| By correcting mistakes we get better ideas

Looks like a *non sequitur*: generating new (and better) ideas is quite distinct from fixing the errors of old ideas -- similar to the difference between writing a new program and debugging an existing one.

| We should prefer non-refuted ideas (no known mistakes) to refuted ideas (known mistakes).

I would argue that we should prefer ideas which successfully solve problems to ideas which solve them less successfully (demarcation! science! :-D)

curi 08 November 2017 09:37:50PM * 0 points



| Ways to point out mistakes? Then the question remains: what is a "mistake"? A finger pointing at the moon is not the moon.

I actually wrote a sentence

| | A mistake is [...]

Do you not read ahead before replying, and don't go back and edit either?

| (6) is quite curious -- is being vague a "mistake"?

In general, yes. It technically depends on context (like the problem specification details). Normally e.g. the context of answering a question is you want an adequately clear answer, so an inadequately clear answer fails.

| In the real world? Then they are falsifiable and we can bring empirical evidence to bear.
| You were very anxious to avoid that.

Ideas solve intellectual problems, and some of those solutions can be used to solve problems we care about in the real world by acting according to a solution. Some problems (e.g. in math) are more abstract and it's unclear what to use the solutions for.

I have nothing against the real world. But even when the real world is relevant, you still have to make an argument saying how to use some evidence in the intellectual debate. The intellectual debate is always primary. You can't just directly look at the world and know the answers, though sometimes the arguments involved with getting from evidence X to rejecting idea Y are sufficiently standard that people don't write them out.

You are welcome to mention some evidence in a criticism of my philosophy claims if you think you see a way to relevantly do that.

| Looks like a non sequitur: generating new (and better) ideas is quite distinct from fixing the errors of old ideas -- similar to the difference between writing a new program and debugging an existing one.

You have idea X (plus context) to solve problem P. You find a mistake, M. You come up with a new idea to solve P which doesn't have M. Whether it's a slightly adjusted version of X (X2) or a very different idea that solves the same problem is kinda immaterial. Both are acceptable. Methodologically, the standard recommendation is to look for X2 first.

| I would argue that we should prefer ideas which successfully solve problems to ideas which solve them less successfully (demarcation! science! :-D)

I consider solving a problem to be binary – X does or doesn't solve P. And I consider criticisms to be binary – either they are decisive (says why the idea doesn't work) or not.

Problems without success/failure criteria I consider inadequately specified. Informally we may get away with that, but when trying to be precise and running into difficult issues then we need to specify our problems better.

Lumifer 09 November 2017 01:31:36AM 0 points



I actually wrote a sentence

That's a curious definition of a "mistake". It's very... instrumental and local. A "mistake" is a function of both an idea and a problem -- therefore, it seems, if you didn't specify a particular problem you can't talk about ideas being mistaken. And yet your examples -- e.g. an internal logical inconsistency -- don't seem to require a problem to demonstrate that an idea is broken.

I have nothing against the real world

Oh, I'm sure it's relieved to hear that

But even when the real world is relevant, you still have to make an argument saying how to use some evidence in the intellectual debate.

Why is that?

The intellectual debate is always primary.

That's an interesting claim. An intellectual debate is what's happening inside your head. You are saying that it's primary compared to the objective reality outside of your head. Am I understanding you correctly?

I consider solving a problem to be binary – X does or doesn't solve P.

Only if a problem has a binary outcome. Not all problems do.

And I consider criticisms to be binary – either they are decisive (says why the idea doesn't work) or not.

A black-and-white vision seems unnecessary limiting.

Consider standard statistics. Let's say we're trying to figure out the influence of X on Y (where both are real values). First, there is no sharp boundary between a solution and a not-

solution. You can build a variety of statistical models which will make different trade-offs and produce different results. There is no natural dividing line between a slightly worse model which would be a not-solution and a slightly better model which will be a solution.

Moreover, since these different models are making trade-offs, you can criticise these trade-offs, but generally speaking it's difficult to say that this one is outright wrong and that one is clearly right. There's a reason they're called trade-offs.

Typically at the end you pick a statistical model or an ensemble of models, but the question "is the problem solved, yes or no?" is silly: it is solved to some extent, not fully, but it's not at the "we have no idea" stage either.

Problems without success/failure criteria I consider inadequately specified.

Life must be very inconvenient for you.

By the way, what about optimization problems? The goal is to maximize Y by manipulating X. There is no threshold, you want Y to be as large as possible. What's the criterion for success?

curi 09 November 2017 01:58:13AM 0 points 

That's a curious definition of a "mistake". It's very... instrumental and local.

This is not local – I specified context matters (whether the context is stated as part of the problem, or specified separately, is merely a matter of terminology.)

You can't determine whether a particular sentence is a correct or incorrect answer without knowing the context – e.g. what is it supposed to answer? The same statement can be a correct answer to one issue and an incorrect answer to a different issue. If you don't like this, you can build the problem and the context into the statement itself, and then evaluate it in isolation.

I'm guessing the reason you consider my view on mistakes "instrumental" is because I think one has to look at the purpose of an idea instead of just the raw data. It's because I add a philosophy layer where you don't. So your alternative to "instrumental" is to say something like "mistakes are when ideas fail to correspond to empirical reality" – and to ignore non-empirical issues, interpretation issues, and that answers to questions need to correspond *to the question* which could e.g. be about a hypothetical scenario. To the

extent that questions, goals, human problems, etc, are *part of reality* then, sure, this is all about reality. But I'm guessing we can both agree that's a difference of perspective.

And yet your examples -- e.g. an internal logical inconsistency -- don't seem to require a problem to demonstrate that an idea is broken.

Self-contradictory ideas are broken for *many* problems. In general, we try to criticize an idea as a solution to a range of problems, not a single one. Those criticisms are more interesting. If your criticism is too narrow, it won't work on a slight variant of the idea. You normally want to criticize all the variants sharing a particular theme.

Self-contradictory ideas can (as far as we know) only be correct solutions to some specific types of problems, like for use in parody or as a discussion example.

But even when the real world is relevant, you still have to make an argument saying how to use some evidence in the intellectual debate.

Why is that?

Because facts are not self-explanatory. Any set of facts is open to many interpretations. (Not equally correct interpretations or anything like that, merely logically possible interpretations. So you have to talk about your interpretation, unless the other person can guess it. And you have to talk about how your interpretation of the evidence fits into the debate – e.g. that in contradicts a particular claim – though, again, in simple cases other people may guess that without you saying it.)

That's an interesting claim. An intellectual debate is what's happening inside your head. You are saying that it's primary compared to the objective reality outside of your head. Am I understanding you correctly?

You may prefer to think of it as the philosophy issues are always prior to the other issues. E.g. the role of a particular piece of evidence in reaching some conclusion is governed by ideas and methodology about the role of evidence in general, an interpretation of the raw data in this case, some general epistemology about how conclusions are reached and judged, etc.

Oh, I'm sure it's relieved to hear that

Please stop the sarcasm or tell me how/why it's productive and non-hostile.

A black-and-white vision seems unnecessary limiting.

it's intentional in order to solve epistemology problems which (I claim) have no other (known) solution. And it's not limiting because things like statistics are used in a secondary role. E.g. you can say "if the following statistical metric gives us 99% or more confidence, i will consider that an adequate solution to my problem". (approaches like that, which use a cutoff amount to determine binary success or failure, are common in science).

First, there is no sharp boundary between a solution and a not-solution.

that depends, as i said, on how the problem is specified.

in the final analysis, when it comes to human action and decision making, for any given issue you decide *yes* to a particular thing and *no* to its rivals. if you hedge, then you're deciding *yes* about that particular hedge.

There is no natural dividing line between a slightly worse model which would be a not-solution and a slightly better model which will be a solution.

depends on the problem domain. e.g. in school sometimes you need an 87 on the test to pass the class, and an 86 will result in failing. so a slightly better test performance can cross a large dividing line. breakpoints like this come up all over the place, e.g. with faster casting speed in diablo 2 (when you hit 37% faster casting speed the casting animation drops by 1 frame. it doesn't drop another frame until 55%. so gear sets totally 40% and 45% FCR are actually equal. (not the actual numbers.)).

Moreover, since these different models are making trade-offs, you can criticise these trade-offs, but generally speaking it's difficult to say that this one is outright wrong and that one is clearly right. There's a reason they're called trade-offs.

it may be difficult, but nevertheless you have to make a decision. the decision should itself be judged in a binary way and be non-refuted – you don't have a criticism of making that particular decision.

i've addressed this stuff at great length. <https://yesornophilosophy.com/argument>

By the way, what about optimization problems? The goal is to maximize Y by manipulating X. There is no threshold, you want Y to be as large as possible. What's the criterion for success?

then do whatever maximizes it. anything with a lower score would be refuted (a mistake to do) since there's an option which gets a higher score. since the problem is to do the thing

with the *best* score (implicitly limited to only options you know of after allocating some amount of resources to looking for better options), second best fails to address that problem.

more typically you don't want to maximize a single factor. i go into this at length in my yes or no philosophy.

[continue this thread »](#)

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

curi comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: ai

Comments (117)

Sort By: Best

You are viewing a single comment's thread. [Show more comments above](#).

curi 09 November 2017 01:58:13AM 0 points

That's a curious definition of a "mistake". It's very... instrumental and local.

This is not local – I specified context matters (whether the context is stated as part of the problem, or specified separately, is merely a matter of terminology.)

You can't determine whether a particular sentence is a correct or incorrect answer without knowing the context – e.g. what is it supposed to answer? The same statement can be a correct answer to one issue and an incorrect

alanforr

Karma Score

17

25



Create new article

Submit a new link

Add new meetup

Preferences

Log out

 Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

answer to a different issue. If you don't like this, you can build the problem and the context into the statement itself, and then evaluate it in isolation.

I'm guessing the reason you consider my view on mistakes "instrumental" is because I think one has to look at the purpose of an idea instead of just the raw data. It's because I add a philosophy layer where you don't. So your alternative to "instrumental" is to say something like "mistakes are when ideas fail to correspond to empirical reality" – and to ignore non-empirical issues, interpretation issues, and that answers to questions need to correspond *to the question* which could e.g. be about a hypothetical scenario. To the extent that questions, goals, human problems, etc, are *part of reality* then, sure, this is all about reality. But I'm guessing we can both agree that's a difference of perspective.

And yet your examples -- e.g. an internal logical inconsistency -- don't seem to require a problem to demonstrate that an idea is broken.

Self-contradictory ideas are broken for *many* problems. In general, we try to criticize an idea as a solution to a range of problems, not a single one. Those criticisms are more interesting. If your criticism is too narrow, it won't work on a slight variant of the idea. You normally want to criticize all the variants sharing a particular theme.

Self-contradictory ideas can (as far as we know) only be correct solutions to some specific types of problems, like for use in parody or as a discussion example.

But even when the real world is relevant, you still have to make an argument saying how to use some evidence in the intellectual debate.

Why is that?

Because facts are not self-explanatory. Any set of facts is open to many interpretations. (Not equally correct interpretations or anything like that, merely logically possible interpretations. So you have to talk about your interpretation, unless the other person can guess it. And you have to talk about how your interpretation of the evidence fits into the debate – e.g. that it contradicts a particular claim – though, again, in simple cases other people may guess that without you saying it.)

That's an interesting claim. An intellectual debate is what's happening inside your head. You are saying that it's primary compared to the objective reality outside of your head. Am I understanding you correctly?

You may prefer to think of it as the philosophy issues are always prior to the other issues. E.g. the role of a particular piece of evidence in reaching some conclusion is governed by ideas and methodology about the role of evidence in general, an interpretation of the raw data in this case, some general epistemology about how conclusions are reached and judged, etc.

Oh, I'm sure it's relieved to hear that

0 points

Thanks, it was not clear to me that it
by [turchin](#) on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by [J_Thomas_Moros](#) on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by [entirelyuseless](#) on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by [turchin](#) on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by [turchin](#) | 0v (4c)

**Fables grow around missed natural
experiments**
by [MaryCh](#) | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by [curi](#) | 1v (41c)

[Link] The Little Dragon is Dead
by [SquirrelInHell](#) | 0v (1c)

[Link] AGI
by [curi](#) | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by [mirefek](#) | 2v (8c)

Please stop the sarcasm or tell me how/why it's productive and non-hostile.

A black-and-white vision seems unnecessary limiting.

it's intentional in order to solve epistemology problems which (I claim) have no other (known) solution. And it's not limiting because things like statistics are used in a secondary role. E.g. you can say "if the following statistical metric gives us 99% or more confidence, i will consider that an adequate solution to my problem". (approaches like that, which use a cutoff amount to determine binary success or failure, are common in science).

First, there is no sharp boundary between a solution and a not-solution.

that depends, as i said, on how the problem is specified.

in the final analysis, when it comes to human action and decision making, for any given issue you decide *yes* to a particular thing and *no* to its rivals. if you hedge, then you're deciding *yes* about that particular hedge.

There is no natural dividing line between a slightly worse model which would be a not-solution and a slightly better model which will be a solution.

depends on the problem domain. e.g. in school sometimes you need an 87 on the test to pass the class, and an 86 will result in failing. so a slightly better test performance can cross a large dividing line. breakpoints like this come up all over the place, e.g. with faster casting speed in diablo 2 (when you hit 37% faster casting speed the casting animation drops by 1 frame. it doesn't drop another frame until 55%. so gear sets totally 40% and 45% FCR are actually equal. (not the actual numbers.)).

Moreover, since these different models are making trade-offs, you can criticise these trade-offs, but generally speaking it's difficult to say that this one is outright wrong and that one is clearly right. There's a reason they're called trade-offs.

it may be difficult, but nevertheless you have to make a decision. the decision should itself be judged in a binary way and be non-refuted – you don't have a criticism of making that particular decision.

i've addressed this stuff at great length. <https://yesornophilosophy.com/argument>

By the way, what about optimization problems? The goal is to maximize Y by manipulating X. There is no threshold, you want Y to be as large as possible. What's the criterion for success?

then do whatever maximizes it. anything with a lower score would be refuted (a mistake to do) since there's an option which gets a higher score. since the problem is to do the thing with the *best* score (implicitly limited to only options you know of after allocating some amount of resources to looking for better options), second best fails to address that problem.

[Link] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians

by curi | 1v (0c)

[Link] Intercellular competition and the inevitability of multicellular aging

by Gunnar_Zarncke | 1v (3c)

Announcing the AI Alignment Prize

by cousin_it | 6v (2c)

Problems as dragons and papercuts

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

more typically you don't want to maximize a single factor. i go into this at length in my yes or no philosophy.

Lumifer 09 November 2017 03:51:35AM * 0 points [-]

| one has to look at the purpose of an idea instead of just the raw data

Oh, I agree. It's just that you were very insistent about drawing the line between unfalsifiable philosophy and other empirically-falsifiable stuff and here you're coming back into the real-life problems realm where things are definitely testable and falsifiable. I'm all for it, but there are consequences.

| So you have to talk about your interpretation, unless the other person can guess it.

Sure, but that's not an intellectual debate. If someone asks how to start a fire and I explain how you arrange kindling, get a flint and a steel, etc. there is no debate -- I'm just transferring information.

| the philosophy issues are always prior to the other issues

Not necessarily. If you put your hand into a fire, you will get a burn -- that's easy to learn (and small kids learn it fast). Which philosophy issues are prior to that learning?

| Please stop the sarcasm

No can do. But tell you what, the fewer silly things you say, the less often you will encounter overt sarcasm :-)

| in order to solve epistemology problems

Which problems you can't solve otherwise?

| for any given issue you decide yes to a particular thing and no to its rivals

There are lot of issues with continuous (real number) decisions. Let's say you're deciding how much money to put into your retirement fund this year and the reasonable range is between \$10K and \$20K. You are not going to treat \$14,999 and \$15,000 as separate solutions, are you?

| breakpoints like this come up all over the place

Sure they do, but **not always**. And your approach requires them.

| the decision should itself by judged in a binary way and be non-refuted

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

I still don't see the need for these rather severe limitations. You want to deal with reality as if it consists of discrete, well-delineated chunks and, well, it just doesn't. I understand that you can impose thresholds and breakpoints any time you wish, but they are artifacts and if your method requires them, it's a drawback.

then do whatever maximizes it

Yes, but you typically have an explore-or-exploit problem. You need to spend resources to look for a better optimum, at each point in time you have some probability of improving your maximum, but there are costs and they grow. At which point do you stop expending resources to look for a better solution?

curi 09 November 2017 04:33:48AM 0 points



It's just that you were very insistent about drawing the line between unfalsifiable philosophy and other empirically-falsifiable stuff

if you have an empirical argument to make, that's fine. but i don't think i'm required to provide evidence for my philosophical claims. (btw i criticize the standard burden of proof idea in Yes or No Philosophy. in short, if you can't criticize an idea then it's non-refuted and demanding some sort of burden of proof is not a criticism since lack of proof doesn't prevent an idea from solving a problem.)

in order to solve epistemology problems

Which problems you can't solve otherwise?

the problem of induction. problems about how to evaluate arguments (how do you score the strength of an argument? and what difference does it really make if one scores higher than another? either something points out why a solution doesn't work or it doesn't. unless you specifically try to specify non-binary problems. but that doesn't really work. you can specify a set of solutions are all equal. ok then either pick any one of them if you're satisfied, or else solve some other more precise problem that differentiates. you can also specify that higher scoring solutions on some metric are better, but then you just pick the highest scoring one, so you get a single solution or maybe a tie again. and whether you've chosen a correct solution given the problem specification, or not, is binary.) and various problems about how you decide what metrics to use (the solution to that being binary arguments about what metrics to use – or in many cases don't use a metric. metrics are overrated but useful sometimes.)

Yes, but you typically have an explore-or-exploit problem. You need to spend resources to look for a better optimum, at each point in time you have some probability of improving your maximum, but there are costs and they grow. At which point do you stop expending resources to look for a better solution?

Yes so then you guess what to do and criticize your guesses. Or, if you wish, define a metric with positive points for a higher score and negative points for resources spent (after you guess-and-criticize to figure out how to put the positive score and all the different types of resources into the same units) and then guess how to maximize *that* (e.g. define a metric about resources allocated to getting a higher score on the first metric, spend that much resources, and then use the highest scoring solution).

multi-factor metrics don't work as well as people think, but are ok sometimes (but you have to make a binary judgement about whether to use a particular metric for a particular situation, or not – so the binary judgement is prior and governs the use of the metric). here's a good article about issues with them:

<https://www.newyorker.com/magazine/2011/02/14/the-order-of-things>

scoring systems are overrated but are allowable in binary epistemology given that their use is governed by binary judgements (should I proceed by doing the thing that scores the highest on this metric? make critical arguments about that and make a binary judgement. so the binary judgement is prior but then things like metrics and statistics are allowable as secondary things which are sometimes quite useful.)

| You are not going to treat \$14,999 and \$15,000 as separate solutions, are you?

depends how precise the problem or context says to be. (or bigger picture, it depends how precise is worth the resources to be – which you should either specify in the problem or consider part of the context.)

if you don't care about single dollar level of precision (cuz you want to save resources like effort to deal with details), just e.g. specify in the problem that you only care about increments of \$500 or that (to save problem solving resources like time) you just want to use the first acceptable solution you come up with that you determine meet some standards of good enough (these are no longer strictly single variable maximization problems).

| breakpoints like this come up all over the place

| Sure they do, but **not always**. And your approach requires them.

they aren't required, you can specify the problem however you want (subject to criticism) so you it makes clear what is a solution or not (or a set of tied solutions you're indifferent btwn which you can then tiebreak arbitrarily if you have no criticism of doing it arbitrarily).

if the problem specifies that some solutions are better than others (not my preferred way to specify problems – i think it's epistemologically misleading), then when you act you should pick one of the solutions in the highest tier you have a solution in, and reject the others. whether this method (pick a highest tier solution) is correct, and whether you've used it in this case, are both binary issues open to criticism.

At which point do you stop expending resources to look for a better solution?

when you guess it's best to stop and your guess is non-refuted and the guess to continue looking is refuted. (you may, if you want to, define some stopping metric and make a subject-to-criticism binary yes-or-no judgement about whether to use that stopping metric.)

the philosophy issues are always prior to the other issues

Not necessarily. If you put your hand into a fire, you will get a burn -- that's easy to learn (and small kids learn it fast). Which philosophy issues are prior to that learning?

i think small kids do guesses and criticism, and use methods of learning (what I would call philosophical methods), even if they can't state those methods in English. i also think ppl who have never studied philosophy use philosophy methods, which they picked up from their culture here and there, even if they don't consciously understand themselves or know the names of the things they're doing. and to the extent ppl learn, i think it's guesses and criticism in some form, since that's the only known method of learning (at a low level, it's evolution – the only known solution to the problem of where the appearance of design comes from – saying it comes from "intelligence" is like attributing it to God or an intelligent designer – it doesn't tell you how god/intelligence does it. my answer to that is, at a low level, evolution. layers of abstraction are built on top of that so it looks more varied at a higher level.).

Lumifer 09 November 2017 05:29:42PM 0 points



i don't think i'm required to provide evidence for my philosophical claims

It depends on what do you want to do with them. If all you want to do is keep them on a shelf and once in a while take them out, dust them, and admire them, then no, you don't. On the other hand, if you want to persuade someone to change their mind, evidence might be useful. And if you want other people to take action based on your claims', ahem, implications, evidence might even be necessary.

the problem of induction. problems about how to evaluate arguments

It seems that the root of these problems is your insistence that truth is a binary category. If you are forced to operate with single-bit values and have to convert every continuous function into a step one, well, sure you will have problems.

The thread seem to be losing shape, so let's do a bit of a summary. As far as I can see, the core differences between us are:

- You think truth (and arguments) are binary, I think both have continuous values;

- You think intellectual debates are primary and empirical testing is secondary, I think the reverse;

Looks reasonable to you?

curi 09 November 2017 06:24:50PM * 0 points 

the two things you listed are ok with me. i'd add induction vs guesses-and-criticism/evolution to the list of disagreements.

do you think there's a *clear, decisive mistake* in something i'm saying?

can you specify how you think induction works? as a fully defined, step-by-step process i can do today?

though what i'd prefer most is replies to the things i said in my previous message.

Lumifer 09 November 2017 07:03:43PM 0 points 

| do you think there's a clear, decisive mistake in something i'm saying?

I would probably classify it as suboptimal. It's not a "clear, decisive mistake" to see only black and white -- but it limits you.

| can you specify how you think induction works?

In the usual way: additional data points increase the probability of the hypothesis being correct, however their influence tends to rapidly decline to zero and they can't lift the probability over the asymptote (which is usually less than 1). Induction doesn't *prove* anything, but then in my system nothing *proves* anything.

What you said in the previous message is messy and doesn't seem to be terribly impactful. Talking about how you can define a loss function or how you can convert scores to a yes/no metric is secondary and tertiary to the core disagreements we have.

curi 09 November 2017 07:09:20PM * 0 points 

| In the usual way: additional data points increase the probability of the hypothesis being correct,

the probability of which hypotheses being correct, how much? how do you differentiate between hypotheses which do not contradict any of the data?

Lumifer 09 November 2017 07:18:49PM * 0 points [-]

| the probability of which hypotheses being correct, how much?

For a given problem I would have a set of hypotheses under consideration. A new data point might kill some of them (in the Popperian fashion) or might spawn new ones. Those which survive -- all of them -- gain some probability. How much, it depends. No simple universal rule.

| how do you differentiate hypotheses which do not contradict any of the data?

For which purpose and in which context? I might not need to differentiate them.

Occam's razor is a common heuristic, though, of course, it is NOT a guide to whether a particular theory is correct or not.

curi 09 November 2017 07:20:43PM * 0 points [-]

Do all the non-contradicted-by-evidence ideas gain *equal* probability (so they are always tied and i don't see the point of the "probabilities"), or differential probability?

EDIT: I'm guessing your answer is you *start* them with different amounts of probability. after that they gain different amounts accordingly (e.g. the one at 90% gains less from the same evidence than the one at 10%). but the ordering (by amount of probability) always stays the same as how it started, apart from when something is dropped to 0% by contradicting evidence. is that it? or do you have a way (which is part of induction, not critical argument?) to say "evidence X neither contradicts ideas Y nor Z, but fits Y better than Z"?

Lumifer 09 November 2017 08:00:28PM 0 points [-]

Different hypotheses (= models) can gain different amounts of probability. They can start with different amounts of probability, too, of course.

| to say "evidence X neither contradicts ideas Y nor Z, but fits Y better than Z"?

Of course. That's basically how all statistics work.

Say, if I have two hypotheses that the true value of X is either 5 or 10, but I can only get noisy estimates, a measurement of 8.7 will add more probability to the "10" hypothesis than to the "5" hypothesis.

[continue this thread »](#)

LessWrong DISCUSSION

MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

Lumifer comments on Questions about AGI's Importance - Less Wrong Discussion

0 curi 31 October 2017 08:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(117\)](#)

Tags: ai

Comments (117)

Sort By: Best

You are viewing a single comment's thread. [Show more comments above](#).

Lumifer 09 November 2017 08:00:28PM 0 points

Different hypotheses (= models) can gain different amounts of probability. They can start with different amounts of probability, too, of course.

to say "evidence X neither contradicts ideas Y nor Z, but fits Y better than Z"?

Of course. That's basically how all statistics work.

Say, if I have two hypotheses that the true value of X is either 5 or 10, but I can only get noisy estimates, a measurement of 8.7 will add more probability to the "10" hypothesis than to the "5" hypothesis.

alanforr

Karma Score

17

25

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Additional info: - <http://curi.us/2067-> by curi on Less Wrong Lacks Representatives and Paths Forward |

curi 09 November 2017 08:03:34PM 0 points [-]

what do you do about ideas which make identical predictions?

gjm 09 November 2017 08:36:14PM 0 points [-]

They get identical probabilities -- if their prior probabilities were equal.

If (as is the general practice around these parts) you give a markedly bigger prior probability to simpler hypotheses, then you will strongly prefer the simpler idea. (Here "simpler" means something like "when turned into a completely explicit computer program, has shorter source code". Of course your choice of language matters a bit, but unless you make wilfully perverse choices this will seldom be what decides which idea is simpler.)

In so far as the world turns out to be made of simply-behaving things with complex emergent behaviours, a preference for simplicity will favour ideas expressed in terms of those simply-behaving things (or perhaps other things essentially equivalent to them) and therefore more-explanatory ideas. (It is at least partly the fact that the world seems so far to *be* made of simply-behaving things with complex emergent behaviours that makes explanations so valuable.)

Lumifer 09 November 2017 08:11:03PM 0 points [-]

I don't need to distinguish between them, then.

curi 09 November 2017 08:15:54PM 0 points [-]

so you don't deal with explanations, period?

Lumifer 09 November 2017 08:38:04PM 0 points [-]

I do, but more or less only to the extent that they will make potential different predictions. If two models are in principle incapable of making different predictions, I don't see why should I care.

curi 09 November 2017 08:41:14PM 0 points [-]

0 points

Thanks, it was not clear to me that it
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

Not going to sign up with some
by **J_Thomas_Moros** on Military AI as
a Convergent Goal of Self-Improving
AI | 1 point

People are weakly motivated
by **entirelyuseless** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

This is our accepted chapter in the
by **turchin** on Military AI as a
Convergent Goal of Self-Improving AI |
0 points

RECENT POSTS

**[Link] Military AI as a Convergent
Goal of Self-Improving AI**
by **turchin** | 0v (4c)

**Fables grow around missed natural
experiments**
by **MaryCh** | 1v (0c)

**Less Wrong Lacks Representatives
and Paths Forward**
by **curi** | 1v (41c)

[Link] The Little Dragon is Dead
by **SquirrelInHell** | 0v (1c)

[Link] AGI
by **curi** | 0v (3c)

**[Link] Kialo -- an online discussion
platform that attempts to support
reasonable debates**
by **mirefek** | 2v (8c)

so e.g. you don't care if trees exist or not? you think people should stop thinking in terms of trees and stick to empirical predictions only, dropping any kind of non-empirical modeling like the concept of a tree?

Lumifer 09 November 2017 08:48:15PM 0 points [-]

you don't care if trees exist or not?

I don't understand what this means.

any kind of non-empirical modeling like the concept of a tree?

The concept of a tree seems pretty empirical to me.

curi 09 November 2017 08:56:07PM 0 points [-]

there are infinitely many theories which say trees don't exist but make identical predictions to the standard view involving trees existing.

trees are not an observation, they are a conceptual interpretation. observations are things like the frequencies of photons at times and locations.

Lumifer 09 November 2017 09:14:32PM 0 points [-]

there are infinitely many theories which say trees don't exist but make identical predictions

Isn't it convenient that I don't have to care about these infinitely many theories?

Since there is an infinity of them, I bet you can't marshal critical arguments against ALL of them :-P

trees are not an observation

I think you're getting confused between actual trees and the abstract concept of a tree.

observations are things like the frequencies of photons at times and locations.

[Link] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians

by curi | 1v (0c)

[Link] Intercellular competition and the inevitability of multicellular aging

by Gunnar_Zarncke | 1v (3c)

Announcing the AI Alignment Prize

by cousin_it | 6v (2c)

Problems as dragons and papercuts

by Elo | 1v (1c)

LATEST OPEN THREAD

By the way, is there an explanation

by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

I learned something new. As a

by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gvern (25)

alanforr (25)

Lumifer (23)

Elo (22)

gjm (19)

I don't think so. Human brains do not process sensory input in terms of " frequencies of photons at times and locations".

curi 09 November 2017 09:16:13PM * 0 points 

Isn't it convenient that I don't have to care about these infinitely many theories?

why not?

Since there is an infinity of them, I bet you can't marshal critical arguments against ALL of them :-P

you can criticize categories, e.g. all ideas with feature X.

I think you're getting confused between actual trees and the abstract concept of a tree.

i don't think so. you can't observe entities. you have to interpret what entities there are (or not – as you advocated by saying only prediction matters)

[continue this thread »](#)

IlyaShpitser (16)
curi (14)
Dagon (11)
NancyLebovitz (6)
satt (6)
cousin_it (6)
Habryka (6)
ChristianKI (6)
SquirrelInHell (5)
root (5)

RECENT KARMA AWARDS

[ABOUT LESS WRONG](#) | [REPORT ISSUES](#) | [PRIVACY & TERMS](#)

Powered by [Reddit](#) 

Less Wrong Lacks Representatives and Paths Forward

1 curi 08 November 2017 07:00PM

In my understanding, there's no one who speaks for LW, as its representative, and is *responsible* for addressing questions and criticisms. LW, as a school of thought, has no agents, no representatives – or at least none who are open to discussion.

The people I've found interested in discussion on the website and slack have diverse views which disagree with LW on various points. None claim LW is true. They all admit it has some weaknesses, some unanswered criticisms. They have their own personal views which aren't written down, and which they don't claim to be correct anyway.

This is problematic. Suppose I wrote some criticisms of the sequences, or some Bayesian book. Who will answer me? Who will fix the mistakes I point out, or canonically address my criticisms with counter-arguments? No one. This makes it hard to learn LW's ideas in addition to making it hard to improve them.

My school of thought (Fallible Ideas – FI – <https://fallibleideas.com>) has representatives and claims to be correct as far as is known (like LW, it's fallibilist, so of course we may discover flaws and improve it in the future). It claims to be the best current knowledge, which is currently non-refuted, and has refutations of its rivals. There are other schools of thought which say the same thing – they actually think they're right and have people who will address challenges. But LW just has individuals who individually chat about whatever interests them without there being any organized school of thought to engage with. No one is responsible for defining an LW school of thought and dealing with intellectual challenges.

alanforr

Karma Score

17

22



[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12
January 2019 02:00PM

LW Netherlands: 03 September 2122
01:45PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.

Less Wrong Study Hall

RECENT COMMENTS

Sorry, did you say weird/esoteric
by [IlyaShpitser](#) on LW 2.0 Open Beta
Live | 1 point

So how is progress to be made? Suppose LW, vaguely defined as it may be, is mistaken on some major points. E.g. Karl Popper refuted induction. How will LW find out about its mistake and change? FI has a forum where its representatives take responsibility for seeing challenges addressed, and have done so continuously for over 20 years (as some representatives stopped being available, others stepped up).

Which challenges are addressed? *All of them*. You can't just ignore a challenge because it could be correct. If you misjudge something and then ignore it, you will stay wrong. Silence doesn't facilitate error correction. For information on this methodology, which I call Paths Forward, see: <https://curi.us/1898-paths-forward-short-summary> BTW if you want to take this challenge seriously, you'll need to click the link; I don't repeat all of it. In general, having much knowledge is incompatible with saying all of it (even on one topic) upfront in forum posts without using references.

My criticism of LW as a whole is that it lacks Paths Forward (and lacks some alternative of its own to fulfill the same purpose). In that context, my criticisms regarding specific points don't really matter (or aren't yet ready to be discussed) because there's no mechanism for them to be rationally resolved.

One thing FI has done, which is part of Paths Forward, is it has surveyed and addressed other schools of thought. LW hasn't done this comparably – LW has no answer to Critical Rationalism (CR). People who chat at LW have individually made some non-canonical arguments on the matter that LW doesn't take responsibility for (and which often involve conceding LW is wrong on some points). And they have told me that CR has critics – true. But which criticism(s) of CR does LW claim are correct and take responsibility for the correctness of? (Taking responsibility for something involves doing some major rethinking if it's refuted – addressing criticism of it and fixing your beliefs if you can't. Which criticisms of CR would LW be shocked to discover are mistaken, and then be eager to reevaluate the whole matter?) There is no answer to this, and there's no way for it to be answered because LW has no representatives who can speak for it and who are participating in discussion and who consider it their responsibility to see that issues like this are addressed. CR is well known, relevant, and makes some clear LW-contradicting claims like that induction doesn't work, so if LW had representatives surveying and responding to rival ideas, they would have addressed CR.

BTW I'm not asking for all this stuff to be perfectly organized. I'm just asking for it to exist at all so that progress can be made.

Yes, the article I was looking for.
by [fblogin](#) on Double Crux — A Strategy for Resolving Disagreement | 0 points

Sorry, but it is. Simple test: open a
by [elharo](#) on LW 2.0 Open Beta Live | 0 points

[Lucifer's version]
by [roland](#) on The "Outside the Box" Box | 0 points

Interesting discussion. Does
by [kimberchoi](#) on Actions and Words: Akrasia and the Fruit of Self-Knowledge | 0 points

RECENT POSTS

LW 2.0 Open Beta Live
by [Vaniver](#) | 22v (36c)

In support of yak shaving part 2
by [Elo](#) | 4v (7c)

Project Hufflepuff: Planting the Flag
by [Raemon](#) | 41v (106c)

European Community Weekend 2017
by [DreamFlasher](#) | 16v (16c)

"Flinching away from truth" is often about *protecting* the epistemology
by [AnnaSalamon](#) | 73v (54c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by [AnnaSalamon](#) | 36v (43c)

Be secretly wrong
by [Benquo](#) | 32v (47c)

CFAR's new focus, and AI Safety
by [AnnaSalamon](#) | 30v (88c)

Anecdotally, I've found substantial opposition to discussing/considering methodology from LW people so far. I think that's a mistake because we use methods when discussing or doing other activities. I've also found substantial resistance to the use of references (including to my own material) – but why should I rewrite a new version of something that's already written? Text is text and should be treated the same whether it was written in the past or today, and whether it was written by someone else or by me (either way, I'm taking responsibility. I think that's something people don't understand and they're used to people throwing references around both vaguely and irresponsibly – but they haven't pointed out any instance where I made that mistake). Ideas should be judged by the idea, not by attributes of the source (reference or non-reference).

The Paths Forward methodology is also what I think individuals should personally do – it works the same for a school of thought or an individual. Figure out what you think is true *and take responsibility for it*. For parts that are already written down, endorse that and take responsibility for it. If you use something to speak for you, then if it's mistaken *you* are mistaken – you need to treat that the same as your own writing being refuted. For stuff that isn't written down adequately by anyone (in your opinion), it's your responsibility to write it (either from scratch or using existing material plus your commentary/improvements). This writing needs to be put in public and exposed to criticism, and the criticism needs to actually get addressed (not silently ignored) so there are good Paths Forward. I hoped to find a person using this method, or interested in it, at LW; so far I haven't. Nor have I found someone who suggested a superior method (or even *any* alternative method to address the same issues) or pointed out a reason Paths Forward doesn't work.

Some people I talked with at LW seem to still be developing as intellectuals. For lots of issues, they just haven't thought about it yet. That's totally understandable. However I was hoping to find some developed thought which could point out any mistakes in FI or change its mind. I'm seeking primarily peer discussion. (If anyone wants to learn from me, btw, they are welcome to come to my forum. It can also be used to criticize FI. <http://fallibleideas.com/discussion-info>) Some people also indicated they thought it'd be too much effort to learn about and address rival ideas like CR. But if no one has done that (so there's no answer to CR they can endorse), then how do they know CR is mistaken? If CR is correct, it's worth the effort to study! If CR is incorrect, someone better write that down in public (so CR people can learn about their errors and reform; and so perhaps they could improve CR to no longer be mistaken or point out errors in the criticism of CR.)

One of the issues related to this dispute is I believe we can always proceed with non-refuted ideas (there is a long answer for how this works, but I don't know how to give a short answer that I expect LW people to understand – especially in the context of the currently-unresolved methodology dispute about Paths Forward). In contrast, LW people typically seem to accept mistakes as just something to put up with, rather than something to try to always fix. So I

Fact Posts: How and Why

by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement

by Duncan_Sabien | 61v (103c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all

by vaultDweller on Rationality Quotes

April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

IlyaShpitser (22)

alanforr (22)

gjm (19)

Elo (17)

Lumifer (16)

curi (15)

cousin_it (7)

gwern (5)

mirefek (5)

OrthernLight (5)

Oscar_Cunningham (5)

Dagon (4)

JenniferRM (4)

morganism (4)

jimrandomh (4)

RECENT KARMA AWARDS

disagree with ignoring some *known* mistakes, whereas LW people seem to take it for granted that they're mistaken in known ways. Part of the point of Paths Forward is not to be mistaken in known ways.

Paths Forward is a methodology for organizing schools of thought, ideas, discussion, etc, to allow for unbounded error correction (as opposed to typical things people do like putting bounds on discussions, with discussion of the bounds themselves being out of bounds). I believe the lack of Paths Forward at LW is preventing the resolution of other issues like about the correctness of induction, the right approach to AGI, and the solution to the fundamental problem of epistemology (how new knowledge can be created).

Article Navigation

[Comments \(85\)](#)

Comments (85)

Sort By: **Best**

Comment

Show help

Lumifer 08 November 2017 09:50:59PM * 5 points



I think I'm the person curi is referring to as "some people" throughout his post. There is a long-ass continuing thread starting about [here](#) in which curi tries to convert me to the Sole True Way of Correct Thinking and isn't making much headway.

So all y'all don't have to worry -- the lazy underdeveloped intellectual is me.

P.S. The tl;dr of OP seems to be:

Take me to your leader!

...

What kind of savages are you that you don't even have a Glorious Leader and a Big Sacred Book?!??

curi 08 November 2017 09:57:55PM 0 points



I think I'm the person curi is referring to as "some people" throughout his post.

You're incorrect. I discussed with various other people, including gjm, on Slack. I also discussed with other people on the forum some, both in the past and recently. Also I don't refer to "some people" throughout the post, only in one paragraph.

I do consider your approach (Lumifer) to have Paths Forward problems. I don't think you are even claiming to be a representative of Less Wrong and to do the various things I say LW lacks. So you're a reasonably typical example, not a counter-example.

gjm 08 November 2017 09:34:43PM 5 points



In my understanding, there's no one who speaks for LW, as its representative, and is *responsible* for addressing questions and criticisms. [...] No one is responsible for defining an LW school of thought and dealing with intellectual challenges.

Correct. There is no Pope of LW, we don't all agree about everything, and no one has any obligation to answer anyone else's objections. That may be inconvenient for some purposes, but that's how it is.

This is also how it is for many other things besides LW. Suppose someone claims that there is something wrong with *science*; there is no Pope of Science any more than there is one of LW, no one responsible for defining how scientists do their thing or answering criticisms. The same goes for all of the following: atheism, Protestantism, conservatism, environmentalism, reductionism, mathematical intuitionism, moral nonrealism, "Big Bang" cosmology. And, in fact, for pretty much everything else.

There *are* movements that have a clearly defined set of doctrines and a clearly defined representative, such that if the representative's positions are refuted then the whole movement is sunk. These movements are generally called cults. Empirically, being such a movement doesn't seem to be conducive to valuable things like good thinking, success in persuading others, or making the movement's members successful outside the movement.

None claim LW is true.

LW is not the *kind of thing* that can be "true". Nor, really, is (say) "the LW philosophy" in so far as there is one -- because a philosophy is a complicated machine with lots of moving parts and almost certainly some things in it aren't quite right.

And, of course, because really there isn't a precisely-defined thing that is "the LW philosophy" or "what LW people believe" or whatever; LW is a community, not a cult, and we don't expect to agree about everything.

You'll probably find plenty of us who will happily endorse some weaker claim, though; something like "most of the ideas found in the Sequences, in Scott's less-blatantly-speculative posts, etc., are pretty good ideas".

Suppose LW, vaguely defined as it may be, is mistaken on some major points. E.g. Karl Popper refuted induction. How will LW find out about its mistake and change?

Well, perhaps we won't. But if we do, it will probably go like this: someone active on LW learns of a problem; posts about it on LW / discusses it in person with other LW people / posts it in some other venue where LW people hang out; it gets discussed, and (perhaps after a few iterations, since "you're doing it all wrong" tends to be a hard lesson to learn even when true) other LW people are gradually convinced; then either we start doing things differently, or (if not a large enough fraction are convinced) some people go off and form their own (sub)community where things are done differently.

Which challenges are addressed? *All of them*. You can't just ignore a challenge because it could be correct. If you misjudge something and then ignore it, you will stay wrong.

There is no procedure that guarantees not staying wrong. Always addressing every challenge is one way to try to avoid that, but it does not guarantee not staying wrong and the fact that *not* always addressing every challenge doesn't guarantee not staying wrong is not a strong argument against it.

CR is well known, relevant, and makes some clear LW-contradicting claims like that induction doesn't work, so if LW had representatives surveying and responding to rival ideas, they would have addressed CR.

But LW does not have "representatives" in the sense you describe, and isn't likely to start having them. However, LW *has* addressed CR in a weaker sense: e.g., some time ago you came here and tried to persuade us to abandon LW-style probabilistic inference in favour of your version of critical rationalism; the community took a look at your ideas, was not impressed, and downvoted your posts to hell.

Of course that may have been a mistake. (Again: there is no guarantee of not making mistakes. Not ever.) But that's what addressing CR looks like for a community like LW: someone proposes it, making the best case they can, then people take a look, argue about it as seems appropriate, and see what they think.

Some people I talked with at LW seem to still be developing as intellectuals.

You say that as if "still developing as intellectuals" implies some sort of immaturity. I hope to continue developing as an intellectual until I die, though empirically it seems most people don't manage to do that :-).

Some people also indicated they thought it'd be too much effort to learn about and address rival ideas like CR.

It is possible that I haven't seen everything you're referring to here. But the things I *have* seen, on LW and on the LW Slack, that I think you're referring to are not accurately described by the words above. What I've seen people say is much more like "I don't think *your particular ideas*, from what I have seen about them, merit *the particular effort* you are saying I should put in". Or, more generally, "It is not feasible to give a deep investigation to every single rival idea that comes along."

These are not the same proposition as "it would be too much effort to learn about and address rival ideas": they are both compatible with the idea that some ideas might justify the deeper investigation that most of the LW community evidently hasn't yet been persuaded by your arguments to give to CR.

| I believe the lack of Paths Forward at LW is preventing the resolution of other issues [...]

You are welcome to believe that. I don't think you should find it surprising that we are largely unconvinced. Not least because the only actual argument you've offered for this, other than saying "go to my website", is the claim that it's always a mistake to leave anything at all unaddressed -- which I suspect seems to others here, as it seems to me, an obvious mistake in its own right.

RealJustinCEO 08 November 2017 09:55:56PM 1 point



| In my understanding, there's no one who speaks for LW, as its representative, and is responsible for addressing questions and criticisms. [...] No one is responsible for defining an LW school of thought and dealing with intellectual challenges.

| Correct. There is no Pope of LW, we don't all agree about everything, and no one has any obligation to answer anyone else's objections.

Asking for someone who thinks some set of ideas is consistent and true, and will address questions about those ideas thoroughly, is not asking for a Pope. It's more like asking for someone who's more than a casual fan.

| That may be inconvenient for some purposes, but that's how it is.

One purpose that the lack of a serious LW advocate is "inconvenient" for is truth-seeking - a rather important case!

gjm 09 November 2017 12:06:48AM 2 points



| Asking for someone who [...] is not asking for a Pope.

No. But curi was asking for more than that. E.g., he wants someone who "speaks for LW". He wants them to do it "as [LW's] representative". He wants them to address arguments against LWish ideas "canonically". He wants someone "responsible for defining an LW school of thought". And so forth.

And, as I said above, this is just not how most communities or schools of thought work, nor should it be, nor I think *could* it be. Except for ones where in order to claim any sort of affiliation you are required to sign up to a particular body of doctrine. That mostly means religions, political parties, etc. And (again, as I said above) groups of that sort don't have an encouraging record of successfully distinguishing truth from error; I don't think we should be emulating them.

curi 09 November 2017 12:11:09AM 0 points [-]

if someone spoke for something smaller than LW, e.g. Bayesian Epistemology, that'd be fine. CR and Objectivism, for example, can be questioned and have people who will answer (unlike science itself).

and if someone wanted to take responsibility for gjm-LW or lumifer-LW or some other body of ideas which is theirs alone, that'd be fine too. but people aren't doing this as a group *or* individually!

ChristianKI 11 November 2017 10:14:26PM 0 points [-]

The fact that objectivism has cultists who want to defend the objectivist way isn't a quality that's worthy of emulation. If CR is copying the same group think structures that's also no argument in favor of it either.

curi 11 November 2017 10:33:05PM * 0 points [-]

I like Ayn Rand's writing, not whatever you think is a "cult". See e.g. <http://curi.us/1930-harry-binswanger-refuses-to-think>

If you have an argument about Ayn Rand's ideas, that would be important.

Regardless, you can get correct answers to tons of common questions about Objectivism at a variety of places online (including both pro-ARI and anti-ARI places). That's good. And Binswanger, linked negatively above, engaged with Popperian criticism more than anyone at LW has. He also has combined seriously writing down ideas with discussing ideas, whereas LW people seem to only do much of one or the other, which I think is a big problem.

ChristianKI 11 November 2017 10:55:48PM 0 points [-]

Speaking for "objectivism" instead of someone personal opinions implies structures that get people think alike in a cultish way.

Lumifer 09 November 2017 01:34:25AM 0 points [-]

You can, of course, go and bother Eliezer. I doubt he would be inclined to listen to you, though.

curi 09 November 2017 02:25:48AM 0 points [-]

Eliezer has already indicated [1] he'd prefer to *take administrative action to prevent discussion* than speak to the issues. No Paths Forward there!

[1] http://lesswrong.com/lw/56m/theconjunctionfallacydoesnot_exist/3wf5

gjm 09 November 2017 11:50:30AM 3 points [-]

That's ... not a very accurate way of describing what happened. Not because there's literally no way to understand it that makes it factually correct, but because it gives entirely the wrong impression.

Here's a more complete description of what happened.

curi came here in early April 2011 (well, he actually first appeared earlier, but before then he made a total of three comments ever) and posted five lengthy top-level posts in five days. They were increasingly badly received by the community, getting scores of -1,-1,-1,-22,-38. The last one was entitled "The conjunction fallacy does not exist" and what it attempted to refute was a *completely wrong statement of what the c.f. is about*, namely the claim (which no one believes) that "people attribute higher probability to X&Y than to Y" *for all X and Y*.

As this was happening, more and more of the comments on curi's posts were along the general lines of [this one](#) saying, in essence: This is not productive, you are just repeating the same wrong things without listening to criticism, so please stop.

It was suggested that there was some reason to think curi was using sockpuppets to undo others' downvotes and keep enough karma to carry on posting.

And *then*, in that context, curi's fifth post -- which attempted to refute the conjunction fallacy but which completely misunderstood what the conjunction fallacy is, and which was sitting on -38 points -- was removed.

Now, *maybe* that's because Eliezer was afraid of curi's ideas and wanted to close down discussion or something of the sort. But a *more plausible* explanation is that he thought further discussion was likely to be a waste of time for the same reason as several commenters.

I don't think removing the post was a good decision, and generally I think Eliezer's moderation has been too heavy-handed on multiple occasions. But I don't think the *kind of explanation* curi is offering for this is at all likely to be correct.

On the other hand, if curi is merely saying that Eliezer is unlikely to be interested if curi contacts him and asks for a debate on Bayes versus CR, then I think he's clearly right about that.

Lumifer 09 November 2017 03:19:04AM 0 points [-]

Yep, sounds like Eliezer. No surprises.

gjm 09 November 2017 12:37:20AM 0 points [-]

Well, both Lumifer and I have (mostly in different venues) been answering a lot of questions and criticisms you've posed. But no, I don't think either of us feels "responsibility" in the specific (and, I think, entirely non-standard) sense you're using here, where to "take responsibility" for a set of ideas is to incur a limitless obligation to answer any and all questions and criticisms made of those ideas.

curi 09 November 2017 12:39:54AM 0 points [-]

there are methods for doing Paths Forward with limited resource use. you just don't want to learn/discuss/use them.

gjm 09 November 2017 01:19:32PM 3 points [-]

The total of what your "paths forward" page says about limited resources: (1) instead of writing your own answers to every criticism, you can point critics to already-written things that address their criticisms; (2) if you have a suitable forum with like-thinking other people there, they may address the criticisms for you.

Perhaps it seems to you that these make it reasonable to have a policy of addressing every criticism and question despite limited resources. It doesn't seem so to me.

I have read your document, I am not convinced by your arguments that we *should* attempt to address every single criticism and question, I am not convinced by your arguments that we *can* realistically do so, and I think the main practical effects of embracing your principles on this point would be (1) to favour obsessive cranks who have nothing else to do with their time than argue about their pet theories, (2) to encourage obsessive-crank-like behaviour, and (3) to make those who embrace them spend more time arguing on the internet. I can't speak for others, but I don't want to give advantages to obsessive cranks, I don't want to become more obsessive and cranky myself, and I think it much more likely that I spend too much time arguing on the internet rather than too little.

I see nothing to suggest that further investigation of "paths forward" is likely to be a productive use of my time.

So: no, I don't want to spend more time learning, discussing, or using "paths forward". I think it would be a suboptimal way to use that time.

Lumifer 09 November 2017 01:12:18AM 1 point [-]

| you just don't want to learn/discuss/use them

By Jove, I think you got it!

:-D

Lumifer 08 November 2017 10:23:47PM 1 point [-]

And who are you, a freshly-minted account with strong opinions?

gjm 08 November 2017 11:41:21PM 2 points [-]

Both here and on the LW slack, "Justin CEO" turned up at about the same time as curi and has done more or less nothing other than agreeing with curi and disagreeing with people who are disagreeing with him.

This is perfectly consistent with Justin not being a sockpuppet of curi, of course.

RealJustinCEO 08 November 2017 11:04:15PM 0 points [-]

BTW I don't even think anything I said was particularly opinionated (by my standards).

And the context makes it funny.

Saying -- *on the LW forum* -- that it'd be good for LW to have a strong advocate? Ooooh, controversial!

Lumifer 09 November 2017 01:33:20AM 0 points [-]

No, not particularly opinionated. But it's an interesting place and time you chose for diving into LW.

curi 09 November 2017 02:08:41AM * 0 points [-]

is it really that interesting? i posted a rough draft of "Less Wrong Lacks Representatives and Paths Forward" to the FI forum for comment. i routinely post links and comments about discussions i'm having to FI. is this surprising?

Lumifer 09 November 2017 03:19:41AM 0 points [-]

Not you -- RealJustinCEO.

curi 09 November 2017 04:00:36AM 0 points [-]

that wasn't clear enough? he is a member of the FI forum. he followed one of my links.

RealJustinCEO 08 November 2017 11:01:25PM 0 points [-]

And who are you, a freshly-minted account with strong opinions?

Yep, I'm a freshly-minted account with strong opinions.

curi 08 November 2017 09:47:53PM 0 points [-]

LW has addressed CR in a weaker sense: e.g., some time ago you came here and tried to persuade us to abandon LW-style probabilistic inference in favour of your version of critical rationalism; the community took a look at your ideas, was not impressed, and downvoted your posts to hell.

downvotes aren't arguments. addressing ideas, in my post, refers to intellectually addressing them – e.g. explaining why an idea is incorrect.

anyway, do you have any suggestion as to a Path Forward to get the intellectual disagreements resolved?

also did you actually read about Paths Forward? If so, why don't you reply to it directly and point out a mistake in it?

RowanE 08 November 2017 11:15:08PM 1 point [-]

Downvoting is not an argument because downvoting is a judgement that an idea is not worthy of "intellectually addressing" (on this forum). That's not not addressing an idea.

gjm 09 November 2017 12:01:22AM 0 points [-]

I did not claim that downvotes are arguments, of course. What they are is *assessments*. As it happens, your posts about CR here got comments as well as downvotes.

do you have any suggestion as to a Path Forward to get the intellectual disagreements resolved?

Not necessarily, just as if we were visited by fundamentalists demanding that everything be "proved from scripture" I would not necessarily have a suggestion as to how to Prove From Scripture that their

fundamentalism was wrong.

And I think that "to get the intellectual disagreements resolved" is a noble but hilariously overoptimistic goal. We are not, realistically, going to end up agreeing about everything, and picking an approach on the basis of whether it could in principle lead to us agreeing about everything is not a good idea.

| did you actually read about Paths Forward?

Yes.

| If so, why don't you reply to it directly and point out a mistake in it?

Because I think many other things I could do with the same time I could use for that would be more productive.

Dagon 09 November 2017 05:02:45AM 2 points



I'm thankful for this state. Nobody represents my beliefs, not even me. They are free-floating and do not need a spokesperson or representative to make them true (or correct them when they're not). Actually, beliefs are only models anyway, and there is no "true" except to the extent that they correlate so some observed states of the universe.

ChristianKI 11 November 2017 10:04:11PM 1 point



| But if no one has done that (so there's no answer to CR they can endorse), then how do they know CR is mistaken?

There are plenty of system of systems of thought out there and there are opportunity costs to spending energy delving into different systems of thoughts. Nothing you did here suggests to me that it's worthwhile to invest a significant amount of time into delving into CR.

| Suppose LW, vaguely defined as it may be, is mistaken on some major points. E.g. Karl Popper refuted induction.

Whether or not individual X did Y is not a major point as far most people on LW are concerned.

| Some people I talked with at LW seem to still be developing as intellectuals

Of course, part of the growth mindset is about constantly developing.

| No one is responsible for defining an LW school of thought

Creating a School of Thought would violate the idea of keeping identity small. We do sometimes use terms like "aspiring rationalist" or speak of LW but that's not the focus of our intellectual pursuits. We only use labels like that when they are useful.

| When there's a disagreement, ask yourself: "Suppose hypothetically that I'm wrong and the other guy is right. In what way would I ever find out and learn better?" If there's no good, realistic answer then you're bad at paths forward.

We have plenty of discussions where people change their minds.

In one of EY recent posts he described how he updated on String theory physicists knowing more than he previously thought because he took a bet.

Betting has the advantage of letting reality decide what's right. That's more important than providing clever arguments in favor of a position and as such it's valued more highly (or at least we try to value it more highly).

curi 11 November 2017 10:45:13PM * 0 points



| There are plenty of system of systems of thought out there and there are opportunity costs to spending energy delving into different systems of thoughts.

This is addressed in Paths Forward. You're just plain ignoring what I said. You aren't engaging with the answers to this that I already provided; you aren't pointing out where my reasoning was mistaken; you're just acting like half my ideas don't exist at all instead of actually arguing with them.

| Creating a School of Thought would violate the idea of keeping identity small.

As I explained already, PF is important on an individual basis and you should all individually do it regardless of whether LW does.

Induction was refuted decades ago and you guys aren't updating and don't have a mechanism to become less wrong about this.

ChristianKI 11 November 2017 10:59:22PM * 0 points



This is addressed in Paths Forward.

And I told you that you haven't made a case that suggests that it's worth reading.

Induction was refuted decades ago and you guys aren't updating and don't have a mechanism to become less wrong about this.

How do you know that we aren't updating? It seems to me like you are using induction to make that assessment. You observe that we don't update towards your arguments and you conclude that we don't update in general.

MrMind 09 November 2017 11:29:34AM 1 point



In my understanding, there's no one who speaks for LW, as its representative, and is *responsible* for addressing questions and criticisms.

Exactly. That is by design. See the title of the site? It doesn't say "MoreRight". Here even Yudkowski, the Founding Father, was frequently disagreed upon.

[This is the School-less school.](#)

metatroll 09 November 2017 05:15:50AM 1 point



Don't bother with these deadbeats. It's been over four years since I mentioned Kirly Prokastian, and there hasn't been a flicker of interest.

curi 09 November 2017 05:21:48AM 1 point



Who is Kirly Prokastian? Is there a typo? I only got one google hit which was an LW thread.

Is there a different site you'd recommend I bother with instead? I've been looking lots of places!

BTW feel free to come post at my forum <http://fallibleideas.com/discussion-info>

metatroll 09 November 2017 07:13:53AM 1 point



Boy am I embarrassed. Not only did I butcher the spelling, it's not even a person, it's a publishing company.
[Information here.](#)

Thank you for your invitation. I'll get in touch if I have anything worth saying.

curi 13 November 2017 08:37:28PM * 0 points



Additional info:

- <http://curi.us/2067-empiricism-and-instrumentalism>
- <http://curi.us/2065-open-letter-to-machine-intelligence-research-institute>
- <http://curi.us/2066-replies-to-gyrodidiot-about-fallible-ideas-critical-rationalism-and-paths-forward>
- <http://curi.us/2063-criticism-of-eliezer-yudkowsky-on-karl-popper>

Viliam 13 November 2017 12:03:41AM * 0 points



In my understanding, there's no one who speaks for LW, as its representative, and is *responsible* for addressing questions and criticisms. LW, as a school of thought, has no agents, no representatives – or at least none who are open to discussion.

As some have already said, this is considered a feature, not a bug. We do not care (or try not to care) about "what is the LW way?". Instead we (try to) focus on "how is it, really?". To [quote](#) Eliezer, who is closest to being *the representative of LW*:

Perhaps your conception of rationality is that it is rational to believe the words of the Great Teacher, and the Great Teacher says, "The sky is green," and you look up at the sky and see blue. If you think: "It may look like the sky is blue, but rationality is to believe the words of the Great Teacher," you lose a chance to discover your mistake.

So, it feels like you would like to have a phone number of the Great Teacher, to ask him about the color of the sky. While this site is -- if I may continue the metaphor -- trying to teach you how to actually look at the sky, and explaining how the human eye perceives colors.

Suppose I wrote some criticisms of the sequences, or some Bayesian book. Who will answer me? Who will fix the mistakes I point out, or canonically address my criticisms with counter-arguments? No one.

If you find that Sequences say "A" and truth is actually "B", what you can do is write an article on LW explaining why "B" is true. (Pointing out that Sequences say "A" is optional; I think it would be better done afterwards, so that

people can debate "B" independently. But do as you wish.)

It may happen is that different people will give different opinions. But then you can let them argue against each other.

| So how is progress to be made?

Here I may be just talking about myself, but I seek progress at a completely different place. I don't care that much about playing with words, which many intelligent people, including you, seem to be so fond of. I see humans, including myself, as deeply imperfect beings. No matter how much I am told "X", no matter how much I in theory agree with "X", if I pay enough attention, I find myself going against "X" all the time. Thus, instead of having yet another debate about virtues of "X", I would rather spend my attention trying to practice "X". Because as long as there is a huge gap between what I profess and what I actually do, it does not matter much whether I profess correct ideas. Actually, talking about rationality, it may be even worse. The ideas I profess can be not only right or wrong, but possibly also irrelevant, or confused, or utterly meaningless.

You linked a website. Let me just look at the first article: "[Why is Reason Important?](#)". You talk about something called "Reason". Do you mean some *hypothetical ideal* of reason, or how smart but imperfect people *actually do* it? Oh wait, let me ask even more important question: Are you even aware that there is a distinction between these two? Because the article does not reflect that.

Still reading the first paragraph: "Reason also rejects the idea that authorities can or should tell us what the truth is. Instead, we should judge ideas ourselves, and based on the content of the idea not the person who said it. Even if I am the person who said an idea, and I have a PhD, that doesn't count for anything"... Really? What is your opinion on the existence of atoms, or theory of relativity? I mean, the Einstein guy is just some unimportant rando; so did you develop the whole theory on your own? Did you do all the relevant experiments to confirm that atoms do indeed exist? Wait, I have a more important question: Even if you have personally verified the theory of relativity, why did you even decide that verifying this theory is worth your time? I mean, (1) there are millions of possible theories, and you certainly cannot verify all of them, and (2) the fact that Einstein and a few others believe in some specific theory "X" means absolutely nothing before you verified it for yourself, right? So, why did you even choose to pay attention to the theory of relativity, if Einstein's words mean nothing, and there were million other potential theories competing for your attention?

...this was just an example of what I meant by "playing with words". You wrote a whole website of arguments that I guess seem convincing to you, and yet I find mistakes in the very first paragraph of the very first article. If you can imagine that this is how I feel about almost each paragraph of each article on your website, you can understand why I am unimpressed, and why I don't want to go this way.

| Which challenges are addressed? All of them.

Okay, I am curious: did someone already tell you something similar to what I just did? If yes, could you please give me a pointer to how it was addressed?

curi 13 November 2017 01:35:17AM 0 points



you missed the intended point about representatives. the point is that anyone takes any responsibility for the ideas they believe are true. the point is e.g. that anyone be available to answer questions about some idea. if the idea has no representatives in the sense of people who think it's good and answer questions about it, then that's problematic. then it's hard to learn and there's no one to improve or advocate it.

If you find that Sequences say "A" and truth is actually "B", what you can do is write an article on LW explaining why "B" is true.

And then people don't like me, b/c i'm a heretic who denies induction, so they ignore it. when there is no mechanism for correcting errors, what you end up with is bias: people decide to pay attention, or not, according to social status, bias, etc.

No matter how much I am told "X", no matter how much I in theory agree with "X", if I pay enough attention, I find myself going against "X" all the time.

For *all* X? E.g. "don't murder"? This part isn't clear.

Do you mean some hypothetical ideal of reason, or how smart but imperfect people actually do it?

The tradition of reason deals with both. It offers some guiding principles and ideals, as well as practical guidance, rules of thumb, tips, etc. People have knowledge of both of these.

Really? What is your opinion on the existence of atoms, or theory of relativity? I mean, the Einstein guy is just some unimportant rando; so did you develop the whole theory on your own?

I am familiar with some science and able to make some judgements about scientific arguments myself. Especially using resources like asking questions of physicists I know and using books/internet. I don't helplessly take people's words for things; i seek out explanations at the level of detail i'm interested in and make a judgement. And science is an interest of mine.

I have no criticism of the atomic theory, no objection to it. I know some stuff about it and I agree. I don't know of any contrary position that's any good. I'm convinced by the reasoning, not the prestige of the reasoners.

I didn't personally do all the experiments. Why should I? I don't accept an experiment merely b/c the person who did it had a PhD, but I don't automatically reject it either. I make a judgement about the experiment (or idea)

instead of about the person's credentials.

I paid attention to physics, initially, because I found the arguments in the book *The Fabric of Reality* high quality and interesting. The book looked interesting to me so I read the opening paragraphs online and I thought they were good so I got the book. I didn't look for the book with the most prestigious author. I don't see what these historical detail matter, but you asked about them. Physics is important (we live in the physical world; we're made of atoms; we move; etc) and worthy of interest (though others are welcome to pursue other matters).

tl;dr: I won't take Einstein's word for it, but I can be impressed by his reasoning.

yet I find mistakes in the very first paragraph of the very first article

let's not jump to conclusions before discussing the matter. *we disagree*, or there is a misunderstanding.

Viliam 13 November 2017 11:28:02PM 0 points



And then people don't like me, b/c i'm a heretic who denies induction, so they ignore it.

Have you tried posting here an article about why induction is wrong? Preferably starting with an explanation of what you mean by "induction", just to make sure we are all debating the same thing.

Of course there is a chance that people will ignore the article, but I would be curious to learn e.g. why evolution gave so many organisms the ability of reinforcement learning, if the fundamental premise of reinforcement learning -- that things in future are likely to be similar to things in the past -- is wrong.

This part isn't clear.

(Yeah, that's me writing at midnight, after my daughter finally decides to go sleep. Sorry for that.)

What I mean was that for me personally, the greatest obstacle in "following reason" is not the reasoning part, but rather the following part. (Using the LW lingo, the greatest problem is not epistemic rationality, but instrumental rationality.) I feel quite confident that I am generally good at reasoning, or at least better than most of the population. What I have problem is to actually follow my own advice. Therefore, instead of developing smarter and smarter arguments, I rather wish to become better at implementing the things I already know.

And I suspect this is the reason why CFAR focuses on things like "trigger-action planning" et cetera, instead of e.g. publishing articles analysing the writings of Popper. The former simply seems to provide much more value than the latter.

Sometimes the lessons seem quite easy -- the map is not the territory; make sure you communicate meaning, not just words; be open to changing your mind in either direction; etc -- yet even after years of trying you are still sometimes doing it wrong. People enjoy "insight porn", but what they need is practicing the boring parts until they become automatic.

I don't accept an experiment merely b/c the person who did it had a PhD, but I don't automatically reject it either.

But do you privilege the hypothesis, if you heard it from a person with PhD?

Oh, I guess this may be another thing that I rarely find outside of LW: reasoning in degrees of gray, instead of black and white. I am not asking whether you take each Einstein's word as sacred. I am asking whether you increase the *probability* of something, if you learn that Einstein said so.

curi 14 November 2017 04:25:18AM 0 points



if i were to provide an anti-induction article, what properties should it have?

apparently it should be different in some way than the ones already provided by Popper and DD, as individual book chapters.

one question is whether it should assume the reader has background knowledge of CR.

if so, it's easy, it'll be short ... and people here won't understand it.

if not, it'll be long and very hard to understand, and will repeat a lot of content from Popper's books.

what about a short logical argument about a key point, which doesn't explain the bigger picture? possible, but people hate those. they don't respond well to them. they don't just want their view destroyed without understanding any alternative. and anyway their own views are too vague to to criticize in a quick, logical way b/c whatever part you criticize, they can do without. there is no clear, essential, philosophical core they are attached to. if advocates of induction actually knew their own position, in exacting detail, inside and out, then you could quickly point out a logical flaw and they'd go "omg, that makes everything fall apart". but when you deal with people who aren't very clear on their own position, and who actually think all their beliefs are full of errors and you just have to muddle through and do your best ... then what kind of short argument will work?

Viliam 16 November 2017 11:58:02PM * 2 points



| if i were to provide an anti-induction article, what properties should it have?

Regardless of the topic, I would say that the article should be easy to read, and relatively self-contained. For example, instead of "go read this book by Popper to understand how he defines X" you could define X using your own words, preferably giving an example (of course it's okay to *also* give a quote from Popper's book).

| one question is whether it should assume the reader has background knowledge of CR.

I don't even know what the abbreviation is supposed to mean. Seriously.

Generally, I think that the greatest risk is people *not even understanding* what you are trying to say. If you include links to other pages, I guess most people will not click them. Aim to *explain*, not to convince, because a failure in explaining is automatically also a failure in convincing.

Maybe it would make sense for you to look at the articles that I believe (with my very unclear understanding of what you are trying to say) may be most relevant to your topic:

- 1) "[Infinite Certainty](#)" (and its mathy sequel "[0 And 1 Are Not Probabilities](#)"), and
- 2) "[Scientific Evidence, Legal Evidence, Rational Evidence](#)".

Because it seems to me that the thing about Popper and induction is approximately this...

Simplicio: "Can science be 100% sure about something?"

Popper: "Nope, that would mean that scientists would never change their minds. But they sometimes do, and that is an accepted part of science. Therefore, scientists are never 100% sure of their theories."

Simplicio: "Well, if they can't prove anything with 100% certainty, why don't we just ignore them completely? It's just another opinion, right?"

Popper: "Uhm... wait a minute... scientists cannot *prove* anything, but they can... uhm... *disprove* things! Yeah, that's what they do; they make many theories, they disprove most of them, and the one that keeps surviving is the official winner, for the moment. So it's not like the scientists proved e.g. the theory of relativity, but rather that they disproved all known competing theories, and failed to disprove the theory of relativity (yet)."

To which I would give the following objection:

- 1) How exactly could it be impossible to prove "X", and yet possible to disprove "not X"? If scientists are able to *falsify* e.g. the hypothesis that "two plus two does not equal four", isn't it the same as *proving* the hypothesis that "two plus two equals four"?

I imagine that the typical situation Popper had in mind included a few explicit hypotheses, e.g. A, B, C, and then a remaining option "something else that we did not consider". So he is essentially saying that scientists can experimentally disprove e.g. B and C, but that's not the same as proving A. Instead, they proved "either A, or something else that we did not consider, but definitely neither B nor C". Shortly: B and C were falsified, but A wasn't proven. And as long as there remains an unspecified category "things we did not consider", there is always a chance that A is merely an approximate solution, and the real solution is still unknown.

But it doesn't always have to be like this. Especially in math. But also in real life. Consider this:

According to Popper, no matter how much scientific evidence we have in favor of e.g. theory of relativity, all it needs is *one experiment* that will falsify it, and then all good scientists should stop believing in it. And recently, theory of relativity was indeed falsified by an experiment. Does it mean we should stop teaching the theory of relativity, because now it was properly falsified?

With the benefit of hindsight, now we know there was a mistake in the experiment. But... that's exactly my point. The concepts of "proving" and "falsifying" are actually much closer than Popper probably imagined. You may have a hypothesis "H", and an experiment "E", but if you say that you falsified "H", it means you have a hypothesis "F" = "the experiment E is correct and falsifies the theory H". To falsify H by E is to prove F; therefore if F cannot be scientifically proven, then H cannot be scientifically falsified. Proof and falsification are not two fundamentally different processes; they are actually two sides of the same coin. To claim that the experiment E falsifies the hypothesis H, is to claim that you have a **proof** that "the experiment E falsifies the hypothesis H"... and the usual interpretation of Popper is that there are no proofs in science.

The answer generally accepted on LessWrong, I guess, is that what *really* happens in science is that people believe theories with greater and greater *probability*. Never 100%. But sometimes with a very high probability instead, and for most practical purposes such high probability works almost like certainty. Popper may insist that science is unable to actually *prove* that moon is not made of cheese, but the fact is that most scientists will behave as if they already had such proof; they are not going to keep an open mind about it.

Short version: Popper was right about inability to prove things with 100% certainty, but then he (or maybe just people who quote him) made a mistake of imagining that *disproving* things is a process fundamentally different from proving things, so you can at least disprove things with 100% certainty. My answer is that you can't even disprove things with probability 100%, but that's okay, because the "100%"

part was just a red herring anyway; what actually happens in science is that things are believed with greater *probability*.

IlyaShpitser 17 November 2017 01:18:45AM * 3 points 

You should probably actually read Popper before putting words in his mouth.

According to Popper, not matter how much scientific evidence we have in favor of e.g. theory of relativity, all it needs is one experiment that will falsify it, and then all good scientists should stop believing in it.

You found this claim in a book of his? Or did you read some Wikipedia, or what?

For example, this is a quote from the Stanford Encyclopedia of Philosophy:

Popper has always drawn a clear distinction between the logic of falsifiability and its applied methodology. The logic of his theory is utterly simple: if a single ferrous metal is unaffected by a magnetic field it cannot be the case that all ferrous metals are affected by magnetic fields. Logically speaking, a scientific law is conclusively falsifiable although it is not conclusively verifiable. Methodologically, however, the situation is much more complex: no observation is free from the possibility of error—consequently we may question whether our experimental result was what it appeared to be.

Thus, while advocating falsifiability as the criterion of demarcation for science, Popper explicitly allows for the fact that in practice a single conflicting or counter-instance is never sufficient methodologically to falsify a theory, and that scientific theories are often retained even though much of the available evidence conflicts with them, or is anomalous with respect to them.

You guys still do that whole "virtue of scholarship" thing, or what?

Viliam 17 November 2017 10:31:11AM 0 points 

You guys still do that whole "virtue of scholarship" thing, or what?

Well, this specific guy has a job and a family, and studying "what Popper believed" is quite low on his list of priorities. If you want to provide a more educated answer to curi, go ahead.

IlyaShpitser 17 November 2017 02:45:21PM * 2 points [-]

If you have a job and a family, and don't have time to get into what Popper actually said, maybe don't offer your opinion on what Popper actually said? That's just introducing bad stuff into a discussion for no reason.

Wovon man nicht sprechen kann, darüber muss man schweigen.

"The virtue of silence."

curi 17 November 2017 06:43:18PM * 0 points [-]

Yeah, good points in both comments. Why don't you come to my forum where we'll appreciate them? :)

<https://groups.yahoo.com/neo/groups/fallible-ideas/info>

IlyaShpitser 17 November 2017 07:26:47PM 2 points [-]

I don't think you and I have much to talk about.

[continue this thread »](#)

Elo 17 November 2017 06:58:27PM 0 points [-]

The virtue of silence is one of our 12 virtues here. That you don't know speaks to ignorance on your part. And perhaps on taking your own advice you might not have made this post at all. And maybe you would have learnt something instead.

curi 17 November 2017 06:55:40PM * 0 points [-]

| I don't even know what the abbreviation is supposed to mean. Seriously.

Do you even know the name of Popper's philosophy? Did you read the discussions about this that already happened on LW?

It seems that you're completely out of your depth, can't answer me, and don't want to make the effort to learn. You can't answer Popper, don't know of anyone or any writing that can, and are content with that. Your fellows here are the same way. So Popper goes unanswered and you guys stay wrong.

FYI Popper has lots of self-contained writing. Many of his book chapters are adapted from lectures, as you would know if you'd looked. I have written recommendations of which specific parts of Popper are best to read with brief comments on what they are about:

<http://fallibleideas.com/books#popper>

| If you include links to other pages, I guess most people will not click them.

Everything you say in your post, about Popper issues, demonstrates huge ignorance, but there are no Paths Forward for you to get better ideas about this. The methodology dispute needs to be settled first, but people (including you) don't want to do that.

Viliam 18 November 2017 03:53:25PM 1 point



| It seems that you're completely out of your depth, can't answer me, and don't want to make the effort to learn.

I generally agree with your judgment (assuming that the "effort to learn" refers strictly to Popper).

But before I leave this debate, I would like to point out that you (and Ilya) were able to make this (correct) judgment only because *I put my cards on the table*. I wrote, relatively shortly and without obfuscation, what I believe. Which allowed you to read it and conclude (correctly) "he is just an uneducated idiot". This allowed a quick resolution; and as a side effect I learned something.

This may or may not be ironically related to the idea of falsification, but at this moment I feel unworthy to comment on that.

Now I see two possible futures, and it is more or less your choice which one will happen:

Option 1:

You may try to describe (a) *your beliefs* about induction, (b) what *you believe* are LW beliefs about induction, and (c) why exactly are the supposed LW beliefs wrong, preferably with a specific example of a situation where following the LW beliefs would result in an obvious error.

This is the "high risk / high reward" scenario. It will cost you more time and work, and there is a chance that someone will say "oh, I didn't realize this before, but now I see this guy has a point; I should probably read more of what he says", but there is also a chance that someone will say "oh, he got Popper or LW completely wrong; I knew it was not worth debating him". Which is not necessarily a bad thing, but will probably feel so.

Yeah, there is also the chance that people will read your text and ignore it, but speaking for myself, there are two typical reasons why I would do that: either is text is written in a way that makes it difficult for me to decipher what exactly the author was actually trying to say; or the text depends on links to outside sources but my daily time budget for browsing internet is already spent. (That is why I selfishly urge you to write a self-contained article using your own words.) But other people may have other preferences. Maybe the best would be to add footnotes with references to sources, but make them optional for understanding the gist of the article.

Option 2:

You will keep saying: "guys, you are so confused about induction; you should definitely read Popper", and people at LW will keep thinking: "this guy is so confused about induction or about our beliefs about induction; he should definitely [read the Sequences](#)", and both sides will be frustrated about how the other side is unwilling to spend the energy necessary to resolve the situation. This is the "play it safe, win nothing" scenario. Also the more likely one.

Last note: Any valid argument made by Popper should be possible to explain without using the word "Popper" in text. Just like Pythagorean theorem is not about the person called Pythagoras, but about squares on triangles, and would be equally valid if instead it would be discovered or popularized by a completely different person; you could simply call it "squares-on-triangles theorem" and it would work equally well. (Related in Sequences: "[Guessing the teacher's password](#)"; "[Argument Screens Off Authority](#)".) If something is true about induction, it is true regardless of whether Popper did or didn't believe it.

curi 18 November 2017 07:18:31PM * 0 points

[1]

(b) what you believe are LW beliefs about induction,

when i asked for references to canonical LW beliefs, i was told that would make it a cult, and LW *does not have beliefs about anything*. since no pro-LW ppl could/would state or link to LW's beliefs about induction – and were hostile to the idea – i think it's unreasonable to ask me to. individual ppl at LW vary in beliefs, so how am i supposed to write a one-size-fits-all criticism?

LW ppl offer neither a one-size-fits-all pro-induction explanation nor do any of them offer it individually. e.g. you have not said how you think induction works. it's your job, not mine, to come up with some version of induction which you think actually works – and to do that while being aware of known issues that make that a difficult project.

again, there are methodology issues. unless LW gives targets for criticism – written beliefs anyone will take responsibility for the correctness of (you can do this *individually*, but you don't want to – you're busy, you don't care, whatever) – then we're kinda stuck (given also the unwillingness to address CR).

your refusal to use outside sources is asking me to rewrite material. why? some attempt to save time on your part. is that the right way to save time? no. could we talk about the right ways to save time? if you wanted to. but my comments about the right way to save time are in outside sources, primarily written by me, which you therefore won't read (e.g. the Paths Forward stuff, and i could do the Popper stuff linking only to my own stuff, which i have tons of, but that's still an outside source. i could copy/paste my own stuff here, but that's stupid. it's also awkward b/c i've intentionally not rewritten essays already written by my colleagues, b/c why do that? so i don't have all the right material written by myself personally, on purpose, b/c i avoid duplication.). so we're kinda stuck there. i don't want to repeat myself for literally more than the 50th time, for you personally (who hasn't offered me anything – not even much sign you'll pay attention, care, keep replying next week, anything), b/c you won't read 1) Popper 2) Deutsch 3) my own links to myself 4) my recent discussions with other LW ppl where i already rewrote a bunch of anti-induction arguments and wasn't answered.

as one example of many links to myself that you categorically don't want to address:

<http://curi.us/1917-rejecting-gradations-of-certainty> (including the comments)

Viliam 19 November 2017 03:55:12PM 0 points [-]

In the linked article, you seem to treat "refutation by criticism" as something absolute. Either something is refuted by criticism, or it isn't refuted by criticism; and in either case you have 100% certainty about which one of these two options it is.

There seems to be no space for situations like "I've read a quite convincing refutation of something, but I still think there is a small probability there was a mistake in this clever verbal construction". It either "was refuted" or it "wasn't refuted"; and as long as you are willing to admit some probability, I guess it by default goes to the "wasn't refuted" basket.

In other words, if you imagine a variable containing value "X was refuted by criticism", the value of this variable at some moment switches from 0 to 1, without any intermediate values. I mean, if you reject gradations of certainty, then you are left with a black-and-white situation where either you have the certainty, or you don't; but nothing in between.

If this is more or less correct, then I am curious about what exactly happens in the moment where the variable actually switches from 0 to 1. Imagine that you are doing some experiments, reading some verbal arguments, and thinking about them. At some moment, the variable is at 0 (the hypothesis was not refuted by criticism yet), and at the very next moment the variable is at 1 (the hypothesis was refuted by criticism). What exactly happened during that last fraction of a second? Some mental action, I guess, like connecting two pieces of a puzzle together, or something like this. But isn't there some *probability* that you actually connected those two pieces incorrectly, and maybe you will notice this only a few seconds (or hours, days, years) later? In other words, isn't the "refutation by criticism" conditional on the probability that you actually understood everything correctly?

If, as I incorrectly said in previous comments, one experiment *doesn't* constitute refutation of a hypothesis (because the experiment may be measured or interpreted incorrectly), then what exactly *does*? Two experiments? Seven experiments? Thirteen experiments and twenty four pages of peer-reviewed scientific articles? Because if you refute "gradations of certainty", then it must be that at some moment the certainty is not there, and at another moment there is... and I am curious about where and why is that moment.

| your refusal to use outside sources is asking me to rewrite material. why?

Throwing books at someone is generally known as "courtier's reply". The more text you throw at me, the smaller probability that I would read them. (Similarly, I could tell you to read Korzybski's *Science and Sanity*, and only come back after you mastered it, because I believe -- and I truly do -- that it is related to some mistakes you are making. Would you?)

There are some situations when things cannot be explained by a short text. For example, if a 10-years old kid would ask me to explain him quantum physics in less than 1 page of text, I would give up. -- So let me ask you; is Popper's argument against induction the kind of knowledge that cannot be explained to an intelligent adult person using less than 1 page of text; not even in a simplified form?

Sometimes the original form of the argument is not the best one. For example, Gödel spent hundreds of pages proving something that kids today could express as "any mathematical theorem can be stored on computer as a text file, which is kinda a big integer in base 256".

(Took him hundreds of pages, because people didn't have computers back then.) So maybe the book where Popper explained his idea is similarly not the most efficient way to explain the idea. Also, if an idea cannot be explained without pointing to the original source, that is a bit suspicious. On the other hand, of course, not everyone is skilled at explaining, so sometimes the text written by a skilled author has this advantage.

Summary:

I believe that your belief in "refutation by criticism" as something that either is or isn't, but doesn't have "gradation of certainty", is so fundamentally wrong that it doesn't make sense to debate further. Because this is the whole point of why probabilistic reasoning, Bayes theorem, etc. is so popular on LW. (Because probabilities is what you use when you don't have absolute certainty, and I find it quite ironic that I am explaining this to someone who read orders of magnitude more of Popper than I did.)

IlyaShpitser 20 November 2017 03:49:36PM * 1 point



Throwing books at someone is generally known as "courtier's reply".

The issue here also is Brandolini's law:

"The amount of energy necessary to refute bullshit is an order of magnitude bigger than to produce it."

The problem with the "courtier's reply" is you could always appeal to it, even if Scott Aaronson is trying to explain something about quantum mechanics to you, and you need some background (found in references 1, 2, and 3) to understand what he is saying.

There is a type 1 / type 2 error tradeoff here. Ignoring legit expert advice is bad, but being cowed by an idiot throwing references at you is also bad.

As usual with tradeoffs like these, one has to decide on a policy that is willing to tolerate some of one type of error to keep the error you care about to some desired level.

I think a good heuristic for deciding who is an expert and who is an idiot with references is credentialism. But credentialism has a bad brand here, due to a "love affair with amateurism" LW has. One of the consequences of this love affair is a lot of folks here

make the above trade off badly (in particular they ignore legit advice to read way too frequently).

[continue this thread »](#)

curi 20 November 2017 12:28:18AM 0 points [-]

I believe that your belief in "refutation by criticism" as something that either is or isn't, but doesn't have "gradation of certainty", is so fundamentally wrong that it doesn't make sense to debate further.

I think there's something really wrong when your reaction to disagreement is to think there's no point in further discussion. That leaves me thinking you're a bad person to discuss with. Am I mistaken?

Making mistakes isn't random or probabilistic. When you make a judgement, there is no way to know some probability that your judgement is correct. Also, if judgements need probabilities, won't your judgement of the probability of a mistake have its own probability? And won't that judgement also have a probability, causing an infinite regress of probability assignments?

Mistakes are unpredictable. At least some of them are. So you can't predict (even probabilistically) whether you made one of the unpredictable types of mistakes.

What you can do, fallibly and tentatively, is make judgements about whether a critical argument is correct or not. And you can, when being precise, formulate all problems in a binary way (a given thing either does or doesn't solve it) and consider criticisms binarily (a criticism either explains why a solution fails to solve the binary problem, or doesn't).

So let me ask you; is Popper's argument against induction the kind of knowledge that cannot be explained to an intelligent adult person using less than 1 page of text; not even in a simplified form?

That'd work fine if they knew everything or nothing about induction. However, it's highly problematic when they already have thousands of pages worth of misconceptions about induction (some of which vary from the next guy's misconceptions). The misconceptions include vague parts they don't realize are vague, non sequiturs they don't realize are non sequiturs, confusion about what induction is, and other mistakes *plus cover up* (rationalizations, dishonesty, irrationality).

Induction would be way easier to explain to a 10 year old in a page than to anyone at LW, due to lack of bias and prior misconceptions. I could also do quantum physics in a page for a ten year old. QM is easy to explain at a variety of levels of detail, if you don't have to include anything to preemptively address pre-existing misconceptions, objections, etc. E.g., in a sentence: "Science has discovered there are many things your eyes can't see, including trillions of other universes with copies of you, me, the Earth, the sun, everything."

[continue this thread »](#)

curi 14 November 2017 03:28:53AM * 0 points [-]

the moderators here actually just threatened my friend with a ban for posting a link to one of my articles about our philosophical disagreements, and deleted the thread. it was this one about empiricism and instrumentalism (not quite induction, but closely related): <http://curi.us/2067-empiricism-and-instrumentalism>

the reason you have trouble applying reason is b/c u understand reason badly. it's easy if u understand it well enuf. the idea/action gap is a matter of *flaws in the ideas* – both having the wrong ideas and also having incomplete ideas. ideas are what you need. nothing but ideas can help/save you.

insight porn sucks because its ideas aren't good enough, and are designed to impress people with standard memes, not to be useful. it's a trap which you shouldn't mix up with real philosophy.

also you ask about posting an anti-induction article. i wrote a number of anti-induction arguments both on the forums and in slack, which have not been answered. i also gave references to more, which have not been answered. why should everything be repeated for each individual who comes along and doesn't want to read references?

repeating arguments for people unwilling to look at the literature is not productive. it takes so much effort to understand philosophy that the effort of doing some reading is table stakes. people who don't want to do that are unserious. and you only have to read until the first mistake, and then comment. and if you're wrong about that first mistake, you can look for the second one and also take the matter more seriously. and by the 5th incorrect mistake i expect your full attention.

the methodology disagreements need to come before the induction disagreement or we won't be discussing induction using the same rules of discussion.

and you ask me to define induction so we're on the same page. that's part of the problem. ppl are LW are not on the same page, and want to all be addressed individually – which is too much work, and anyway none of them take responsibility for finding the truth, they all just quit after a small amount of discussion, as i expect you to as well. if you want to learn, join FI (<http://fallibleideas.com/discussion-info>) and ask and ppl will help you. or read. <http://fallibleideas.com/books> or look through the discussions i already had here (both recently and years ago) and answer the points that others did not. you can find logs of the slack chats at <https://groups.yahoo.com/neo/groups/fallible-ideas/info>

more broadly, inductivists vary so much – and most barely know anything about induction. so there's no really short one-size-fits-all way to address the issue. it's a big topic. hence lots of important arguments – and, perhaps more importantly, extensive explanation of the alternative.

if you really want an anti-induction article, one of the best things you could do, first, is *give me a pro-induction article* you endorse, and stand behind, and take responsibility for. shouldn't that come first? but when i asked for canonical LW material that would be appropriate to respond to, and that anyone would care if it was mistaken ... i was flamed. lay out your positive claims in a serious way – stick your neck out as CR has – before asking for refutation of your unspecified positive claims.

and no i don't give ideas probabilities. <https://yesornphilosophy.com>

Lumifer 14 November 2017 05:20:39PM 0 points [-]

| just threatened my friend

Justin used to be merely "a member of the FI forum. he followed one of my links". But now it turns out you're a team?

| the reason you have trouble applying reason is b/c u understand reason badly

Mirrors. They are a thing, you should look into one.

| why should everything be repeated ... is not productive ... people ... are unserious

Clearly, your valuable time is wasted here. You probably should go find emptier vessels to fill with your wisdom.

| first, is give me a pro-induction article you endorse

I described to you my approach to induction. Were there any fatal flaws you noticed but didn't mention?

and no i don't give ideas probabilities

Yes, we understand your approach has problems :-P

curi 14 November 2017 07:38:24PM 0 points [-]

Stop making hostile assumptions, I wasn't even talking about Justin.

username2 14 November 2017 11:42:29AM 0 points [-]

Please keep posting here. Your powers of persuasion are amazing.

RealJustinCEO 10 November 2017 12:26:41AM 0 points [-]

i wrote some commentary on Paths Forward, thought people might appreciate it:

<http://justinmallone.com/2017/01/paths-forward-comments-part-1/>

Elo 09 November 2017 07:48:15AM 0 points [-]

There are paths forward but it's no one else's job to show them to you or help you find them.

The attitude of expectation is one that does not fit here and will not encourage anyone to help you. I also expect it runs into walls in other cultures too.

curi comments on Less Wrong Lacks Representatives and Paths Forward - Less Wrong

1 curi 08 November 2017 07:00PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(85\)](#)

Comments (85)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above.](#)

curi 20 November 2017 12:28:18AM 0 points



I believe that your belief in "refutation by criticism" as something that either is or isn't, but doesn't have "gradation of certainty", is so fundamentally wrong that it doesn't make sense to debate further.

I think there's something really wrong when your reaction to disagreement is to think there's no point in further discussion. That leaves me thinking you're a bad person to discuss with. Am I mistaken?

Making mistakes isn't random or probabilistic. When you make a judgement, there is no way to know some probability that your judgement is correct. Also, if judgements need probabilities, won't your judgement of the probability of a mistake have its own probability? And won't that judgement also have a probability, causing an infinite regress of probability assignments?

Custom Search

alanforr

Create new article

Karma Score

Submit a new link

17

Add new meetup

22

Preferences



Log out

Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 01:45PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.

Less Wrong Study Hall

RECENT COMMENTS

Sorry, did you say weird/esoteric
by IlyaShpitser on LW 2.0 Open Beta
Live | 1 point

LessWrong

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

IlyaShpitser comments on Less Wrong Lacks Representatives and Paths Forward - Less Wrong

1 curi 08 November 2017 07:00PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(85\)](#)

Comments (85)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above.](#)

IlyaShpitser 17 November 2017 07:26:47PM 2 points [-]

I don't think you and I have much to talk about.

curi 17 November 2017 08:03:44PM 0 points [-]

Why?

Elo 17 November 2017 08:30:22PM 2 points [-]

a. virtue of silence

b. it's your job to work that out.

Custom Search

alanforr

Create new article

Karma Score

Submit a new link

17

Add new meetup

22

Preferences



Log out

Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 01:45PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Sorry, did you say weird/esoteric
by **IlyaShpitser** on LW 2.0 Open Beta Live | 1 point

curi 17 November 2017 09:05:21PM * 1 point [-]

What happened to NVC (Non-Violent Communication)? Your comments are purely intended to hurt me.

Elo 17 November 2017 09:36:00PM 1 point [-]

No. That's your interpretation. You have agency too to interpret what I say with clarity. You also value bold conjecture. So that's again your problem to work out what I mean and how to apply it.

IlyaShpitsner 18 November 2017 08:00:26PM 0 points [-]

Everything you say in your post, about Popper issues, demonstrates huge ignorance.

Do you even know the name of Popper's philosophy?

It seems that you're completely out of your depth.

The reason you have trouble applying reason is b/c u understand reason badly.

I have a thought. Since you are a philosopher, would your valuable time not be better spent doing activities philosophers engage in, such as writing papers for philosophy journals?

Rather than arguing with people on the internet?

If you are here because you are fishing for people to go join your forum, may I suggest that this place is an inefficient use of your time? It's mostly dead now, and will be fully dead soon.

curi 18 November 2017 08:08:29PM 0 points [-]

I have a low opinion of academic philosophers and philosophy journals. I was hoping to find a little intelligence somewhere. I have tried a lot of places. If you have better suggestions than philosophy journals or LW, let me know.

Yes, the article I was looking for.
by **fblogin** on Double Crux — A Strategy for Resolving Disagreement | 0 points

Sorry, but it is. Simple test: open a
by **elharo** on LW 2.0 Open Beta Live | 0 points

[Lucifer's version]
by **roland** on The "Outside the Box" Box | 0 points

Interesting discussion. Does
by **kimberchoi** on Actions and Words: Akrasia and the Fruit of Self-Knowledge | 0 points

RECENT POSTS

LW 2.0 Open Beta Live
by Vaniver | 22v (36c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

European Community Weekend 2017
by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about *protecting* the epistemology
by AnnaSalamon | 73v (54c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)

Mistakes are unpredictable. At least some of them are. So you can't predict (even probabilistically) whether you made one of the unpredictable types of mistakes.

What you can do, fallibly and tentatively, is make judgements about whether a critical argument is correct or not. And you can, when being precise, formulate all problems in a binary way (a given thing either does or doesn't solve it) and consider criticisms binarily (a criticism either explains why a solution fails to solve the binary problem, or doesn't).

So let me ask you; is Popper's argument against induction the kind of knowledge that cannot be explained to an intelligent adult person using less than 1 page of text; not even in a simplified form?

That'd work fine if they knew everything or nothing about induction. However, it's highly problematic when they already have thousands of pages worth of misconceptions about induction (some of which vary from the next guy's misconceptions). The misconceptions include vague parts they don't realize are vague, non sequiturs they don't realize are non sequiturs, confusion about what induction is, and other mistakes *plus cover up* (rationalizations, dishonesty, irrationality).

Induction would be way easier to explain to a 10 year old in a page than to anyone at LW, due to lack of bias and prior misconceptions. I could also do quantum physics in a page for a ten year old. QM is easy to explain at a variety of levels of detail, if you don't have to include anything to preemptively address pre-existing misconceptions, objections, etc. E.g., in a sentence: "Science has discovered there are many things your eyes can't see, including trillions of other universes with copies of you, me, the Earth, the sun, everything."

Viliam 20 November 2017 01:09:29AM * 0 points



I think there's something really wrong when your reaction to disagreement is to think there's no point in further discussion.

It's like you believe "A" and "A implies B" and "B implies C", while I believe "non-A" and "non-A implies Q". The point we should debate is whether "A" or "non-A" is correct; because as long as we disagree on this, of course each of us is going to believe a different chain of things (one starting with "A", the other starting with "non-A").

I mean, if I hypothetically would believe that absolute certainty is possible and relatively simple to achieve, of course I would consider the probabilistic reasoning to be interesting but inferior form of reasoning. We wouldn't have this debate. And if you would accept that certainty is impossible (even certainty of refutation), then probability would probably seem like the next best thing.

When you make a judgement, there is no way to know some probability that your judgement is correct.

Yes, the article I was looking for.
by **fblogin** on Double Crux — A Strategy for Resolving Disagreement | 0 points

Sorry, but it is. Simple test: open a
by **elharo** on LW 2.0 Open Beta Live | 0 points

[Lucifer's version]
by **roland** on The "Outside the Box" Box | 0 points

Interesting discussion. Does
by **kimberchoi** on Actions and Words: Akrasia and the Fruit of Self-Knowledge | 0 points

RECENT POSTS

LW 2.0 Open Beta Live
by Vaniver | 22v (36c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

European Community Weekend 2017
by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about *protecting* the epistemology
by AnnaSalamon | 73v (54c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)

Okay, imagine this: I make a judgment that feels completely correct to me, and I am not aware of any possible mistakes. But of course I am a fallible human, maybe I actually made a mistake somewhere, maybe even an embarrassing one.

Scenario A: I made this judgement at 10 AM, after having a good night of sleep.

Scenario B: I made this judgement at 2 AM, tired and sleep deprived.

Does it make sense to say that the *probability* of making the mistake in the judgment B is *higher* than the probability of making the mistake in the judgment A? In both cases I believe at the moment that the judgment is correct. But in the latter case my ability to notice the possible mistake is smaller.

So while I couldn't make an exact calculation like "the probability of the mistake is exactly 4.25%", I can still be aware that there is *some* probability of the mistake, and sometimes even estimate that the probability in one situation is greater than in another situation. Which suggests that there is a number, I just don't know it. (But if we could somehow repeat the whole situation million times, and observe that I was wrong in 42500 cases, that would suggest that the probability of the mistake is about 4.25%. Unlikely in real life, but possible as a hypothesis.)

Also, if judgements need probabilities, won't your judgement of the probability of a mistake have its own probability?

It definitely will. Notice that those are two different things: (a) the probability that I am wrong, and (b) *my estimate* of the probability that I am wrong.

Yes, what you point out is a very real and very difficult problem. Estimating probabilities in a situation where everything (including our knowledge of ourselves, and even our knowledge of math itself) is... complicated. Difficult to do, and even more difficult to justify in a debate.

This may even be a hard limit on human certainty. For example, if at every moment of time there is a 0.00000000001 probability that you will go insane, that would mean you can never be sure about anything with probability greater than 0.99999999999, because there is always the chance that however logical and reasonable something sounds to you at the moment, it's merely because you have become insane at this very moment. (The cause of insanity could be e.g. a random tumor or a blood vessel breaking in your brain.) Even if you would make a system more reliable than a human, for example a system maintained by hundred humans, where if anyone goes insane, the remaining ones will notice it and fix the mistake, the system itself could achieve higher certainty, but you, as an individual, reading its output, could not. Because there would always be the chance that you just got insane, and what you believe you are reading isn't actually there.

Relevant LW article: "[Confidence levels inside and outside an argument](#)".

Fact Posts: How and Why

by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement

by Duncan_Sabien | 61v (103c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all
by vaultDweller on Rationality Quotes
April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

IlyaShpitser (22)
alanforr (22)
gjm (19)
Elo (17)
Lumifer (16)
curi (15)
cousin_it (7)
gwern (5)
mirefek (5)
OrthernLight (5)
Oscar_Cunningham (5)
Dagon (4)
JenniferRM (4)
morganism (4)
jimrandomh (4)

RECENT KARMA AWARDS

And you can, when being precise, formulate all problems in a binary way (a given thing either does or doesn't solve it) and consider criticisms binarily (a criticism either explains why a solution fails to solve the binary problem, or doesn't).

Suppose the theory predicts that an energy of a particle is 0.04 whatever units, and my measurement detected 0.041 units. Does this falsify the theory? Does 0.043, or 0.05, or 0.08? Even when you specify the confidence interval, it is ultimately a probabilistic answer. (And saying " $p < 0.05$ " is also just an arbitrary number; why not " $p < 0.001$ "?)

You can have a "binary" solution only as long as you remain in the realm of words. ("Socrates is a human. All humans are mortal. Therefore Socrates is mortal. Certainty of argument: 100%.") Even there, the longer chain of words you produce, the greater chance that you made a mistake somewhere. I mean, if you imagine a syllogism going over thousand pages, ultimately proving something, you would probably want to check the whole book at least two or three times; which means you wouldn't feel a 100% certainty after the first reading. But the greater problems will appear on the boundary between the words and reality. (Theory: "the energy of the particle X is 0.04 units"; the experimental device displays 0.041. Also, the experimental devices sometimes break, and your assistant sometimes records the numbers incorrectly.)

it's highly problematic when they already have thousands of pages worth of misconceptions

Fair point.

(BTW, I'm going offline for a week now; for reasons unrelated to LW or this debate.)

EDIT:

For the record: Of course there are things where I consider the probability to be so high or so low that I treat them *for all practical purposes* as 100% or 0%. If you ask me e.g. whether gravity exists, I will simply say "yes"; I am not going to role-play Spock and give you a number with 15 decimal places. I wouldn't even know exactly how many nines are there after the decimal dot. (But again, there is a difference between "believing there is a probability" and "being able to tell the exact number".)

The most obvious impact of probabilistic reasoning on my behavior is that I generally don't trust long chains of words. Give me 1000 pages of syllogisms that allegedly prove something, and my reaction will be "the probability that *somewhere* in that chain is an error is so high that the conclusion is completely unreliable". (For example, I am not even trying to understand Hegel. Yeah, there are also other reasons to distrust him specifically, but I would not trust such long chain of logic without experimental confirmation of intermediate results from any author.)

curi 20 November 2017 01:41:40AM 0 points

[]

Does it make sense to say that the probability of making the mistake in the judgment B is higher than the probability of making the mistake in the judgment A?

It may or may not make sense, depending on terminology and nuances of what you mean, for *some types* of mistakes. Some categories of error have some level of predictability b/c you're already familiar with them. However, it *does not* make sense for *all* types of mistakes. There are some mistakes which are simply unpredictable, which you know nothing about in advance. Perhaps you can partly, in some way, see some mistakes coming – but that doesn't work in all cases. So you can't figure out any overall probability of some judgement being a mistake, because at most you have a probability which addresses some sources of mistakes but others are just unknown (and you can't combine "unknown" and "90%" to get an overall probability).

I am a fallibilist who thinks we can have neither 100% certainty nor 90% certainty nor 50% certainty. There's always framework questions too – e.g. you may say according to your framework, given your context, then you're unlikely (20%) to be mistaken (btw my main objections remain the same if you stop quantifying certainty with numbers). But you wouldn't know the probability your framework has a mistake, so you can't get an overall probability this way.

Difficult to do, and even more difficult to justify in a debate.

if you're already aware that your system doesn't really work, due to this regress problem, why does no one here study the philosophy which has a solution to this problem? (i had the same kind of issue in discussions with others here – they admitted their viewpoint has known flaws but stuck to it anyway. knowing they're wrong in some way wasn't enough to interest them in studying an alternative which claims not to be wrong in any known way – a claim they didn't care to refute.)

This may even be a hard limit on human certainty.

the hard limit is we don't have certainty, we're fallible. that's it. what we have, knowledge, is something else which is (contra over 2000 years of philosophical tradition) different than certainty.

Suppose the theory predicts that an energy of a particle is 0.04 whatever units, and my measurement detected 0.041 units. Does this falsify the theory? Does 0.043, or 0.05, or 0.08? Even when you specify the confidence interval, it is ultimately a probabilistic answer. (And saying " $p < 0.05$ " is also just an arbitrary number; why not " $p < 0.001$ "?)

you have to make a decision about what standards of evidence you will use for what purpose, and why that's the right thing to do, and expose that meta decision to criticism.

the epistemology issues we're talking about are prior to the physics issues, and don't involve that kind of measurement error issue. we can talk about measurement error after resolving epistemology. (the big picture is that probabilities and statistics have some use in life, but they aren't probabilities of truth/knowledge/certainty, and their use is governed by non-probabilistic judgements/arguments/epistemology.)

see <http://curi.us/2067-empiricism-and-instrumentalism> and <https://yesornophilosophy.com>

| You can have a "binary" solution only as long as you remain in the realm of words.

no, a problem can and should specify criteria of what the bar is for a solution to it. lots of the problems ppl have are due to badly formulated (ambiguous) problems.

| which means you wouldn't feel a 100% certainty after the first reading

i do not value certainty as a *feeling*. i'm after objective knowledge, not feelings.

gjm 20 November 2017 03:18:13AM 2 points



| If you're already aware that your system doesn't work, due to this regress problem,

That isn't what Viliam said, and I suggest that here you're playing rhetorical games rather than arguing in good faith. It's as if someone took your fallibilism and your rejection of probability, and said "Since you admit that you could well be wrong and you have no idea how likely it is that you're wrong, why should we take any notice of what you say?".

| why does no one here study the philosophy which has a solution to this problem?

You mean "the philosophy which claims to have a solution to this problem". (Perhaps it really does, perhaps not; but all someone can know in advance of studying it is that it claims to have one.)

Anyway, I think the answer depends on what you mean by "study". If you mean "investigate at all" then the answer is that several people here *have* considered some version of Popperian "critical rationalism", so your question has a false premise. If you mean "study in depth" then the answer is that by and large those who've considered "critical rationalism" have decided after a quick investigation that its claim to have the One True Answer to the problem of induction is not credible enough for it to be worth much further study.

My own epistemic state on this matter, which I mention not because I have any particular importance but because I know my own mind much better than anyone else's, is that I've read a couple of Deutsch's books and some of his other writings and given Deutsch's version of "critical rationalism" hours, but not weeks, of thought, and that since you turned up here I've given some further attention to your version; that c.r. seems to me to contain some insights and some outright errors; that I do not find it credible that c.r. "solves" the problem of getting information from observations in any strong sense; that I find the claims made by some c.r. proponents that (e.g.) there is no such thing as induction, or that it is a mistake to assign probabilities to statements that aren't explicitly about random events, even less credible; that the "return on investment" of further in-depth investigation of Popper's or Deutsch's ideas is likely worse than that of other things I could do with the same resources of time and brainpower, not because they're all bad ideas but because I think I already grasp them well enough for my purposes.

the epistemology issues [...] are prior to the physics issues, and don't involve that kind of measurement error issue.

A good epistemology needs to deal with the fact that observations have errors in them, and it makes no sense to try to "resolve epistemology" in a way that ignores such errors. (Perhaps that isn't what you meant by "we can talk about measurement error after resolving epistemology", in which case some clarification would be a good idea.)

What we have, knowledge, is something else which is (contra over 2000 years of philosophical tradition) different than certainty.

You say that as if you expect it to be a new idea around here, but it isn't. See e.g. [this old LW article](#). For the avoidance of doubt, I'm not claiming that what that says about knowledge and certainty is *the same as you would say* -- it isn't -- nor that what it says is original to its author -- it isn't. Just that *distinguishing knowledge from certainty* is something we're already comfortable with.

I do not value certainty as a *feeling*.

You would equally not *be entitled to* a 100% certainty, or have any other sort of 100% certainty you might regard as more objective and less dependent on feelings. (Because in the epistemic situation Viliam describes, it would be *very likely* that at least one error had been made.)

Of course, in principle you admit exactly this: after all, you call yourself a fallibilist. But, while you admit the possibility of error and no doubt actually change your mind sometimes, you refuse to try to *quantify* how error-prone any particular judgement is. I think this is "obviously" a mistake (i.e., obviously when you look at things rightly, which may not be an easy thing to do) and I think Viliam probably thinks the same.

(And when you complain above of an infinite regress, it's precisely about what happens when one tries to quantify these propensities-to-error, and your approach avoids this regress not by *actually handling it any better* but by *simply declaring that you aren't going to try to quantify*. That might be OK if your approach handled such uncertainties just as well by other means, but it doesn't seem to me that it does.)

curi 20 November 2017 03:57:00AM 0 points [-]

you haven't cared to try to write down, with permalink, any errors in CR that you think could survive critical scrutiny.

by study i mean look at it enough to find something wrong with it – a reason not to look further – or else keep going if you see no errors. and then write down what the problem is, ala Paths Forward.

| the claims made by some c.r. proponents

it's dishonest (or ignorant?) to refer to Popper, Deutsch and myself (as well as Miller, Bartley, and more or less everyone else) as "some c.r. proponents".

| you refuse to try to quantify how error-prone any particular judgement is.

no. i have tried and found it's impossible, and found out *why* (arguments u don't wish to learn).

anyway i don't see what your comment is supposed to accomplish. you have 1.8 of your feet out the door. you aren't really looking to have a conversation to resolve the matter. why speak at all?

gjm 21 November 2017 03:29:29AM * 0 points [-]

| you haven't cared to [...]

Correct: I am not interested in jumping through the idiosyncratic set of hoops you choose to set up.

| it's dishonest (or ignorant?) [...]

Why?

| arguments you don't wish to learn

Don't wish to *learn* them? True enough. I don't see your relationship to me as being that of teacher to learner. I'd be interested to hear what they are, though, if you could drop the superior attitude and try having an actual discussion.

| I don't see what your comment is supposed to accomplish.

It is supposed to point out some errors in things you wrote, and to answer some questions you raised.

| you have 1.8 of your feet out the door.

Does that actually mean anything? If so, what?

| you aren't really looking to have a conversation to resolve the matter.

I am very willing to have a conversation. I am not interested in straitjacketing that conversation with the arbitrary rules you keep trying to impose ("paths forward"), and I am not interested in replacing the (to me, potentially interesting) conversation about probability and science and reasoning and explanation and knowledge with the (to me, almost certainly boring and fruitless) conversation about "paths forward" that you keep trying to replace it with.

| why speak at all?

See above. You said some things that I think are wrong, and you asked some questions I thought I could answer. It's not *my* problem that you're unable or unwilling to address any of the actual content of what I say and only interested in meta-issues.

[EDITED because I noticed I wrote "conservation" where I meant "conversation" :-)]

curi 21 November 2017 04:13:57AM 0 points

[-]

you have openly stated your unwillingness to

1) do PF

2) discuss PF or other methodology

that's an impasse, created by you. you won't use the methodology i think is needed for making progress, and won't discuss the disagreement. a particular example issue is your hostility to the use of references.

the end.

| I am very willing to have a conversation.

given your rules, including the impasse above.

gjm 21 November 2017 01:37:35PM 2 points [-]

| you have openly stated your unwillingness [...]

Yup. I'm not interested in jumping through the idiosyncratic set of hoops you choose to set up.

| that's an impasse, created by you.

Curiously, I find myself perfectly well able to conduct discussions with pretty much everyone else I encounter, including people who disagree with me at least as much as you do. That would be because they don't try to lay down a bunch of procedural rules and refuse to engage unless I either follow their rules or get sidetracked onto a discussion of those rules. So ... nah, I'm not buying "created by you". I'm not the one who tried to impose the absurdly over-demanding set of procedural rules on a bunch of other people.

| your hostility to the use of references

You just made that up. I am not hostile to the use of references.

(Maybe I objected to something you did that involved the use of references; I don't remember. But if I did, it wasn't because I am hostile to the use of references.)

Lumifer 20 November 2017 03:50:14PM 0 points [-]

| you aren't really looking to have a conversation to resolve the matter

Your understanding of "resolve the matter" is very peculiar -- as far as I can see it means "go read what I tell you to read so that you will agree with me".

I notice that you show considerable lack of flexibility: you follow a certain pattern of interaction which, to no great surprise, tends to end up in the same place, you get nowhere and accuse people of bad faith and unwillingness to learn.

You've been hanging around the place for a few weeks by now -- how about you, did you learn anything? Or this is strictly a bring-civilization-to-the-savages expedition from your point of view?

LessWrong

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

IlyaShpitser comments on Less Wrong Lacks Representatives and Paths Forward - Less Wrong

1 curi 08 November 2017 07:00PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(85\)](#)

Comments (85)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above.](#)

IlyaShpitser 17 November 2017 07:26:47PM 2 points 

I don't think you and I have much to talk about.

curi 17 November 2017 08:03:44PM 0 points 

Why?

Elo 17 November 2017 08:30:22PM 2 points 

a. virtue of silence

b. it's your job to work that out.

Custom Search

alanforr

Create new article

Karma Score

Submit a new link

17

Add new meetup

22

Preferences



Log out

 Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 01:45PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

Sorry, did you say weird/esoteric
by **IlyaShpitser** on LW 2.0 Open Beta Live | 1 point

curi 17 November 2017 09:05:21PM * 1 point [-]

What happened to NVC (Non-Violent Communication)? Your comments are purely intended to hurt me.

Elo 17 November 2017 09:36:00PM 1 point [-]

No. That's your interpretation. You have agency too to interpret what I say with clarity. You also value bold conjecture. So that's again your problem to work out what I mean and how to apply it.

IlyaShpitsner 18 November 2017 08:00:26PM 0 points [-]

Everything you say in your post, about Popper issues, demonstrates huge ignorance.

Do you even know the name of Popper's philosophy?

It seems that you're completely out of your depth.

The reason you have trouble applying reason is b/c u understand reason badly.

I have a thought. Since you are a philosopher, would your valuable time not be better spent doing activities philosophers engage in, such as writing papers for philosophy journals?

Rather than arguing with people on the internet?

If you are here because you are fishing for people to go join your forum, may I suggest that this place is an inefficient use of your time? It's mostly dead now, and will be fully dead soon.

curi 18 November 2017 08:08:29PM 0 points [-]

I have a low opinion of academic philosophers and philosophy journals. I was hoping to find a little intelligence somewhere. I have tried a lot of places. If you have better suggestions than philosophy journals or LW, let me know.

Yes, the article I was looking for.
by **fblogin** on Double Crux — A Strategy for Resolving Disagreement | 0 points

Sorry, but it is. Simple test: open a
by **elharo** on LW 2.0 Open Beta Live | 0 points

[Lucifer's version]
by **roland** on The "Outside the Box" Box | 0 points

Interesting discussion. Does
by **kimberchoi** on Actions and Words: Akrasia and the Fruit of Self-Knowledge | 0 points

RECENT POSTS

LW 2.0 Open Beta Live
by Vaniver | 22v (36c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

European Community Weekend 2017
by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about *protecting* the epistemology
by AnnaSalamon | 73v (54c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)

IlyaShpitser comments on Less Wrong Lacks Representatives and Paths Forward - Less Wrong

1 curi 08 November 2017 07:00PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

Article Navigation

[Comments \(85\)](#)

Comments (85)

Sort By: **Best**

You are viewing a single comment's thread. [Show more comments above.](#)

IlyaShpitser 20 November 2017 03:49:36PM * 1 point 

Throwing books at someone is generally known as "courtier's reply".

The issue here also is Brandolini's law:

"The amount of energy necessary to refute bullshit is an order of magnitude bigger than to produce it."

The problem with the "courtier's reply" is you could always appeal to it, even if Scott Aaronson is trying to explain something about quantum mechanics to you, and you need some background (found in references 1, 2, and 3) to understand what he is saying.

There is a type 1 / type 2 error tradeoff here. Ignoring legit expert advice is bad, but being cowed by an idiot throwing references at you is also bad.

Custom Search

alanferr

Create new article

Karma Score

Submit a new link

17

Add new meetup

22

Preferences



Log out

 [Subscribe to RSS Feed](#)

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 01:45PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.

Less Wrong Study Hall

RECENT COMMENTS

Sorry, did you say weird/esoteric
by **IlyaShpitser** on LW 2.0 Open Beta Live | 1 point



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

David Deutsch on How To Think About The Future

4 curi 11 April 2011 07:08AM

<http://vimeo.com/22099396>

What do people think of this, from a Bayesian perspective?

It is a talk given to the Oxford Transhumanists. Their previous speaker was Eliezer Yudkowsky. Audio version and past talks here: <http://groups.spaces.com/oxfordtranshumanists/pages/past-talks>

Article Navigation

Comments (197)

Tags: deutsch popper Delete

Comments (197)

Sort By: New

Enter a comment here

Comment

Show help

timtyler 15 April 2011 02:50:51PM * 0 points

Deutsch gives Malthus as an example of a failed pessimistic prediction - at 23:00. However, it *still* looks as though Malthus is likely to have been correct. Populations increase exponentially, while resources expand at most in a polynomial fashion - due to the light cone. Deutsch discusses this point 38:00 minutes in, claiming relativistic time dilation changes this conclusion, which I don't think it really does: you still wind-up with most organisms being resource-limited, just as Malthus described.

timtyler 15 April 2011 02:21:50PM * 0 points

Martin Rees is misrepresented 4:04 in. What Rees actually said was:

the odds are no better than 50-50 that our present civilisation on Earth will survive to the end of the present century without a serious setback'

...whatever a "serious setback" is supposed to mean.

vallinder 14 July 2011 05:56:07PM 1 point

Do you have a reference for that? My copy of Our Final Hour contains the same sentence minus "without a serious setback".

timtyler 14 July 2011 10:35:07PM * 1 point

Our Final Century, page 8 line 4.

It seems as though Rees - rather confusingly - said different things on the topic in Our Final Century and [Our Final Hour](#).

Custom Search

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

Create

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

- Denver Area LW November Meetup: 07 November 2017 08:00PM
- Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

- Yeah, that's a bug. I am planning to by [Habryka](#) on LW 2.0 Open Beta Live | 0 points
- Woo! Also if anyone else gets a by [korin43](#) on LW 2.0 Open Beta Live | 0 points
- Running through this to check that by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points
- No, you don't need update you by [trickster](#) on What Evidence Filtered Evidence? | 0 points
- I think that the core of religion—that by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

The Conjunction Fallacy Does Not Exist

-38 curi 10 April 2011 10:35PM

The conjunction fallacy says that people attribute higher probability to X&Y than to Y.

This is false and misleading. It is based on bad pseudo-scientific research designed to prove that people are biased idiots. One of the intended implications, which the research does nothing to address, is that this is caused by genetics and isn't something people can change except by being aware of the bias and compensating for it when it will happen.

In order to achieve these results, the researchers choose X, Y, and the question they ask in a special way. Here's what they don't ask:

What's more likely this week, both a cure for cancer and a flood, or a flood?

Instead they do it like this:

http://en.wikipedia.org/wiki/Conjunction_fallacy

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?

Linda is a bank teller.

Linda is a bank teller and is active in the feminist movement.

Or like this:

http://lesswrong.com/lw/ji/conjunction_fallacy/

"A complete suspension of diplomatic relations between the USA and the Soviet Union, sometime in 1983."

"A Russian invasion of Poland, and a complete suspension of diplomatic relations between the USA and the Soviet Union, sometime in 1983."

These use different tricks. But both are biased in a way that biases the results.

By the way, this is a case of the general phenomenon that bad research often gets more impressive results, which is explained in *The Beginning of Infinity* by David Deutsch. If they weren't bad researchers and didn't bias their research, they would have gotten a negative result and not had anything impressive to publish.

The trick with the first one is that the second answer is more evidence based than the first one. The first answer choice has nothing to do with the provided context. The second answer choice has something to do with the provided context: it is partially evidence based. Instead of taking the question really literally as to be about the mathematics of probability, they are deciding which answer makes more sense and saying that. The first one makes no sense (having nothing to do with the provided information). The second one partially makes sense, so they say it's better.

A more literally minded person would catch on to the trick. But so what? Why should people learn to split hairs so that they can give literally correct answers to bad and pointless questions? That's not a useful skill so most people don't learn it.

The trick with the second one is that the second answer is a better explanation. The first part provides a reason for the second part to happen. Claims that have explanatory reasons are better than claims that don't. People are helpfully reading "and" as expressing a relationship -- just as they would do if their friend asked them about the possibility of Russia invading Poland and the US suspending diplomacy. They think the two parts are relevant, and make sense together. With the first one, they don't see any good explanation offered so they reject the idea. Did it happen for no reason? Bad claim. Did it happen without an invasion of Poland or any other notable event worth mentioning? Bad claim.

Custom Search

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by Habryka on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by korin43 on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by Multipartite on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by trickster on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by adjuant on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

People are using valuable real life skills, such as looking for good explanations and trying to figure out what reasonable question people intend to ask, rather than splitting hairs. This is not a horrible bias about X&Y being more likely than Y. It's just common sense. All the conjunction fallacy research shows is that you can miscommunicate with people and then and then blame them for the misunderstanding you caused. If you speak in a way such that you can reasonably expect to be misunderstood, you can then say people are wrong for not giving correct answers to what you meant and failed to communicate to them.

The conjunction fallacy does not exist, as it claims to, for all X and all Y. That it does exist for specially chosen X, Y and context is incapable of reaching the stated conclusion that it exists for all X and Y. The research is wrong and biased. It should become less wrong by recanting.

This insight was created by philosophical thinking of the type explained in *"The Beginning of Infinity"* by David Deutsch. It was not created by empirical research, prediction, or Bayesian epistemology. It's one of many examples of how good philosophy leads to better results and helps us spot mistakes instead of making them. It also wasn't discovered by empirical research. As Deutsch explained, bad explanations can be rejected without testing, and testing them is pointless anyway (because they can just make ad hoc retreats to other bad explanations to avoid refutation by the data. Only good explanations can't do that.).

Please correct me if I'm wrong. Show me an unbiased study on this topic and I'll concede.

Article Navigation

Comments (113)

Tags: [popper](#) | [deutsch](#) | [conjunction](#) | [Delete](#)

Comments (113)

Sort By: New

Risto_Saarelma 11 April 2011 12:43:13PM 2 points [\[-\]](#)

I wonder what the downvote pattern here would be if people wouldn't see the existing downvotes in the -20 range along with the post, but would actually need to read the text and individually decide whether to downvote, upvote or abstain with no knowledge of the other votes.

What if the displayed karma, at least to anyone else other than the author, would be shown as just 0 or -1 if it was any amount into the negative? Then there wouldn't be an invitation to join the happy fun massive downvote pile-on party without even reading the post.

[deleted] 12 April 2011 10:21:45AM 2 points [\[-\]](#)

I read the post in the RSS feed, clicked through to downvote, and then *didn't* because I thought "s/he's suffered enough". Arguably, with a monumentally *wrong* post like the OP, we might see more people downvoting if they didn't know others had.

Jonathan_Graehl 11 April 2011 10:25:38PM 2 points [\[-\]](#)

Before reading the post, I skimmed the comments. I see that the author's replies to comments are clear evidence of failure. I don't know, or care, if it's trolling or merely defensiveness (I guess the latter is by far more likely).

As for the claim that the particular questions are designed to elicit wrong answers (and yes, the wrong answers are objectively wrong), so as to sex up the results, that's obviously true. That obviousness is probably the reason for the initial downvotes.

I think the author has completely misunderstood the intent of "conjunction fallacy". It's not that people are more prone to favor a conjunction as more plausible in general, just because it's a conjunction, it's that it's *obviously wrong* to consider a conjunction more likely than any one of its components.

[LW 2.0 Open Beta Live](#)
by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)
by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)
by Raemon | 41v (106c)

[European Community Weekend 2017](#)
by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)
by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)
by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)
by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)
by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)
by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)
by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all' by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- [gwern](#) (32)
- [ingres](#) (30)
- [Lumifer](#) (27)
- [turchin](#) (23)
- [Habryka](#) (22)
- [Yosarian2](#) (21)
- [IlyaShtitser](#) (19)
- [SaidAchmiz](#) (19)
- [gjm](#) (19)
- [fowlertm](#) (17)
- [SquirrelInHell](#) (15)
- [Dagon](#) (15)
- [Elo](#) (14)
- [Stuart_Armstrong](#) (13)
- [Gunnar_Zamcke](#) (12)

RECENT KARMA AWARDS

curi 11 April 2011 06:21:02PM 2 points 

I don't know about front page posts, but a lot of comments stop getting downvotes at -2 or -3. I think they turn invisible for some people by default, so that's why, rather than people thinking the point is already made.

On Hacker News, early voting trends usually continued to inspire more votes of the same kind. Except someones when someone would comment "Why is this getting downvoted" which could reverse it. They capped things to -4 points minimum now. Upvote orgies are still common. Some are due to quality, sure -- some things deserve a lot of upvotes -- but not all of them.

FeepingCreature 12 April 2011 03:15:29PM 0 points 

Similar thing happens on reddit. I think it's widespread across vote-based sites. Any counterexamples?

Risto_Saarelma 11 April 2011 07:35:51PM 1 point 

Hacker News seems to have a definite idea on curbing downvote sports by capping the point display when things go to negative. This might be a good idea. Another intentional thing they have is having no downvoting at all on the toplevel posts, but their posts are more links and short questions than original articles, so this might not fit LW that well.

Spurious upvoting is probably also going on, but I don't see it as problematic as downvotes since it doesn't come with a direct effect of anonymously telling people that they don't belong on the forum.

AdeleneDawner 11 April 2011 01:08:26PM 5 points 

My model predicts the opposite of what you seem to be suggesting, actually. Downvotes are a limited resource, even if the limit isn't a restrictive one, so I expect that at least some users will decline to downvote in order to conserve that resource if they see that the point has been made. Even without that particular pressure, it seems likely to me that some people just won't *bother* to downvote if it's already clear that something isn't approved of. The latter is a bit like the bystander effect - if one knows that someone else has already taken care of the problem, one won't feel much pressure to do so as well.

We're probably both right, in a sense - when this kind of question has come up before, there have tended to be a mix of responses, so both things are probably happening. I do think that on net we'd see more downvotes if the existing balance wasn't public until people voted, though, even if 'abstain' counted as a vote so that people who were just curious weren't forced to weigh in.

Risto_Saarelma 11 April 2011 01:22:20PM * 1 point 

Yeah, I don't actually have a very solid anticipation on what would happen, though I would be more surprised if the result was even more downvotes. Still, you do make a case why it might well be.

I'm not sure I'd mind a lot of people reading something and individually deciding to downvote it, that would just do what the score is originally supposed to and sum a lot of individual assessments. What I don't like is getting the pile-on effect from both showing the current score and the downvote options. It makes downvoting stuff already in the negative further feel icky. It also makes it feel a lot more symbolic to upvote something you see far in the negative or downvote something far in the positive.

And I didn't even know downvotes cost something. I see a mention of a cap on an old thread, but I don't see my downvote mana score anywhere in the UI.

AdeleneDawner 11 April 2011 02:20:28PM 2 points 

What I don't like is getting the pile-on effect from both showing the current score and the downvote options. It makes downvoting stuff already in the negative further feel icky. It also makes it feel a lot more symbolic to upvote something you see far in the negative or downvote something far in the positive.

I'm not sure I'm parsing this properly - it sounds to me like you're saying that seeing a very low score along with the option to lower it further makes you less likely to actually downvote, because it feels like ganging up on someone, and more likely to upvote it, compared to not seeing the existing score?

Does it change anything if you specifically focus on the fact that upvotes and downvotes are supposed to mean "I want to see more/less like this" rather than "I think this is good/bad" or "I think this is right/wrong" or "I like/dislike the point being made"?

And I didn't even know downvotes cost something. I see a mention of a cap on an old thread, but I don't see my downvote mana score anywhere in the UI.

The downvote limit isn't displayed anywhere, until you hit it. (And I'm not *sure* you get a proper notification then, either.) But the algorithm is that you can make four times as many downvotes as your karma score, and then you can't make any more downvotes until your karma goes up. It's not something that most people will ever run into, I suspect, but knowing that it's possible seems likely to have an effect - and possibly a stronger one given the uncertainty of the size of one's buffer, in fact.

Risto_Saarelma 11 April 2011 04:24:45PM 1 point

I'm not sure I'm parsing this properly - it sounds to me like you're saying that seeing a very low score along with the option to lower it further makes you less likely to actually downvote, because it feels like ganging up on someone, and more likely to upvote it, compared to not seeing the existing score?

Pretty much. And yeah, looks like I'm saying that I like behave the opposite way that I expect the most of the group to behave. But then again, I'm imagining both reacting more to a big, shiny sign saying "Look! This is what Us People are expected to think about this thing!" than the actual thing itself. I'm just generally reacting to anything with that sort of mindless, mob-like tone with "screw You People."

Does it change anything if you specifically focus on the fact that upvotes and downvotes are supposed to mean "I want to see more/less like this" rather than "I think this is good/bad" or "I think this is right/wrong" or "I like/dislike the point being made"?

It's easy enough to imagine that when I'm the one doing or receiving the voting. When I'm seeing this stuff unfold elsewhere, I need to imagine that the downvoters and the downvotees are also seeing things in that way. Given that this requires extra thought, and the simple, instant interpretation is upvote good, downvote bad, I'm not having a very easy time imagining that's the case.

Also given how human group psychology works, I'm not sure there's that much of a difference there. Saying that you'd like to see less of the sort of thing the author implicitly thought people are interested in seeing by posting it doesn't sound like that much less of a rebuke than just saying the thing was bad or wrong. You could probably make a case for it being a bigger rebuke, since it implies the author is too incompetent to figure out what's good content, and the social inferior whom the downvoters can pass judgment on, rather than someone with equal social standing having a dispute with the others on whether something is good and right or bad and wrong.

lessdazed 11 April 2011 04:15:38PM 1 point

Does it change anything if you specifically focus on the fact that upvotes and downvotes are supposed to mean "I want to see more/less like this" rather than "I think this is good/bad" or "I think this is right/wrong" or "I like/dislike the point being made"?

This is at most only true in one respect: many implementors of the system primarily had such intent.

["It doesn't even seem to have occurred to them that a voter might attach their own meaning to a vote for Stephen Colbert - that a voter might have their own thoughts about whether a vote for Stephen Colbert was "wasted" or not."](#)

Count me among those who saw downvoting this as pointless. It was already downvoted enough to be removed from the main page, and downvotes are limited. Also, all else equal, I feel bad when engaging in piling on. [This post](#) is what I think of as ideal: it responds to the OP by finding the limited case for which an argument resembling that of the OP would be right, so the OP can see the contrast without feeling attacked.

AdeleneDawner 11 April 2011 10:50:48AM * 0 points

Eliezer's post about [defying the data](#) (sometimes called 'defying the evidence') seems relevant here.

Maelin 11 April 2011 09:34:33AM 2 points

So I'm curious. Pretending for a moment that you have actually approached this situation rationally, you must have had some significant amount of evidence that humans do *not* commit any kind of conjunction fallacy (except when "tricked").

Given that you are so confident in this contention that you feel you can safely dismiss every one of the significant number of corroborating scientific studies a priori as "wrong and biased", I'm wondering where got this evidence that humans never implement faulty probability estimation techniques except when being tricked, and why nobody else seems to have seen it.

curi 11 April 2011 05:55:38PM -1 points

you must have had some significant amount of evidence that humans do not commit any kind of conjunction fallacy

Nope. I don't have any evidence you don't have. I don't think the disagreement is primarily about evidence but philosophical understanding and explanation.

Given that you are so confident in this contention that you feel you can safely dismiss every one of the significant number of corroborating scientific studies

I've looked at some. Found they weren't even close to meeting the proper standards of science. I've run into the same problem with other kinds of studies in various fields before. I have general purpose explanations about bad studies. They apply. And no one, on being challenged, has been able to point me to any study that is valid.

I'm wondering where got this evidence that humans never implement faulty probability estimation techniques except when being tricked

I have *explanations* that probability estimating isn't how people think in the first place, in general. See: Popperian epistemology.

nobody else seems to have seen it.

You mean, nobody else *here* seems to have seen it, on a site devoted to the idea of applying probability math to human thought.

Eliezer_Yudkowsky 11 April 2011 07:08:01AM 2 points

Post removed from main and discussion on grounds that I've never seen anything voted down that far before. Page will still be accessible to those who know the address.

Icelus 12 April 2011 07:06:51AM 3 points

It seems vote scores of EY posts about site-related actions and discussion turn into agree/disagree, so based on that the consensus is this was a good action.

But I would like to have a comment here at least voicing the opinion (my opinion on this coincidentally) that just displaying a low vote score is enough to warn people that don't want to read things that get low vote scores.

[deleted] 11 April 2011 11:00:54AM * 4 points

Outliers are interesting. If it was me, I'd be checking it out.

lessdazed 11 April 2011 04:32:46PM 6 points



I agree. "...on grounds that I've never seen anything voted down that far before..." isn't literally true in the strictest possible sense. Implicitly, it means the editor *also* did not think of any sufficiently compelling counter-arguments upon reading the piece.

So unless removal for being downvoted by orders of magnitude more than anything else (including those previously removed, to avoid [evaporative cooling](#)) was the rule before the editor removed the post, the content of the post had a good deal to do with its removal.

I am not saying it is the wrong decision because of this, but I would prefer the formulation: "Post removed from main and discussion on grounds that I've never seen anything voted down that far before, and it seems to me the downvotes are probably justified."

It's important not to imply the post's removal wasn't a matter of editorial judgement, since it was, [not that there's anything wrong with that](#).

Arguably no one will be misled, and the editorial explanation by EY is perfect and Gricean.

There's ["still something sort of awesome about a guy who could take on the entire western world from a cave somewhere"](#). I began writing but did not post three posts of my own for this site because I think them not good enough. One is about Newcomb's problem, one is about tattoos, and one is about cooking.

[deleted] 12 April 2011 12:12:00AM * 3 points



It's important not to imply the post's removal wasn't a matter of editorial judgement, since it was, not that there's anything wrong with that.

I have no issue with people making editorial judgements provided they are upfront about it. On the face of it, EY is saying he removed the post purely on the basis of popularity, as though he is deferring judgement to others. You assert he wasn't and that no-one would be misled but if you just leave things implied there is a lot of scope for misunderstanding and it is a way of not taking responsibility.

lessdazed 11 April 2011 05:48:27AM 14 points



The strangest thing about this post is how everything is anthropomorphised. It's as if what is at stake isn't determining a relatively isolated matter of fact, but rather ostracizing a social rival.

The conjunction fallacy says...

The conjunction fallacy does not exist, as it claims to...

The research is wrong and biased. It should become less wrong by recanting.

...bad explanations can be rejected without testing, and testing them is pointless anyway (because they can just make ad hoc retreats to other bad explanations to avoid refutation by the data. Only good explanations can't do that.).

Explanations are models, not social beings you can hope to subdue or isolate by making them unpopular.

Something that causes angst needn't be intending to do so. The frustration this idea causes you is in you, not an evil part of this idea. The facts in question are just the way the world is, not a kid in school who is to be slandered for being your rival, whom you can disempower by convincing others s/he is a very biased bad person.

A conjecture: in general, the chances of being right about something are inversely proportional to how closely your model for that thing matches your self-model, which is what you use by default when you don't understand the specific causal mechanisms at play. This post is an extreme example of that phenomenon.

If you're getting angry at patterns of human cognition for "claiming to exist" because they "really" "do not exist", you're simply doing it wrong.

curi 11 April 2011 07:30:40AM -1 points



I'm not angry. Why do you assume people you disagree with are angry?

What kind of words do you think should be used to describe the contents of theories? e.g. take

| The many worlds interpretation says there are multiple universes.

You think that's anthropomorphic and should be replaced by what sentence?

lessdazed 11 April 2011 09:54:13PM 3 points



| Why do you assume people you disagree with are angry?

This could be unpacked in several ways: "Why do you always assume people you disagree with are angry?", "Why do you sometimes but not always assume people you disagree with are angry?", and an equivocation which pretends to the advantages of each statement.

"Why do you always assume people you disagree with are angry?", would be silly for you to say after only having read my few LW posts. If true, it would be a severe reasoning flaw on my part.

"Why do you *sometimes but not always* assume people you disagree with are angry?", in turn has several possible meanings. "Assumption" could mean an arbitrary **prior** I have before considering the subject I am analyzing, or alternatively it could mean an **intermediate conclusion** that is both reasoned and a basis for other conclusions. It may also mean something I'm not thinking of, and like other words, it of course may be an equivocation in which the speaker seeks to mean different parts of coherent but incompatible definitions.

If "assumption" means "prior", your statement highlights that my assumptions may be silly for an additional reason than their being false, as read this way it would suggest that my reasoning is not just misaligned from the reality about which I am trying to reason, but that my reasoning isn't even consistent, with my reasoning dependent on which of several arbitrary reading assumptions I select without considering the external world. As an analogy, not only do I attempt to map cities around mine without looking at them, my maps contradict each other by allocating the same space to different cities in an impossible way. "Prior" is often the intended way to interpret the word "assumption", so a definition along these lines springs naturally to mind. However, you would have even less justification for this meaning than the previous one.

If "assumption" means "**intermediate conclusion** that is both reasoned and a basis for other conclusions", which it sometimes does, then your statement wouldn't be at all unflattering to me or relevant.

Something that would at first appear to be an equivocation might itself not be logically flawed, but its meaning might be difficult to track. For instance, one could say "Why do you think some people you disagree with are angry?" "Think" and "some" don't presuppose anything about the reason behind the thought. Knowing that I think you are angry, you would be correct to use a term that encompasses me either assuming or concluding that you are angry, as you don't know which it is.

"Why do you assume people you disagree with are angry?" depends on being interpreted as "Why do you sometimes think, whether as a prior or an intermediate conclusion, that some people you disagree with are angry?" to seem a grammatical question that does not make unreasonable assumptions. However, in the context of responding and trying to rebut a criticism of your thoughts, your statement only performs the work it is being called on to do if its answer casts doubt on my thinking.

"Why do you assume all people you disagree with are angry?", "Why do you arbitrarily assume some people you disagree with are angry?", and "How could you reasonably conclude all people you disagree with are angry?" are questions to which no one could give a reasonable explanation, taking for granted their assumptions. These are powerful rebuttals, if they can be validly stated without making any assumptions.

Though there is no explanation to them, there is an explanation for them, a response to them, and an answer to them, namely: you are making incorrect assumptions.

Here, your statement attempts to borrow plausibility from the true but irrelevant formulation without assumptions from which there is no conclusion that my thinking is flawed, while simultaneously relying on the context to cast it as a damning challenge if true - which a *different* sentence in fact would be. Your equivocation masks your assumptions from you.

As to why I think that...well, I started with the second paragraph, so we can start there.

It is based on bad pseudo-scientific research designed to prove that people are biased idiots. One of the intended implications, which the research does nothing to address...

(Pejorative) research you disagree with is designed to prove people are (pejorative). Supposed intended implications that you despise, which are so important they aren't at all addressed...which of course would mean that you have no evidence for them...and are confabulating malice both where and as an explanation behind what you do not fully understand or agree with.

curi 12 April 2011 02:57:28AM * 1 point

To clarify the part you were curious about:

You apparently sometimes assume people are angry on minimal evidence and without explaining your reasoning. You did it with me, and I don't think this is a personal grudge but something you might do with someone else in similar circumstances. In particular, I think you may have tried to judge my emotional state from the *on topic content* and perhaps also *style* of my writing. I understand that in some cases, with some people, such judgments are accurate. I don't think this is such a case, and I don't think you had any good evidence or good explanation to tell you otherwise.

I think that you had some kind of reason for thinking I was angry, but that you left it *unstated*, which shields it from critical analysis.

The way I read your statements, you assumed I was angry (based presumably on unstated reasoning) and drew some conclusions from that, rather than asserting I was angry and arguing for it. E.g. you suggested that I got angry and it was causing me to do it wrong. (BTW that looks to me like, by your standards, stronger evidence that you were angry at me than the evidence of my anger in the first place. It is pejorative. I suspect you actually meant well -- mostly -- but so did I.)

(Pejorative) research you disagree with is designed to prove people are (pejorative).

People can dislike things without being angry. Right? If you disagree with me about the substance of my criticisms, ad hominems against my emotions are not the proper response. Right?

katydee 11 April 2011 04:53:54AM 6 points

Why are you on this website? Not joking.

jsalvatier 11 April 2011 06:30:12PM * 5 points

"What is your goal here?", or something like it might have been less confrontational phrasing.

curi 11 April 2011 05:12:46AM -12 points

How will I benefit from telling you?

David_Allen 11 April 2011 09:33:27PM 3 points

By explaining your reasons for posting to this site you may get feedback suggesting how to better use this site to achieve your goals.

katydee 11 April 2011 08:19:34AM 7 points

You might, for instance, use this moment as an opportunity to signal that you aren't a troll or indicate why you shouldn't be banned, as I fear this fate awaits you if you continue on your present course.

CuSithBell 11 April 2011 05:22:56AM 2 points

That, of course, depends on your reason - but it could be that you will receive information on how better to

achieve your goals, indication that they are impossible or prohibitively costly to accomplish (thus saving you time and effort), the good will of those interacting with you, confirmation that you are acting effectively....

Your goals may conflict with this, however. For instance, if they are to reap discord and vitriol, or to test the community, or end civilization, or to hide your goals, or bring low the proud, or find a worthy sparring partner, you may not benefit from such transparency.

[deleted] 11 April 2011 05:28:41AM * 4 points [-]

Have a care. Remember absence of evidence is only weak evidence of absence, and that, as you have mentioned, there are many possible explanations.

Probability measure is spread pretty thin in these parts.

lessdazed 19 May 2011 01:22:38AM 0 points [-]

Probability measure is spread pretty thin in these parts.

Awesome. This needs to be one of the first lines a protagonist hears after entering a cantina in a science fiction Space Opera...maybe rationalist fanfiction like HPATMOR.

endoself 11 April 2011 06:19:38AM 4 points [-]

absence of evidence is only weak evidence of absence

Not always. If someone has no symptoms, the probability that they have, say, pneumonia drops considerably.

JoshuaZ 11 April 2011 04:50:26AM 1 point [-]

Others have addressed most of the issues but I'd like to address another issue.

Claims that have explanatory reasons are better than claims that don't.

This isn't at all relevant. In the Soviet-Poland experiment the two groups were asked to give a probability estimate. Whether they are better in some sense is a distinct claim about whether or not the probability of one or the other is higher. Even if one prefers claims with explanatory power to those that don't, this doesn't make those claims more probable. Indeed, the fact that they are easier to falsify in the Popper sense can be thought of as connected in some way to the fact that in a Bayesian sense they are more improbable.

But you can't argue that because the class of hypotheses is preferred people should be willing to assign a higher probability to them.

curi 11 April 2011 04:59:01AM -8 points [-]

"Probability estimate" is a technical term which most people don't know. That isn't a bias or a fault. When asked to give one, especially using a phrase like "which is more probable" which is ambiguous (often used in non-mathematical sense. whereas the term "probability estimate" isn't), they guess what you want and try to give that instead. Make sense so far?

Maelin 11 April 2011 05:18:57AM * 12 points [-]

But this is nullified in the colour die experiment of Tversky and Kahneman, as detailed in the [Conjunction Fallacy](#) article, as [linked](#) by HonoreDB in a comment in this very topic, as so far unaddressed by you.

Consider a regular six-sided die with four green faces and two red faces. The die will be rolled 20 times and the sequences of greens (G) and reds (R) will be recorded. You are asked to select one sequence, from a set of three, and you will win \$25 if the sequence you chose appears on successive rolls of the

die. Please check the sequence of greens and reds on which you prefer to bet.

1. RRRRR
2. GRRRR
3. RRRRR

65% *still* chose sequence 2 despite it being a conjunction of sequence 1 with another event. There was actual money on the line if the students chose a winning sequence. *There is no plausible way that the students could have misinterpreted this question because of ambiguous understandings of phrases involving "probability".*

I find your claim that people interpret "which is more probable" as asking something other than which is more probable to be dubious at best, but it completely fails to explain the above study result.

curi 11 April 2011 05:23:01AM * -4 points

You're calling lack of mathematical knowledge a bias? Before it was about how people think, now it's just ignorance of a concrete skill...?

Maelin 11 April 2011 05:48:02AM * 7 points

I'm not calling it anything, and I didn't use the word "bias" once. Perhaps people who have studied probability theory *would* have fared better on the task. But this is kind of the point, isn't it? The issue is that when a person intuitively tries to make a judgement on probabilities, their intuition gives them the wrong result, because it seems to use a heuristic based on representativeness rather than actual probability.

Nobody is saying that this heuristic isn't most likely more than adequate in the majority of normal day-to-day life. You seem to be conflating the claim that *people use some kind of representativeness heuristic, rather than correct probability analysis, in probability estimation problems* - which appears to be true, see above - with some kind of strawman claim you have invented that *people are stupid idiots whose probability estimation faculties are so bad they reliably get every single probability estimation problem wrong every time, lolmorons*. People here are posting in defense of the first one, you keep attacking as if they are defending the second one.

The hypothesis

a representativeness heuristic was easier for evolution to implement than a full, correct model of probability theory, and it did the job well enough to be a worthwhile substitute, thus humans will incorrectly choose representative cases rather than more probable cases when *attempting* to choose more probable cases

predicts the results of all the studies strongly. Your contrary hypothesis, which to me appears to be along the lines of

humans DO implement correct probability (i.e. they do not implement it incorrectly) but psychologists are determined to make humans look stupid in order to laugh at them, thus they craft experiments that exploit ambiguity over "probability" to make it look like humans use a representativeness heuristic when they in fact do not

predicts the results of some of the studies, but not the red/green die rolling one.

So now we are waiting for your post hoc justification of how test subjects who *knew* they were seeking a strictly more probable case (because they were hoping to win money) still got the wrong answer.

curi 11 April 2011 05:58:04AM * -4 points

The issue is that when a person intuitively tries to make a judgement on probabilities

Not intuitively. They're in a situation outside their daily routine that they *don't* have good intuition about. They are making a judgment not because they think they know what they are doing but because they were asked to.

Can you see how artificial situations are not representative of people's ability to cope with life? How

putting people in artificial situations designed to fool them is itself a biased technique which lends itself to reaching a particular conclusion?

Psychohistorian 11 April 2011 01:03:33PM 2 points 

Your objection seems to boil down to, "Experimental settings are not identical to actual settings, therefore, everything derived from them is useless."

It seems pretty unlikely that people will miscalculate probability in an experiment with money on the line, but then always calculate it with perfect accuracy when they are not specifically in an experimental setting.

In short *how* is the experimental setting so *different* that we should completely ignore experimental results? If you have a detailed argument for that, then you'd actually be making a point.

curi 11 April 2011 07:36:11PM -1 points 

I think, in general, they don't calculate it at all.

They don't do it in normal life, and they don't do it in most of the studies either.

In short how is the experimental setting so different that we should completely ignore experimental results?

It's different in that it's *specifically designed* to elicit this mistake. It's designed to trick people. How do I know this? Well apart from the various arguments I've given, they *said* so in their paper.

Psychohistorian 11 April 2011 07:48:59PM 3 points 

It's designed to trick people.

So, your position is that it is completely unrealistic to try to trick people, because that *never happens* in real life? Really? Or is it a moral condemnation of the researchers. They're bad people, so we shouldn't acknowledge their results?

People try to trick each other all the time. People try to trick *themselves* all the time. People routinely make informal estimates of what's likely to happen. These experiments all show that *under certain circumstances*, people will systematically be wrong. Not one thing you've said goes to counter this. Your entire argument appears predicated on the confusion of "all" with "some."

curi 11 April 2011 08:08:22PM -2 points 

If you succeed at tricking people, you can get them to make mistakes.

What those mistakes are is an interesting issue. There is no argument the mistakes actually have anything to do with the conjunction fallacy. They were simply designed to look like they do.

[continue this thread »](#)

[deleted] 11 April 2011 06:09:25AM * 6 points 

How putting people in artificial situations **designed to fool them** is itself a biased technique which lends itself to reaching a particular conclusion? (emphasis mine)

I understand this criticism when applied to the Linda experiment, but not when it is applied to the color experiment. There was no "trickery" here.

To avoid more post-hockery, please propose a concrete experiment testing the conjunction fallacy,

but controlling for "tricks" inherent in the design.

curi 11 April 2011 06:18:34AM * -4 points [\[-\]](#)

Not all the experiments are the same thing, just because they reached the same conclusion.

The color experiment is different: it's about asking people to do something they don't know how to do, and then somehow interpreting that as a bias.

The paper doesn't describe the test conditions (that makes it unscholarly. scientific papers are supposed to have things like "possible sources of error" sections and describe their experimental procedure carefully. It doesn't even say if it was double blind. Presumably not. If not, an explanation of why it doesn't need to be is required but not provided.). There's various issues there, e.g. pressure not to ask questions could easily have been present.

Setting that aside, it's putting people in an artificial situation that they don't understand well and getting them to make a guess at the answer. This doesn't simulate real life well. It has a consistent bias: people are better at real life than artificial situations designed in such a way that people will fail.

EDIT: Look at this sentence in the 1983 paper:

Our problems, of course, were constructed to elicit conjunction errors, and they do not provide an unbiased estimate of the prevalence of these errors.

See. Biased. Specifically designed to fool people. They directly admit it. You never read the paper, right? You just heard a summary of the conclusions and trusted that they had used the methods of science, which they had not.

AdeleneDawner 11 April 2011 07:01:36AM 4 points [\[-\]](#)

Our problems, of course, were constructed to elicit conjunction errors, and they do not provide an unbiased estimate of the prevalence of these errors.

See. Biased. Specifically designed to fool people.

What that says is that the test in question addresses only one of two obvious questions about the conjunction fallacy. It addresses the question of whether that kind of mistake is one that people do sometimes make; it does not address the question of how often people make that kind of mistake in practice. It's like if they were trying to find out if overdosing on a certain vitamin can be fatal: They'll give their (animal) subjects some absurd amount of it, and see what happens, and if that turns out to be fatal, *then* they go trying different amounts to see where the effect starts. (That way they know to look at, say, kidney function, if that's what killed the first batch of subjects, rather than having to test all the different body systems to find out what's going on.) All the first test tells you is whether that kind of overdose is possible at all.

Just because it doesn't answer every question you'd like it to doesn't mean that it doesn't answer any questions.

curi 11 April 2011 07:04:52AM * -3 points [\[-\]](#)

Do you understand that the situation of "someone using his intelligence to try to fool you" and the situation "living life" are different? Studies about the former do not give results about the latter. The only valid conclusion from this study is "people can sometimes be fooled, on purpose". But that isn't the conclusion it claims to support. Being tricked by people intentionally is different than being inherently biased.

[continue this thread »](#)

[deleted] 11 April 2011 06:33:23AM * 3 points [\[-\]](#)

The color experiment is different: it's about asking people to do something they don't know how to do, and then somehow interpreting that as a bias.

Setting that aside, it's putting people in an artificial situation that they don't understand well and getting them to make a guess at the answer. This doesn't simulate real life well.

This is exhausting. Whatever this "real life" of yours is, it's so boring, predictable, and uniform as to be not worth living. Who hasn't had to adapt to new (yes, even "artificial") situations before? Who hasn't played a game? Who hasn't bet on the outcome of an event?

Propose an experiment already. My patience is waning.

EDIT: I have, in fact, read the 1983 paper.

curi 11 April 2011 06:38:48AM -3 points

You think I have to propose a better experiment? We should just believe the best experiment anyone did, even if it's no good?

[continue this thread »](#)

AdeleneDawner 11 April 2011 06:06:23AM 7 points

They're in a situation outside their daily routine that they don't have good intuition about. They are making a judgment not because they think they know what they are doing but because they were asked to.

This is not exactly an uncommon situation in day-to-day life.

[deleted] 11 April 2011 05:27:31AM 1 point

It is unclear how this relates to grandparent's point.

JoshuaZ 11 April 2011 05:08:54AM 6 points

No. Absolutely not. First of all, the Soviet experiment was done (and has been replicated in other versions) with educated experts. Second of all, it isn't very technical at all to ask for a probability. People learn these in grade school. How do you think you can begin to talk to the Bayesians when you think probability estimate is a technical term? They use far more difficult math than that as parts of their basic apparatus. Incidentally, this hypothesis also is ruled out by the follow-up studies which have already been linked to in this thread that show that the conjunction fallacy shows up when people are trying to bet.

I have to wonder if your fanatical approach to Popper is causing problems in which you think or act that you think that any criticism, no matter how weak, or ill-informed allows the rejection of a claim until someone responds to that specific criticism. This is not a healthy attitude if one is interested in an exchange of ideas.

jimrandomh 11 April 2011 03:03:10AM 22 points

Curi, you are just confused. But rather than asking questions or attempting to resolve your own confusion, you've gone and loudly, falsely accused some good research of being shoddy. You have *got* to learn to notice when you're confused. You are trying to deal with confusing topics, and you have approached them with insufficient humility. You have treated criticisms as foes to be fought, and your own ignorance as a castle to be defended. Until you can stop yourself from doing that, you will make no progress, and we cannot help you.

shokwave 11 April 2011 02:52:07AM * 6 points 

Why should people learn to split hairs so that they can give literally correct answers to bad and pointless questions?

Nature will sometimes ask bad and pointless questions and reward anything less than a literally correct answer with grievous injury or death.

Thus it is said that taking humans as your opponents only leads to overconfidence. Nature will not be fair in the challenges she puts before you.

curi 11 April 2011 04:33:49AM -6 points 

We don't need everyone to have that skill.

shokwave 11 April 2011 11:29:17PM 1 point 

Actually, we *really* do, given things like democracy and politics.

[deleted] 11 April 2011 02:51:35AM 12 points 

In a way, this post is correct: the conjunction fallacy is not a cognitive bias. That is, thinking "X&Y is more likely than X" is not the actual mistake that people are making. The mistake people make, in the instances pointed out here, is that they compare probability using the representativeness heuristic -- asking how much a hypothesis resembles existing data -- rather than doing a Bayesian calculation.

However, we can't directly test whether people are subject to this bias, because we don't know how people arrive at their conclusions. The studies you point out try to assess the effect of the bias indirectly: they construct specific instances in which the representativeness heuristic would lead test subjects to the wrong answer, and observe that the subjects do actually get that wrong answer. This isn't an ideal method, because there could be other explanations, but by conducting different such studies we can show that those other explanations are less likely.

But how do we construct situations with an obviously wrong answer? One way is to find a situation in which the representativeness heuristic leads to an answer with a clear logical flaw -- a mathematical flaw. One such flaw is to assign a probability to the event X & Y greater than the probability of X. This *logical error* is known as the conjunction fallacy.

(Another solution to this problem, incidentally, is to frame experiments in terms of bets. Then, the wrong answer is clearly wrong because it results in the test subject losing money. Such experiments have also been done for the representativeness heuristic)

nhamann 11 April 2011 02:50:43AM * 8 points 

This post that you have excreted has essentially zero content. You restate the core idea behind the representativeness heuristic repeatedly, and baldly assert that there are good reasons for people having the intuitions that they do, that people are "using valuable real life skills" when they give incorrect answers to questions. No ones arguing that it hasn't been an evolutionarily useful heuristic, just that it happens to be incorrect from time to time. I cannot figure out where in your post you actually made an argument that the conjunction fallacy "doesn't exist", and I am overjoyed that you no longer have the karma to make top-level posts.

Please stop posting and read the sequences.

HonoreDB 11 April 2011 01:49:26AM 12 points 

Check out the second example here, where the subjects bet with real money in a way demonstrating the bias.

Dustin 11 April 2011 01:02:49AM * 10 points 

I like this post because it exposes some of my own faulty thinking.

1. I find myself less likely to look in to what David Deutsch says about anything because of the manner you present your ideas. This isn't fair, maybe you're misrepresenting what he actually says. Maybe he explains it in a less grating fashion. Whatever his actual thoughts, it's not fair to him that I'm discounting what he has to say...and yet, I can't help it!
2. Very similar to point #1, I had a very difficult time reading all of your post because of things like:

This is false and misleading. It is based on bad pseudo-scientific research designed to prove that people are biased idiots.

... To me, the language of this sentence seems combative and off-putting, and in fact much of the post does. I wasn't able to put much thought into what you had to say because of this. The attitude conveyed seems to be "Haha, you guys are so dumb for believing the conjunction fallacy exists!".

I should be able to read past the attitude it felt like I was getting from your writing and understand the points you were trying to make...but I confess it was very difficult to do so!

1. And finally (this is supposed to be item #3, but the software makes it a #1...)

The conjunction fallacy does not exist, as it claims to, for *all X and all Y*.

Does it? I was never under the impression that this was what the conjunction fallacy showed. I may be wrong, but it doesn't matter to my broken mind. This was another strike against your post, and I wish it hadn't made it more difficult to read the rest of it. You could have expanded on why you felt the conjunction fallacy made such a sweeping generalization, but you didn't and it made it harder for me to read.

[deleted] 12 April 2011 10:26:30AM 2 points

Deutsch's *The Fabric Of Reality*, the only book of his I've read, is very well-written and has much that LessWrongers would agree with or find interesting, though I disagree with some of it. The version of Deutsch's ideas that curi is presenting seems to me utterly wrong-headed. This suggests to me that either Deutsch has suffered some catastrophic failure of reasoning (possibly due to severe brain injury) in the intervening decade-and-a-half, or that curi is misrepresenting his ideas.

curi 14 April 2011 05:03:48AM * -3 points

Any specific details? Why do you think I'm misrepresenting his ideas? Do you know anything about my relationship with him? Or about how much I've studied his ideas, compared to how much you have? Can you point out anything I said which contradicts anything from FoR?

It seems to me you are claiming the authority of having read his book, to know what he thinks. In this regard you are outclassed. And as to rational argument ... you don't give any.

curi 11 April 2011 01:28:45AM -15 points

One reason I don't edit out "combative" language is I don't interpret it as combative or emotional. It is the literal truth -- at least my best guess at such. I don't read things in the emotional way that some people do, and I'm not other-people-oriented, and I don't want to be, so it's hard to remember the parochial and arbitrary rules others want me to follow to sound nice to them. And I don't mind if the people too biased to read past style issues ignore me -- I consider it a benefit that they will self-select themselves for me like that.

I don't want to or like to spend my mental effort remember other people's biases in order to circumvent them. And I'm content with the results. The results I've gotten here are not representative (that is, I don't offend everyone. I have control over that). And the assumptions you're probably making about my goals here (required to judge if I got what I wanted, or not) are not correct either.

Does it?

It's supposed to be a *universal* issue (which happens some significant portion of the time, not every time). That's the only way it can demonstrate that it has anything to do with X&Y vs Y. If it's merely tricks of the type I describe, the X&Y vs Y part is irrelevant and arbitrary; random things of the form X&Y vs Y in general won't work for it, and you could easily do it with another logical form.

I find myself less likely to look in to what David Deutsch says about anything because of the manner you present your ideas. This isn't fair, maybe you're misrepresenting what he actually says. Maybe he explains it in a less grating fashion.

If you read his books you will find the style less grating.

Psychohistorian 11 April 2011 12:58:38PM 4 points



This basically reads as, "The point of writing is for other people to divine what's inside my head. There's no point in me actually trying to communicate this clearly. I'll write whatever I feel like and the burden is on them to figure it all out."

Using lots of conclusory and inflammatory language makes your writing vastly less compelling. It's not a far cry from saying, "Everyone who disagrees with me is a doo-doo head!" Writing to convince people has rules and forms you need to follow, because if you don't, you will fail at convincing people. If your goal is merely the satisfaction of your own ego, I suppose this isn't a problem.

In other words, if you have no desire to actually communicate successfully or persuade people, by all means, keep doing what you're doing.

Dustin 11 April 2011 01:44:23AM 8 points



And I don't mind if the people too biased to read past style issues ignore me -- I consider it a benefit that they will self-select themselves for me like that.

Do you feel as if *anyone* who reads Less Wrong is getting any benefit out of your post?

And the assumptions you're probably making about my goals here (required to judge if I got what I wanted, or not) are not correct either.

I don't think I'm making any assumptions about this. What are your goals? I can think of several off the top of my head:

1. Convince others of your ideas.
2. Get feedback on your ideas to help you further refine them.
3. Perform a social experiment on LW (use combative tone, see results)

Goal #3 and related "trollish" goals are the only ones I can think of off the top of my head that benefit from a combative tone.

I mean...why write at all if you don't want most people to look at the actual content of your writing?

These aren't attacks in question form. These are honest questions.

curi 11 April 2011 02:36:43AM -12 points



Do you feel as if anyone who reads Less Wrong is getting any benefit out of your post?

Haven't you noticed some people said positive things in comments?

You say you aren't making "any assumptions" but then you give some. As an example, I'm not here to do (1) -- if it happens, cool, but it's not my goal. That was, apparently, your assumption.

I mean...why write at all if you don't want most people to look at the actual content of your writing?

The halfway sane people can read the content. I haven't actually said anything I shouldn't have. I haven't actually said anything combative, except in a few particular instances where it was intentional (none in this discussion topic). I've only said things that I'm vaguely aware offend some silly cultural biases. The reason I said "biased idiots" is not to fight with anyone but because it *clearly and accurately* expresses the *idea* I had in mind (and also because the concept of bias is a topic of interest here). There is a serious disrespect for people as rational beings prevalent in the culture here.

Do people really get offended because of language? Maybe sometimes. I think most of it is not because of that. It's the substance. I think the attitude behind the conjunction fallacy, and the studies, and various

other things popular here, is anti-human. It's treating humans like dirt, like idiots, like animals. That's false and grossly immoral. I wonder how many people are now *more* offended due to the clarification, and how many less.

These aren't attacks in question form. These are honest questions.

That's what I thought. And if I didn't, saying it would change nothing because asserting it is not an argument.

Unless, perhaps, it hadn't even occurred to me at all. I think there's a clash of world views and I'm not interested in changing mine *for the purpose of alleviating disagreement*. e.g. my writing in general is focussed on better people than the ones who it wouldn't occur to to consider that questions weren't attacks. They are the target audience I care about, usually. You're so used to dealing with bad people that they occupy attention in your mind. I usually don't want them to get attention in mind. There's better people in the world, and better things to think about.

wnoise 11 April 2011 03:51:02AM * 10 points

There is a serious disrespect for people as rational beings prevalent in the culture here.

People aren't completely rational beings. Pretending they are is showing more disrespect than acknowledging that we have flaws.

It's treating humans like dirt, like idiots,

Not like dirt. Not like idiots. As though we sometimes act as idiots, yes. Because we sometimes do. You seem to have trouble confusing "always" and "sometimes".

like animals.

We are, in fact, animals. We're a type of ape that has a very big brain, as primates go. We have many differences from the rest of the animals, but the similarities with other animals should be clear enough that it would be a severe mistake to not call us animals.

That's false and grossly immoral.

Ah, now this is a very honest and revealing statement. These are two separate issues, of course. Statements can be true or false, and actions can be grossly immoral, and not the reverse. Yet they're linked in your mind. Why is that? Did you decide these statements(*) were false and thus holding them grossly immoral (or to be charitable, likely to lead to grossly immoral actions), or were you offended at the statements and thus decided they must be false?

I wonder how many people are now more offended due to the clarification, and how many less.

Not offended. Just saddened.

(*) Rather, your misunderstanding of them. "People can think $P(X\&Y) > P(X)$ " is not the same as "People always think $P(X\&Y) > P(X)$ for all X and Y". Yes, of course special X and Y have to be selected to demonstrate this. There is a wide range between "humans aren't perfect" and "humans are idiots". The point of studies like this is not to assert humans are idiots, or bad at reasoning, or worthless. The point is to find out where and when normal human reasoning breaks down. Not doing these studies doesn't change the fact that humans aren't perfectly rational, it merely hides the exact contours of when our reasoning breaks down.

curi 11 April 2011 03:53:43AM -12 points

Ah, now this is a very honest and revealing statement. These are two separate issues, of course.

They are not separate issues. The reason for the false statements is the moral agenda.

wnoise 11 April 2011 09:18:19AM 2 points

The possible (asserted, not shown) existence of an agenda might indeed have something to do with what studies were done and how the interpreters presented them. The morality or immorality

of this agenda is what is a separate issue.

[deleted] 11 April 2011 03:28:55AM 5 points

I think there's a clash of world views and I'm not interested in changing mine for the purpose of alleviating disagreement.

Are you interested in changing your worldview if [it happens to be incorrect](#)?

They are the target audience I care about, usually. You're so used to dealing with bad people that they occupy attention in your mind. I usually don't want them to get attention in mind. There's better people in the world, and better things to think about.

That's a good way to get caught in an [affective death spiral](#).

curi 11 April 2011 03:40:56AM 2 points

Are you interested in changing your worldview if it happens to be incorrect?

Yes of course.

Dreaded_Anomaly 11 April 2011 03:15:26AM 8 points

I think the attitude behind the conjunction fallacy, and the studies, and various other things popular here, is anti-human. It's treating humans like dirt, like idiots, like animals. That's false and grossly immoral.

This seems terribly backwards to me. In order to become more rational, to think *better*, we must examine our built-in biases. We don't do and use such studies simply to point out the results and laugh at "idiots"; we must first understand the problems we face, the places where our brains don't work well, so that we can *fix them* and improve ourselves. Utilizing [the lens that sees its flaws](#) is inherently pro-human. (Note that "pro-human" should be taken to mean something reasonable like "wanting humans to be the best that they can be," which does involve admitting that humans aren't at that point right now. That is only "anti-human" in the sense that Americans who want to improve their country are "anti-America.")

curi 11 April 2011 04:13:17AM -5 points

This seems terribly backwards to me. In order to become more rational, to think better, we must examine our built-in biases.

I wasn't talking about examining biases that exist. I was talking about making up biases, and finding ways to claim the authority of science for them.

These studies, and others, are wrong (as I explained in my post). And not randomly. They don't reach random conclusions or make random errors. All the conclusions fit a particular agenda which has a low opinion of humans. (Low by what standard? The average man-on-the-street American Christian, for example. These studies have a significantly lower opinion of humans than our culture in general does.)

We don't do and use such studies simply to point out the results and laugh at "idiots";

Can you understand that I didn't say anything about laughing? And that replies like this are not productive, rational discussion.

People who demean *all* humans are demeaning themselves too. Who are they to laugh at?

Utilizing the lens that sees its flaws is inherently pro-human.

Do you understand that I didn't say that the "lens that sees its flaws" was the anti-human?

we must first understand the problems we face, the places where our brains don't work well,

Can you understand that the notion that our brains don't work well, in some places, is itself a substantive assertion (for which you provide no argument, and for which these bad studies attempt to provide authority)?

Can you see that this claim is in the direction of thinking humans are a less awesome thing than they would otherwise be?

I'll stop here for now. If you can handle these basic introductory issues, I'll move on to some further explanation of what I was talking about. If you, for example, try to interpret these statements as my substantive arguments on the topic, and complain that they aren't very good arguments for my initial claim, then I will stop speaking with you.

Dreaded_Anomaly 11 April 2011 07:39:16PM 2 points

These studies, and others, are wrong (as I explained in my post). And not randomly. They don't reach random conclusions or make random errors. All the conclusions fit a particular agenda which has a low opinion of humans. (Low by what standard? The average man-on-the-street American Christian, for example. These studies have a significantly lower opinion of humans than our culture in general does.)

Your post only served to highlight your own misunderstanding of the topic. From this, you've proceeded to imply an "agenda" behind various research and conclusions that you deem faulty. To be blunt: you sound like a conspiracy theorist at this point.

Can you understand that the notion that our brains don't work well, in some places, is itself a substantive assertion (for which you provide no argument, and for which these bad studies attempt to provide authority)?

<http://wiki.lesswrong.com/wiki/Bias>.

Also, at this point it sounds like you're throwing out all studies of cognitive bias based on your problems with one specific area.

Can you see that this claim is in the direction of thinking humans are a less awesome thing than they would otherwise be?

Having a low opinion of default human mental capabilities does not necessarily extend to having a low opinion of humans in general. Humans are not static. Having an undeservedly high opinion of default human mental capabilities is a barrier to self-improvement.

benelliott 11 April 2011 06:48:42PM 1 point

Can you understand this Paul Graham [essay](#)

And see where on it your above comment falls?

curi 11 April 2011 06:58:16PM * -2 points

I'm familiar with it. Why don't you say which issue you had in mind.

Do you understand that my goal wasn't to be as convincing as possible, and to make it as easy as possible for the person I was talking to?

benelliott 11 April 2011 07:04:11PM * 1 point

You constantly attack trivial side-points, and peculiarities of wording, putting you somewhere in the DH3-5 region, not bad but below what we are aiming for on this site, which is DH6 or preferably Black Belt Bayesian's new DH7 (deliberately improving you opponents arguments for them, refuting not only the argument but the strongest thing which can be constructed from its corpse).

Can you understand that I didn't say anything about laughing? And that replies like this are not productive, rational discussion.

This is a good example. Yes, he used the word laugh and you didn't, but that is not important to the point he was making, nor does it significantly change the meaning or the strength of the argument to remove it. If wish to reach the higher levels of the DH you should try to ignore things like that and refute the main point.

Do you understand that I didn't say that the "lens that sees its flaws" was the anti-human?

This is another example. You never used the phrase 'lens that sees its flaws' but since the heuristics and biases program is absolutely central to the lens that sees its flaws and you accused that of being anti-human his central point still stands.

curi 11 April 2011 07:25:57PM * -1 points



Side-points are not trivial, and they are not unimportant, nor irrelevant (some can sometimes be irrelevant. These aren't, in my view.)

Every mistake matters. Learning is hard. Understanding people very different than you is hard. Agreeing about which points are trivial, with people who see the world differently, is hard to.

To deal with the huge difficulty of things like learning, changing minds, making progress, it's important to do things like *respect every mistake* and try to fix them all, not dismiss some as too small and gloss them over. Starting small is the best approach; only gradual progress works; trying to address the big issues straight away does not work.

Black Belt Bayesian's new DH7

New? Do you really think that's a new idea? Popper talked about it long ago.

Yes, he used the word laugh and you didn't, but that is not important

But I regarded it as important in several ways.

nor does it significantly change the meaning or the strength of the argument to remove it.

But 1) I think it does 2) I think that by rewriting it without the mistake he might learn something; he might find out it does matter; we might have a little more common ground. I don't think *skipping steps* like this is wise.

[continue this thread »](#)

Cyan 11 April 2011 04:29:58AM 12 points



... I will stop speaking with you.

Please, please, please stop speaking with us. Uncle! Uncle!

Dustin 11 April 2011 02:58:42AM 1 point



You say you aren't making "any assumptions" but then you give some. As an example, I'm not here to do (1) -- if it happens, cool, but it's not my goal. That was, apparently, your assumption.

I didn't say I was making any of those assumptions. In fact, I specifically said I wasn't assuming any of them.

curi 11 April 2011 03:04:39AM -4 points



You didn't say that. Apparently you meant you were stating *possible*, not *actual* goals. But you didn't say that either. Your bad. Agreed?

Dustin 11 April 2011 03:49:08AM * 0 points

| Agreed?

No, but I don't see any point in continuing this.

| I don't think I'm making any assumptions about this.

wedrifid 11 April 2011 02:10:04AM 3 points

| Do you feel as if anyone who reads Less Wrong is getting any benefit out of your post?

I personally enjoyed confirming that this thread sent curi below the threshold at which anyone who is not a prominent SIAI donor can make posts. Does that count?

Unnamed 11 April 2011 12:50:50AM * 14 points

I think that the between-subjects versions of the conjunction fallacy studies actually provide stronger evidence that people are making the mistake. For instance*, one group of subjects is given the description of Linda and asked to rank the following statements in order, from the one that is most likely to be true of Linda to the one that is least likely to be true of her:

- Linda is a teacher in elementary school
- Linda works in a bookstore and takes Yoga classes
- Linda is a psychiatric social worker
- Linda is a member of the League of Women voters
- Linda is a bank teller and is active in the feminist movement
- Linda is an insurance salesperson

A second group of subjects does the same task, but statement (e) is replaced by "Linda is a bank teller."

In this version of the experiment, there should be no ambiguity about what is being asked for, or about the meaning of the words "likely" or "and" - the subjects can understand them, just as statisticians and logicians do. But 'feminist bank teller' still gets judged as more likely than 'bank teller' relative to the alternatives; that is, subjects in the first group give statement (e) an earlier (more likely) ranking than subjects in the second group.

Some people don't like the between-subjects design because it's less dramatic, and you can't point to any one subject and say "that guy committed the conjunction fallacy", but I think it actually provides more clear-cut evidence. Subjects who saw a conjunctive statement about Linda being a bank teller and a feminist considered it more likely than they would have if the statement had not been a conjunction (and had merely included "bank teller"). This isn't a trick, it's an error that people will consistently make - overestimating the likelihood that a very specific or rare statement applies to a situation, even considering it more likely than they would have considered a broader statement which encompasses it, because it sounds like it fits the situation.

* This study design has been used, with these results, I think in Kahneman and Tversky's original paper on representativeness. But I didn't take the time to find the paper and reproduce the exact materials used, I just googled and took the 6 statements from [here](#).

Gray 11 April 2011 05:22:13AM * 4 points

Looking at some of this, I wonder if people are biased towards thinking that the more concrete statement is more likely? Somehow, in my mind, "feminist" is more abstract than "feminist book keeper". The latter seems closer to being a person, whereas the former seems to be closer to a concept. The more descriptive you are about a subject, the more concrete it sounds, and thus the more likely it is, because it sounds closer to reality. The less

descriptive, the more abstract it sounds, and therefore the less likely it is, because it sounds more hypothetical or "theoretical".

Of course, the more descriptive account is going to have more conjunctions, and therefore has lesser or the same probability. I just wonder if this has been taken into account.

curi 11 April 2011 01:17:38AM -12 points

Note for example the contextual hints in

d) Linda is a member of the League of Women

which bias it.

Psychohistorian 11 April 2011 12:50:42AM * 9 points

The point of the conjunction fallacy is that, under specific circumstances, people's ability to estimate probability will reliably misfire, i.e. people are biased.

This does not require that people's estimates *always* misfire. You seem to have some odd personal stake in this; the conjunction fallacy is not some great criticism of people's moral worth. It is merely an observation that the human brain does not function optimally, much like the existence of blind spots.

I can think of two ways this article represents a coherent point: first, it is pointing out that we do not always miscalculate. This is an obvious fact and does not conflict with the bias in question.

Second, you might be complaining that the circumstances under which people exhibit this bias are so unlike reality as to be irrelevant. You provide no evidence and virtually no argument to this effect. Indeed, people make off-the-cuff estimates without consulting the laws of probability all the time. If your point is that it is of no practical consequence, you did not get anywhere near demonstrating that.

Also, you said that a study is unbiased only if it confirms your view, then asked for unbiased studies contradicting your view. I hope the problem there is now apparent to you.

neq1 11 April 2011 12:31:06AM * 8 points

The first one is flawed, IMO, but not for the reason you gave (and I wouldn't call it a 'trick'). The study design is flawed. They should not ask everyone "which is more probable?" People might just assume that the first choice, "Linda is a bank teller" really means "Linda is a bank teller and not active in the feminist movement" (otherwise the second answer would be a subset of the first, which would be highly unusual for a multiple choice survey).

The Soviet Union study has a better design, where people are randomized and only see one option and are asked how probable it is.

HughRistik 11 April 2011 05:32:02AM * 5 points

Yes. The bank teller example is probably flawed for that reason. When real people talk to each other, they obey the cooperative principle of [Grice](#) or flout it in obvious ways. A cooperative speaker would ask whether Linda was a bank teller who was active in the feminist movement, or if Linda was bank teller *regardless* of whether she was active in the feminist movement (showing that one answer included the other).

Eliezer addresses this possibility in [Conjunction Controversy](#):

For one thing, dear readers, I offer the observation that most bank tellers, even the ones who participated in anti-nuclear demonstrations in college, are probably not active in the feminist movement.

Yet Eliezer is strangely tossing out a lot of information. We don't just know that she is a bank teller and anti-nuclear, we also know: "She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice."

At least where I went to college, the conditional probability that someone was a feminist given that they were against nuclear power, majored in philosophy, and were concerned with discrimination and social justice was

probably pretty damn high.

Owen_Richardson 10 April 2011 11:18:43PM * 11 points

I think it's worthwhile that you brought this up (and I'd love to see a discussion on the same issue of questionable phrasing in the Wason selection test, where I suspect your criticism might be valid), but I don't think you're right.

I mean, yes, most people would get the [cancer cure/cancer cure+flood right]. It's easy. But the example of Linda the [bank teller/bank teller+active feminist] *does* have *exactly the same logical structure*. And people get it wrong because of the way they're used to using language without thinking very hard in this context, yes.

Kinda like the way that people can fall for "if all X's are A, B, and C, then anything that's A, B, and C must be an X," but not "if all dalmations have spots, then anything with spots is a dalmation". Cuz we've got this screwy verb "to be" that sometimes means 'is identical with', sometimes 'is completely within the set of', sometimes 'is one example of something with this property' etc etc, and we have to rely on concrete examples oftentimes in order to figure out which applies in a particular context.

And maybe this is largely genetic, or maybe humans would reliably come to see the correct answers as obvious if they were raised in an environment where people talked about these issues a lot in a language that was clear about it. I dunno.

But I *do* know that I have run in to problems following the same pattern as the Linda the [bank teller/bank tell+active feminist] "in the wild", and made mistakes because of them. I have created explanations and plans with lots of burdensome details and tricked myself into thinking that made them *more* likely to be right, *more* likely to succeed. And I've been wrong and failed because of that.

And I don't think I'm unusual in this.

And I *do* know that I've avoided making this mistake at least several times after I learned about these studies and understood the explanation for why it's wrong to rate "linda is a bank teller and an active feminist" as more likely.

And I don't think I'd be unusual in that.

So I think the conjunction fallacy does exist.

curi 10 April 2011 11:31:17PM 0 points

I mean, yes, most people would get the [cancer cure/cancer cure+flood right]. It's easy. But the example of Linda the [bank teller/bank teller+active feminist] *does* have exactly the same logical structure.

But it doesn't have the same structure *as presented to the person being asked*. They specifically put in a bunch of context designed so that if you try to use the contextual information provided you get the "wrong" answer (literally wrong, ignoring the fact that the person isn't answering the question literally written down).

It's like if you added the right context to the flood one, you could get people to say both. Like if you talked about how a flood is practically guaranteed for a while, and then ask, some people are going to look at the answers and go "he was talking about floods being likely. this one has a flood. the other one is irrelevant". Why? Well they don't know or care what you mean by the technical question, "Which is more probable?". Answering a different question than the one the researcher claims to have communicated to his subjects is different than people actually, say, doing probability math wrong.

Owen_Richardson 11 April 2011 06:03:46AM * 7 points

Yes, they put in a bunch of irrelevant context, and then asked the question. And rather than answering the question that was actually asked, people tend to get distracted by all those details and make a bunch of assumptions and answer a different question.

Say the researchers put it this way instead: "When we surround a certain simple easy question with irrelevant distracting context, we end up miscommunicating very badly with the subject.

Furthermore, even in 'natural' conditions where no-one is deliberately contriving to create such bad communication, people still often run into similar situations with irrelevant distracting context 'in the wild', and react in the same sorts of ways, creating the same kind of effects.

Thus people often make the mistake of thinking that adding *more* details to explanations/plans makes them *more likely to be good*, and thus do people tend to be insufficiently careful about pinning down every step very firmly from multiple directions."

That's bending over backwards to put the 'fault' with the researchers (and the entire rest of the world), and yet it still leads to the same significant, important real world day-to-day phenomena.

That's why I think you're "wrong" and the "conjunction fallacy" does actually "exist".

Or, to put it another way, "conjunction fallacy exists - y/n" correlates with some actual feature that the territory can meaningfully have or not have, and I'm pretty damn sure that the territory *does* have it.

However! I *would* like to talk about how the same criticism might apply to the [Wason selection task](#), and the hypothesis that humans are naturally bad at reasoning about it generally, but have a specialized brain circuit for dealing with it in social contexts.

Check it. Looking at the wikipedia article, and zeroing in on this part:

Which card(s) should you turn over in order to test the truth of the proposition that if a >card shows an even number on one face, then its opposite face is red?

I know I had to translate that to "even number cards always have red on the other side" before I could solve it, thinking, "well, they didn't specifically say anything about odd number cards, so I guess that means they can be any color on the other side".

Because that's the way people would *normally* phrase such a concept (usually seeking a quick verification of the assumption about it being intended to mean odds can be whatever, if there's anyone to ask the question).

And we do that cuz English aint got an easy distinction between 'if' and 'if and only if', and instead we have these vague and highly variable rules for what they *probably* mean in different contexts (ever noticed that "let's" always seems to be an 'inclusive we' and "let us" always seems to be an 'exclusive we'? Maybe you haven't cuz that's just an oddity of my dialect! I dunno! :P).

And maybe in that sort of context, for people who are used to using 'if' normally and not in a special math jargon way, it seems to be more likely to mean 'if and only if'.

And if the people in the experiment are encouraged to ask questions to make sure they understand the question properly, or presented with the phrasing "even number cards always have red on the other side" instead, maybe in that case they'd be as good at reasoning about evens and odds and reds and browns on cards as they are at reasoning about minors and majors and cokes and beers in a bar.

But that seems like an obvious thing to check, trying to find a way of phrasing it so that people get it right even though it's *not* involving anything "social", and I'd be surprised if somebody hasn't already. Guess I oughta post my own discussion topic on that to find out! But later. Right now, I'ma go snooze.

lessdazed 11 April 2011 06:25:52AM 5 points



Great post.

The thing that bemuses me most at the moment is how bad I am at predicting the voting response to a post of mine on LW. Granted that I and most people upvote and downvote to get post ratings to where we think they should be, rather than rating them by what we thought of them and letting the aggregate chips all where they may, I still can't do well predicting it (that factor should actually be improving my accuracy).

This post of yours is truly fantastic and a moral success; you avoided the opportunities to criticize flaws in the OP and found a truly interesting instance of when an argument resembling that in the OP would actually be valid.

wedrifid 11 April 2011 06:37:31AM 1 point



Granted that I and most people upvote and downvote to get post ratings to where we think they should be,

(This factor applies most significantly to mid-quality contributions. There is a threshold beyond which votes tend to just spiral off towards infinity.)

lessdazed 11 April 2011 07:25:11PM 1 point

I think this is the only example I have seen of a negative infinity spiral on LW. The OP shouldn't be too discouraged, if what you are saying is right (and I think it is). Content rated approaching negative infinity isn't (adjudged) that much worse than content with a high negative rating.

AlephNeil 10 April 2011 11:16:05PM 7 points

Lost me at:

| designed to prove that people are biased idiots.

novalis 10 April 2011 11:10:13PM * 9 points

This post is pretty incoherent, but I was able to get from it that you had done basically no research on the conjunction fallacy. So here, [have a random paper](#).

Zack_M_Davis 10 April 2011 11:03:17PM * 19 points

| The conjunction fallacy does not exist, as it claims to, for all X and all Y. That it does exist for specially chosen X, Y and context is incapable of reaching the stated conclusion that it exists for all X and Y.

But the research *isn't* claiming that the conjunction fallacy exists for *all* X and Y. The claim, as I understand it, is that people rely on the [representativeness heuristic](#), which is indeed useful in many contexts but which does not in general obey the probability-theoretic law that $P(X\&Y) < P(X)$.

Further evidence for the conjunction fallacy was covered on this site in the post "[Conjunction Controversy](#)".

CuSithBell 10 April 2011 11:03:07PM * 2 points

I agree that these studies don't necessarily show what they intend to. In fact, I could easily imagine giving a higher probability for X&Y than Y myself... if I were asked about Y first, and if X were evidence that Y is probable that I had not considered. (Of course, after the fact I'd adjust my probability for Y).

HughRistik 11 April 2011 05:42:38AM 4 points

Then you're talking about conditional probability ($Y | X$), rather than joint probability.

It's quite possible that if you thought the questions were asking about conditional probability, participants in these studies might have, too. Let's take the Russia question:

"A complete suspension of diplomatic relations between the USA and the Soviet Union, sometime in 1983."

"A Russian invasion of Poland, and a complete suspension of diplomatic relations between the USA and the Soviet Union, sometime in 1983."

Taken more literally, this question is asking about the joint probability: $P(\text{invasion \& suspension of diplomatic relations})$

But in English, the question could be read as: "A Russian invasion of Poland and, *given this invasion*, a complete suspension of diplomatic relations between the USA and the Soviet Union, sometime in 1983."

in which case participants making that interpretation might give $P(\text{suspension of diplomatic relations} | \text{invasion})$

It's not at all clear that the participants interpreted these questions the way the experimenters thought.

CuSithBell 11 April 2011 06:10:26AM * 3 points

What I mean is, I might not have any particular notion of what could cause a complete suspension of diplomatic relations, and give it, say, .01 probability. Then, when asked the second question, I might think "Oh! I hadn't thought of that - it's actually quite likely (.5) that there'll be an invasion, and that would be likely (.5) to cause a suspension of diplomatic relations, so $A \wedge B$ has a probability of .25 (actually, slightly more). Of course, this means that B has a probability higher than .25, so if I can go back and change my answer, I will."

(These numbers are not representative of anything in particular!)

I do agree, however, that the literature overall strongly suggests the veracity of this bias.

wedrifid 11 April 2011 06:34:56AM 5 points

In the case of Russia/Poland question the subjects were professional political analysts. For this case in particular we can assume that for these subjects the amount of information about political relations included in the question itself is insignificant.

CuSithBell 11 April 2011 03:00:36PM 2 points

That's... somewhat discouraging.

wedrifid 11 April 2011 03:42:04PM 3 points

Enough so that I had to triple check the source to be sure I hadn't got the details wrong.

CuSithBell 11 April 2011 04:04:05PM 2 points

It certainly contradicts the claim that these studies test artificial judgments that the subjects would never face in day-to-day life.

Manfred 10 April 2011 10:59:29PM 9 points

The conjunction fallacy does not exist, as it claims to, for all X and all Y.

Where do people claim that? I've seen the opposite claimed quite often. You seem to be strawmanning.

drethelin 10 April 2011 10:51:06PM 12 points

Fallacies are not hard and fast rules about the way people think. The conjunction fallacy is an example of something people **SOMETIMES** do wrong when thinking, and saying that it only happens in specific situations makes it not exist is like saying ad hominem attacks don't exist because you've never used one.

vallinder 18 July 2011 10:52:06AM 1 point

Ah, that's interesting. Thanks for clarifying.

Perplexed 10 April 2011 07:51:56PM * 5 points

Thanks for posting this. I would definitely enjoy seeing a debate between Deutsch and Yudkowsky.

The part that dealt with ethics was incredibly naive. About 47 minutes in, for example, he is counseling us not to fear ET, because ET's morality will inevitably be superior to our own. And the slogan: "All evils are due to lack of knowledge". Why does this kind of thing remind me of George W. Bush?

But I agreed with some parts of his argument for the superiority of a Popperian approach over a Bayesian one when 'unknown unknowns' regarding the growth of knowledge are involved. For example, 42:30 in when he quotes Popper advising us to drop the hopeless search for an inerrant source of knowledge, and to instead search for a fairly reliable method of eliminating error once it has become established. Maybe a good idea.

I have mixed feelings, though, about his advocacy of optimism. He argues that Malthus's pessimistic predictions failed simply because Malthus had no way of foreseeing the positive effects of the growth of knowledge. But by the same token, optimistic predictions of a positive future for mankind are also liable to fail because they attempt to predict that the growth of knowledge will include specific breakthroughs.

timtyler 15 April 2011 03:41:59PM * 0 points

The part that dealt with ethics was incredibly naive. About 47 minutes in, for example, he is counseling us not to fear ET, because ET's morality will inevitably be superior to our own.

This seems pretty daft to me too. It looks like a kind of moral realism - according to which being eaten by aliens might well be "good" - since it leads to more "goodness".

Perplexed 15 April 2011 03:54:43PM 1 point

Right. But moral realism is not necessarily daft. It only becomes so when you add in universalism and a stricture against self-indexicality.

timtyler 15 April 2011 06:21:58PM * 1 point

I have some sympathies for the idea that convergent evolution is likely to *eventually* result in a *universal morality* - rather than, say, pebble sorters and baby eaters. If true, that might be considered to be a *kind* of moral realism.

Perplexed 15 April 2011 06:49:59PM 2 points

It is a kind of moral realism if you add in the proclamation that one *ought* to do now that which we all converge toward doing later. Plus you probably need some kind of argument that the limit of the convergence is pretty much independent of the starting point.

My own viewpoint on morality is closely related to this. I think that what one morally ought to do now is the same as what one prudentially and pragmatically ought to do in an ideal world in which all agents are rational, communication between agents is cheap, there are few, if any, secrets, and lifetimes are long. In such a society, a strongly enforced "social contract" will come into existence, which will have many of the characteristics of a universal morality. At least within a species. And to some degree, between species.

timtyler 16 April 2011 01:14:41PM * 1 point

It is a kind of moral realism if you add in the proclamation that one ought to do now that which we all converge toward doing later.

LW 2.0 Open Beta Live

by Vaniver | 21v (33c)

In support of yak shaving part 2

by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag

by Raemon | 41v (106c)

European Community Weekend 2017

by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about "protecting" the epistemology

by AnnaSalamon | 73v (53c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"

by AnnaSalamon | 36v (43c)

Be secretly wrong

by Benquo | 32v (47c)

CFAR's new focus, and AI Safety

by AnnaSalamon | 30v (88c)

Fact Posts: How and Why

by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (32)

ingres (30)

Lumifer (27)

turchin (23)

Habryka (22)

Yosarian2 (21)

IlyaShpitser (19)

SaidAchmiz (19)

gjm (19)

fowlertm (17)

SquirrelInHell (15)

Dagon (15)

Elo (14)

Stuart_Armstrong (13)

Gunnar_Zarncke (12)

RECENT KARMA AWARDS

...or if you think what we ought to be doing is helping to create the thing with the universal moral values.

I'm not really convinced that the convergence will be complete, though. If two advanced alien races meet, they probably won't agree on all their values - perhaps due to moral spontaneous symmetry breaking - and small differences can become important.

Eugine_Nier 10 April 2011 10:54:24PM 6 points

And the slogan: "All evils are due to lack of knowledge". Why does this kind of thing remind me of George W. Bush?

Well, it reminds me of Plato, which is much more damning.

curi 10 April 2011 08:09:20PM 1 point

. And the slogan: "All evils are due to lack of knowledge".

You should read his book, *The Beginning of Infinity*. It's not a slogan but a philosophical position which he explains at length. Learn *why* he thinks it. He's not an idiot.

Since you partly agree with him, and have mixed feelings, I think it'd be worth looking into for you, so I wanted to let you know it's much more than a slogan! And "optimism" to DD does not mean "predicting a positive future", it's not about wearing rose colored glasses.

NancyLebovitz 10 April 2011 09:37:36AM 2 points

My first reaction to his unlimited progress riff was "every specific thing I care about will be gone". The answer is presumably that there will be more things to care about. However, that initial reaction is probably common enough that it might be worth working on replies to people who are less inclined to abstraction than I am.

I'll take the edge off his optimism somewhat by pointing out that individuals and cultures can be rolled over by change, even if the whole system is becoming more capable, and we care about individuals and cultures (especially if they're us or ours) as well as the whole system.. Taking European diseases to the New World happened by accident.

Still, the pursuit of knowledge and competence may well be the least bad strategy the vast majority of the time (rather than a guarantee of things becoming more wonderful for what we personally care about), and I'm intrigued by the idea of explicitly intending to increase the returns for cooperation.

Manfred 10 April 2011 02:00:55AM * 10 points

I stopped listening fairly quickly, after determining that it was rubbish from a Bayesian perspective. Specifically I stopped listening when he says that the future of humanity is different from russian roulette because the future can't be modeled by probability. This is the belief that there is a basic "probability-ness" that dice have and gun chambers have but people don't, and that things with "probability-ness" can be described by probability, but things without "probability-ness" can't be. But of course, we're all fermions and bosons in the end - there is no such thing as "probability-ness," probability is simply what happens when you reason from incomplete information.

NancyLebovitz 10 April 2011 09:39:45AM 5 points

Deutsch is arguing (and I think correctly) that there's a difference between knowing the full range of possibilities in a system and not knowing it.

Manfred 10 April 2011 04:05:23PM * 4 points

That seems pretty reasonable. "What will the future be like" is a pretty undetermined question.

However, he was applying this same logic to "will civilization be destroyed," where "destroyed" and "not destroyed" are a pretty complete range of possibilities.

Unless maybe he meant that you have to know every possible way civilization could be destroyed in order to estimate a probability, which seems like searching for a reason that civilization doesn't have probability-ness.

Vladimir_Nesov 10 April 2011 01:51:32AM * 5 points

[...]

It is a talk given to the Oxford Transhumanists. Their previous speaker was Eliezer Yudkowsky.

To clarify what I originally misinterpreted on reading this description: according to [this page](#), Yudkowsky was giving a talk on 25 Jan 2011, while Deutsch on 10 Mar 2011, so "previous speaker" doesn't refer to giving talks in succession.

Vladimir_Nesov 09 April 2011 11:45:22PM 13 points

[...]

I think this talk [motivates](#) a Yudkowsky-Deutsch debate on bloggingheads.

alexflint 10 April 2011 05:17:11PM 1 point

[...]

Oh boy oh boy oh boy that would rock my socks

Larks 09 April 2011 11:18:52PM 1 point

[...]

It's slow loading for me due to a slow internet connection, but if the questions at the end are included, I was the one who asked about insurance companies.

I don't think his response was very satisfactory, though I have a better version of my question.

Suppose I give you some odds $p:q$ and force you to bet on some proposition X (say, Democrats win in 2012) being true, but I let you pick which side of the bet you take; a payoff of p if X is true, or a payoff of q if X is false. For some (unique) value of p/q , you'll switch which side you want to take.

It seems this can force you to assign probabilities to arbitrary hypothesis.

Eugene_Nier 10 April 2011 04:31:30PM 2 points

[...]

Suppose I give you some odds $p:q$ and force you to bet on some proposition X (say, Democrats win in 2012) being true, but I let you pick which side of the bet you take; a payoff of p if X is true, or a payoff of q if X is false. For some (unique) value of p/q , you'll switch which side you want to take.

It seems this can force you to assign probabilities to arbitrary hypothesis.

So, how precise should these probabilities be? Any why can't I apply this argument to force the probabilities to have arbitrary high precision?

Larks 10 April 2011 06:57:27PM 1 point

[...]

Not that I can think of, besides memory/speed constraints, and how much updating you can have done with the evidence you've received.

Eugene_Nier 10 April 2011 07:31:53PM 2 points

[...]

and how much updating you can have done with the evidence you've received.

Why can't it happen that you have so little and/or such weak evidence, that the amount of precision you should have is none at all?

Larks 10 April 2011 08:03:45PM 0 points [-]

Well, your prior gives you a unique value, and bayes theorem is a function, so it gives you a unique value for every input.

Eugene_Nier 10 April 2011 08:50:33PM 2 points [-]

Well, your prior gives you a unique value,

So the claim is that you have arbitrary precision priors. What are they, and where are they stored?

Larks 10 April 2011 09:38:21PM 0 points [-]

Sorry, I haven't been very clear. A perfect bayesian agent would have a unique real number to represent it's level of belief in every hypothesis.

The betting-offer system I described about can force people (and force any hypothetical agent) to assign unique values.

Of course, an actual person won't be capable of this level of precision or coherence.

Eugene_Nier 10 April 2011 08:17:05PM 1 point [-]

Yes, but actually computing that function is computationally intractable in all but the simplest examples.

Manfred 10 April 2011 08:01:44PM * 0 points [-]

Imagine that you had to give a probability density to each probability estimate you could make of Obama winning in 2012 being the correct one. You'd end up with something looking like a bell curve over probabilities, centered somewhere around "Obama has a 70% (or something) chance of winning." Then to make a decision based on that distribution using normal decision theory, you would average over the possible results of an action, weighted by the probability. But this is equivalent to taking the mean of your bell curve - no matter how wide or narrow the bell curve, all that matters to your (standard decision theory) decision is the location of the mean.

Less evidence is like a wider bell curve, more evidence like a sharper one. But as long as the mean stays the same, the average result of each decision stays the same, so your decision will also be the same.

So there are two kinds of precision here: the precision of the mean probability given your current (incomplete) information, which can be arbitrarily high, and the precision with which you estimate the true answer, which is the width of the bell curve. So when you say "precision," there is a possible confusion. Your first post was about the "how precise can these probabilities be," which was the first (and boring, since it's so high) kind of precision, while this post seems to be talking about the second kind, the kind that is more useful because it reflects how much evidence you have.

[deleted] 11 April 2011 02:46:14PM 0 points [-]

Imagine that you had to give a probability density to each probability estimate you could make of Obama winning in 2012 being the correct one. You'd end up with something looking like a bell curve over probabilities

Bell curves prefer to live on unbounded intervals! It would be less jarring, (and less convenient for you?), if he ended up with something looking like a uniform distribution over probabilities.

Manfred 11 April 2011 06:07:57PM 0 points [-]

It's equally convenient, since the mean doesn't care about the shape. I don't think it's particularly jarring - just imagine it going to 0 at the edges.

The reason you'll probably end up with something like a bell curve is a practical one - the [central limit theorem](#). For complicated problems, you very often get what looks something like a bell curve. Hardly watertight, but I'd bet decent amounts of money that it is true in this case, so why not use it to add a little color to the description?

Eugene_Nier 10 April 2011 08:48:20PM 2 points

So there are two kinds of precision here: the precision of the mean probability given your current (incomplete) information, which can be arbitrarily high, and the precision with which you estimate the true answer, which is the width of the bell curve.

I'm not sure what you mean by the "true answer". After all, in some sense the true probability is either 0 or 1 it's just that we don't know which.

Manfred 10 April 2011 09:09:11PM 1 point

That's a good point. So I guess the second kind of precision doesn't make sense in this case (like it would if the bell curve were over, say, the number of beans in a jar), and "precision" should only refer to "precision with which we can extract an average probability from our information," which is very high.

curi 09 April 2011 11:23:36PM -4 points

That does not require probabilities. You could also come up with an explanation of at what value to switch.

Larks 09 April 2011 11:29:55PM 8 points

In that case, we're done. Standard probability theory/Cox Theorem/de Finette would give us a ready made criticism of any conjecture that wasn't isomorphic to probability theory, so we'd have isomorphism, which is all we need. Once we have functional equivalence, we can prove results in probability theory, apply Bayes theorem, etc., and then at the end translate back into Popperesque.

(Also, IIRC, Jaynes only claimed to have proven that rational reasoning must be isomorphic to probability theory)

curi 09 April 2011 11:46:54PM 2 points

I don't quite get your point. You are saying that if you bring up betting (a real life scenario where probability is highly relevant), then *given* your explanations that help you come up with priors (background knowledge needed to be able to do any math about it), you shouldn't act on those explanations in ways that violates math. OK, so what? probability math is useful in some limited cases, given some explanatory knowledge to get set up. no one said otherwise.

ata 10 April 2011 06:40:55PM 6 points

You are saying that if you bring up betting (a real life scenario where probability is highly relevant) Every decision is a bet.

Sniffnoy 10 April 2011 02:47:19PM 4 points

I think you are beginning to get the point. :) The key missing fact here is that in fact the resulting math is

highly constraining, to the point that if you actually follow it all the way you will be acting in a manner isomorphic to a Bayesian utility-maximizer.

curi 10 April 2011 07:04:21PM -3 points

But the background knowledge part is highly not-constraining (just given your math). When a math algorithm gives constrained output, but you have wide scope for choice of input, it's not so good. you need to do stuff to constrain the inputs.

it seems to me you just dump all the hard parts of thinking into the priors and then say the rest follows. but the hard parts are still there. we still need to work out good explanations to use as input for the last step of not doing stuff that violates math/logic.

Vladimir_Nesov 09 April 2011 08:14:40PM * 27 points

Deutsch argues that the future is fundamentally unpredictable, that for example expected utility considerations can't be applied to the future, because we are ignorant of the possible outcomes and intermediate steps leading to those outcomes, and the options that will be available; and there is no way to get around this. The very use of the concept of probability in this context, Deutsch says, is invalid.

As illustration, among other things, he lists some failed predictions made by smart people in the past, attributing failure to unavailability of the ideas relevant for the predictions, ideas that will only be discovered much later.

<18:57> [Science can't] predict any phenomenon whose course is going to be affected by the growth of knowledge, by the creation of new ideas. This is the fundamental limitation on the reach of scientific explanation and prediction.

<20:33> [Predictions that are serious attempts to extract unknowable answers from existing knowledge] are going to be biased towards bad outcomes.

(If it's unknowable, how can we know that a certain prediction strategy is going to be systematically biased in a known direction? Biased with respect to what knowable standard?)

Deutsch explains:

And the basic reason for that is that, as I said, the growth of knowledge is good, so that kind of prophesy, which can't imagine it, is going to be biased against prophesying good.

<24:54> Reason and science are the means to progress. They are not means to prophesy.

On a more constructive if not clearly argued note:

<25:33> Merely pulling the trigger less often doesn't change the inevitability of doom. [...] One of the most important uses of technology is to counteract disasters and to recover from disasters, both from foreseen and unforeseen evil. Therefore, the speed of progress itself is one of the things that is a defense against catastrophe.

<26:53> The speed of progress is one of the things that gives the good guys the edge over the bad guys, because good guys make faster progress.

(Possibly an example of the [halo effect](#): the good guys are good, the progress is good, so the good guys will make faster progress than the bad guys. Quite probably, there was better reasoning behind this argument, but Deutsch doesn't give it, and doesn't hint at its existence, probably because he considers the conclusion obvious, which is in any case a flaw of the talk.)

For the next 10 minutes or so he argues for the possibility of essentially open-ended technological progress.

<39:27> The amount of knowledge in an environment of rational thought that allows it to grow, grows exponentially relative to the speed of computation.

[...] It's a mistake to think of the so-called singularity as being a shock, where we find that we can't cope with life, because iPhone updates are coming [...] every second. That's a mistake, because when progress reaches

that speed, our technologically enhanced speed of thinking will have increased in proportion, and so subjectively again we will experience mere exponential growth.

Here, Deutsch seemingly makes the same mistake he discussed at the beginning of the talk: making detailed predictions about future technology that depend on the set of technology-defining ideas presently available (which, by his own argument, can lead to underestimation of progress).

The conclusion is basically a better version of Kurzweil's view of Singularity, that ordinary technological progress is going to continue indefinitely (Deutsch's progress is exponential in *subjective time*, not in physical time). Yudkowsky [wrote](#) in 2002:

I've come to the conclusion that what Kurzweil calls the "Singularity" is what we would call "the ordinary progress of technology." In Kurzweil's world, the Grinding Gears of Industry churn out AI, superhuman AI, uploading, brain-computer interfaces and so on, but these developments do not affect the nature of technological progress except insofar as they help to maintain Kurzweil's curves *exactly on track*.

Deutsch considers Popper's views on the process of development of knowledge, pointing out that there are no reliable sources of knowledge, and so instead we should turn to finding and correcting errors. From this he concludes:

<44:48> Optimism demands that we not try to extract prophesies of everything that could go wrong in order to forestall it from our existing scanty and misconception-laden existing knowledge. Instead, we need policies and institutions that are capable of correcting mistakes and recovering from disasters when they happen. When, not if.

(This doesn't terribly help with existential risks. Also, this optimism thing seems to be one magically reliable source of knowledge, strong enough to ignore whatever best conclusions it is possible to draw [using the best tools currently available](#), however poor they seem on the great cosmic scale.)

<46:00> The way to prevent that nightmare of rogue AI apocalypse is not try to enslave our AIs, because if the AIs are creating new knowledge (and that's a definition of AI), then successfully enslaving them would require foretelling (prophesying) the ideas that they could have, and the consequences of those ideas, which is impossible.

This was addressed in [Knowability of Friendly AI](#) and many later Yudkowsky's writings, most recently in his [joint paper with Bostrom](#). Basically, you can't predict the moves of a good chess AI, otherwise you'd be at least that good chess player yourself, and yet you know it's going to win the game.

Deutsch continues:

So instead, just as for our fellow humans, and for the same reason, we must allow AIs to integrate into the institutions of our open society.

(Or, presumably, so Optimism demands, since the AIs are unpredictable, and technology.)

<47:55> The only moral values that permit sustained progress are the objective values of an open society and more broadly of the enlightenment. No doubt, the [extraterrestrials'] morality would not be the same as ours, but nor will it be the same as that of 16th century conquistadors. It will be better than ours.

Finally, Deutsch summarizes the meaning of the overarching notion of "optimism" he has been using throughout the talk:

<49:50> Optimism in this sense that I have argued for is not a feeling, is not a bias or spin that we put on facts, like, you know, half-full instead of half-empty, nor on predictions, it's not hope for the best, nor blind expectation of the best (in some sense it's quite the contrary, we expect errors). It is a cold, hard, far-reaching implication of rejecting irrationality, nothing else. Thank you for listening.

(No good questions in the quite long Q&A session. No LWers in the audience, I guess, or only the shy ones.)

timtyler 15 April 2011 06:29:56PM * 0 points

[...]

Possibly an example of the halo effect: the good guys are good, the progress is good, so the good guys will make faster progress than the bad guys.

This is surely a real effect. The government is *usually* stronger than the mafia. The army is stronger than the terrorists. The cops usually beat the robbers, etc.

timtyler 15 April 2011 02:31:58PM 2 points

Deutsch argues that the future is fundamentally unpredictable, that for example expected utility considerations can't be applied to the future, because we are ignorant of the possible outcomes and intermediate steps leading to those outcomes, and the options that will be available; and there is no way to get around this. The very use of the concept of probability in this context, Deutsch says, is invalid.

This bit starts about 12 minutes in. It is *complete* nonsense - Deutsch does not have a clue about the subject matter he is talking about :-)

XiXiDu 10 April 2011 11:17:27AM * 1 point

Basically, you can't predict the moves of a good chess AI, otherwise you'd be at least that good chess player yourself, and yet you know it's going to win the game.

I just realized you tried to make a different point here. That one can prove the behavior of computationally unpredictable systems. Reminds me of the following:

6) Disproving mathematical proofs within the terms of their own definitions. This falls within the realm of self-contradiction. No transparent has disproved the Pythagorean Theorem for Euclidean spaces as defined by classical Greek mathematicians, for instance, or disproved Godel's Incompleteness Theorem on its own terms. ([Encyclopedia Galactica - Limits of Transparent Power](#))

Sounds reasonable but I have no idea to what extent one could prove "*friendliness*" while retaining a degree of freedom that would allow a seed AI to recursively-selfimprove towards superhuman intelligence quickly. Intuitively it seems to me that the level of abstraction of a definition of "*friendliness*" will be somehow correlated with the capability of an AGI.

XiXiDu 10 April 2011 11:02:18AM * 2 points

Basically, you can't predict the moves of a good chess AI, otherwise you'd be at least that good chess player yourself, and yet you know it's going to win the game.

This is a really good point. When I read it I first thought I would have to disagree, after all we've designed the chess AI and therefore do understand it. But since I am currently reading Daniel Dennett's '*Darwin's Dangerous Idea*' my next thought was that disagreeing with it seems to be a general bias assuming that a design is always inferior to its designer. But it should be obvious that our machines are faster and stronger than us, why not better thinkers too?

Unlike [the blind idiot God](#) we can pinpoint our own flaws and devise solutions but are also unable to apply them to ourselves effectively, which will be realized by the next level of self-redesigning things. But even now our designs can be superior to us as they mirror our own improved upon capabilities, our skills minus our flaws. We are still able to understand our machines but unable to mimic their capabilities as we've been able to recreate some of our skills but haven't been able to benefit from the improvements we devised. We know that steel is tougher than bones, beware "*steel*" that knows this fact as well.

curi 09 April 2011 11:24:59PM 0 points

(Presumably, since the AIs are unpredictable, and technology, Optimism demands that we all live happily ever after.)

No. Deutsch's "principle of optimism" states:

All evils are caused by insufficient knowledge.

optimism demands that they *can* live happily ever after *if they learn how*. it does not predict that they *will*.

Vladimir_Nesov 09 April 2011 11:57:26PM * 0 points

Agreed. The "we all live happily ever after" inference does contradict Deutsch's idea, which I noticed a little

after writing this, and so corrected the wording (before seeing your comment) thusly:

(Or, presumably, so Optimism demands, since the AIs are unpredictable, and technology.)

curi 09 April 2011 11:16:30PM -9 points

Basically, you can't predict the moves of a chess AI, otherwise you'd be at least that good chess player yourself, and yet you know it's going to win the game.

As someone who has beaten chess programs, I have noticed that this sentence as written is false. Would you care to refine it so that it's no longer straightforwardly false?

PhilGoetz 10 April 2011 12:39:48AM 0 points

-5 points seems harsh for a statement that is technically correct.

Desrtopa 10 April 2011 02:49:34PM * 4 points

Yes, but it's the worst sort of manifestation of [this](#) sort of behavior; if someone will attempt to generate conflict by nitpicking when they could so easily have interpreted the argument themselves in such a way as to render it unnecessary, can they be trusted to take arguments as seriously as they deserve to be?

wedrifid 10 April 2011 01:34:29AM 8 points

I would prefer not to see any more comments by curi in conversations by Popper. The quality of discussion makes continued exposure unpleasant. This makes a decision to downvote all such comments appropriate.

curi 10 April 2011 01:03:42AM 3 points

Yes indeed. But also -- and maybe this is only a Popperian thing you guys think is wrong? -- I find that correcting statements, instead of just saying them wrong and leaving it at that, often leads to better understanding. Sometimes you find it's not as easy to correct as you assumed, and maybe change your conclusion a bit.

I think it's easy to make mistakes without realizing it -- happens all the time -- and that not making blatant mistakes -- or at least caring about them and correcting them when you do, rather than deeming it unimportant -- is a good start for dealing with the harder ones.

Emile 10 April 2011 02:36:08PM 3 points

Yes indeed. But also -- and maybe this is only a Popperian thing you guys think is wrong?

This has nothing to do with Popper (I hope, not having read much Popper myself), and everything to do with obnoxious nitpicking in bad faith.

JoshuaZ 10 April 2011 01:24:56AM * 12 points

Yes indeed. But also -- and maybe this is only a Popperian thing you guys think is wrong? -- I find that correcting statements, instead of just saying them wrong and leaving it at that, often leads to better understanding. Sometimes you find it's not as easy to correct as you assumed, and maybe change your conclusion a bit.

No. You are missing the point. The easy correction would be for you to say "Well, the chess claim might not be true. But your point still goes through if I used Go and one of the world's best Go players or some chess variant like [Andernach chess](#) or [cylindrical chess](#) or [Capablanca chess](#)." And then respond to the argument in that form.

It isn't helpful to pick out a small problem with an argument someone makes and then ignore the rest of the argument until they've responded to doing so. It might feel fun, and it might be rhetorically impressive in some circumstances, but it doesn't really help resolving disagreement or improving understanding of what people are trying to communicate.

curi 10 April 2011 05:07:40AM -11 points

It isn't helpful to pick out a small problem with an argument

But that's just my point: it is.

We need to learn not to make small mistakes.

That's why the big problems are so hard: too many small mistakes everywhere.

Fix the little stuff. Then fix a little more. You make progress.

People should appreciate every single little mistake they make being pointed out, and should strive to stop making them. If not making little mistakes is too hard for someone, not making big ones is out of reach.

JoshuaZ 10 April 2011 05:18:15AM 7 points

curi, please reread what I wrote. Please note that the whole sentence is not as you quoted "It isn't helpful to pick out a small problem with an argument" but "It isn't helpful to pick out a small problem with an argument someone makes and then ignore the rest of the argument until they've responded to doing so." Please also reread the first paragraph where I outlined the ideal approach, noting the issue, and supplying a correction yourself, and then replying to the corrected form.

curi 10 April 2011 05:31:18AM * -4 points

I sometimes try to fix people's arguments for them.

What sometimes happens is they don't like or want the fixed form.

That especially happens a lot when it's a person with different methods of judging what is a good argument, and a rather different worldview than my own.

That issue applies here. Having people fix their own stuff is the less ambitious and less error prone approach. It's the one that is more resilient to frequent miscommunication.

I am especially not inclined to fix people's arguments when they are wrong either way. Fixing from wrong to still wrong is weird. In that case, I think it would be wise for people to fix mistakes in their argument, one by one. And starting with an easy one is good, not bad. If they won't even fix that, what is going to happen with a more subtle issue?

If I jump ahead to the fully clarified version of their argument -- written perhaps in a way that also helps make it easier to see why it's wrong -- people complain. If I write it in a way that makes it hard to see why it's wrong (as they've been doing by unconscious bias), then they might be a bit happier but we'll still have the problem of having to walk them through improvements of it until they get it to a better state.

Learning isn't super easy. You need patience and persistence. You need to be happy fixing one mistake at a time. You shouldn't complain that people aren't helping you *skip steps*. That it's too slow. That if only I would reply to what I know they meant, instead of what they actually said, we'd make progress faster. Attempts at mind reading increase miscommunication difficulties and misunderstanding. They usually seem to work well because the two people do it share a ton of background knowledge, cultural assumptions, biases, and so on.

If I was replying to what people really meant it'd just create a mess. I would reply to a lot of their unconscious biases they weren't aware they had and they'd just get confused. And yet many of their statements express those unconscious ideas. To learn, they need to engage with the process of improving their ideas, not just insist I should be able to improve their stuff to the point

it's true -- without changing the conclusion -- and then concede.

As to skipping ahead while smaller issues are still pending, I wonder why you think building on rotten foundations is wise. I think it can work sometimes, but it's a bit ambitious.

Sniffnoy 10 April 2011 02:24:55PM 4 points

I am especially not inclined to fix people's arguments when they are wrong either way. Fixing from wrong to still wrong is weird.

Not really, it's a common method for showing that someone is very wrong. It's just the common "But let's suppose we fix that [alternatively, that I spot you that for now] - even then there's still a problem, as..."

Regarding much of the rest of the post: The idea is not to silently reply to a corrected version, but to *explicitly* note the correction and reply to that! Then people can, rather than just being confused about your correction, actually evaluate your corrected version and verify whether or not it still conforms to their intentions.

As to skipping ahead while smaller issues are still pending, I wonder why you think building on rotten foundations is wise. I think it can work sometimes, but it's a bit ambitious.

Hence you *fix those foundations*, rather than silently building on top of them.

curi 10 April 2011 07:31:09PM -4 points

I am especially not inclined to fix people's arguments when they are wrong either way

Not really, it's a common method

I don't like attributing to people false ideas they didn't actually write. I think that's a recipe for disaster. You disagree?

I wasn't talking about silent corrections either.

JoshuaZ 10 April 2011 12:54:27AM 9 points

I disagree. It is a good example where it is obvious or close to obvious what was intended. The remark simply damaged the signal to noise ratio while avoiding grappling with the point.

PhilGoetz 10 April 2011 02:59:26AM 2 points

True - but I don't think it would ordinarily have been down-voted that hard, for that sin.

JoshuaZ 10 April 2011 03:01:33AM 7 points

True - but I don't think it would ordinarily have been down-voted that hard, for that sin.

It is possible that some general annoyance with the user also resulted in the total.

Vladimir_Nesov 10 April 2011 12:00:41AM 5 points

As someone who has beaten chess programs

Corrected to "good chess AI". [LCPW](#) applies.

Larks 09 April 2011 11:24:11PM 2 points [\[-\]](#)

trivial refinement: "a chess AI that is much better than you", or "any chess AI that has beaten a grandmaster", as I assume neither you nor Nesov are grandmasters.

curi 09 April 2011 11:29:02PM -4 points [\[-\]](#)

If you understand that a program plays chess well, then you have an understanding of the matter. It's not prophecy apply your understanding.

The chess computer isn't even relevant here. I can understand something about how Kasparov plays, and how I play, and then predict he'll beat me. So what?

Larks 09 April 2011 11:33:07PM 4 points [\[-\]](#)

It shows you can know general facts about a system that creates new knowledge, despite not knowing all the specific facts/bits of knowledge that it will create. We can know Kasparov will beat us despite not knowing exactly what move he'll take; we can know that an AGI will destroy/save/whatever us despite not knowing exactly how.

curi 09 April 2011 11:36:53PM * -4 points [\[-\]](#)

Chess playing programs don't create new knowledge.

So, the argument is wrong without me fixing it (human chess players do).

Small amounts of new knowledge in very limited areas is predictable. Like writers can predict they will finish writing a book (even if they haven't worked out 100% of the plot yet) in advance.

This doesn't have much to do with large scale prediction that depends on new *types* of knowledge, does it?

JoshuaZ 09 April 2011 11:52:56PM 4 points [\[-\]](#)

Whether you call it new knowledge or not it not relevant. Nor is new types of knowledge what is generally relevant (aside from the not at all small issue that "type" isn't a well-defined notion in this context.)

Like writers can predict they will finish writing a book (even if they haven't worked out 100% of the plot yet) in advance.

Actually, writers sometimes start a book and find part way through that they don't want to finish, or the book might even change genres in the process of writing. If you prefer an example, I can predict that Brandon Sanderson's [next Mistborn book](#) will be awesome. I can predict that it will sell well, and get good reviews. I can even predict a fair number of plot points just based on stuff Sanderson has done before and various comments he has made. But, at the same time, I can't write a novel nearly as well as he does, and if he and I were to have a novel writing contest, he will beat me. I don't know how he will beat me, but he will.

Similarly, a sufficiently smart AI has the same problem. If it decides that human existence is non-optimal for it to carry out its goals, then it will try to find ways to eliminate us. It doesn't matter if all the ways it comes up with of doing so are in a fairly limited set of domains. If it is really good at chemistry it might make nasty nanotech to reduce organic life into constituent atoms. If it is really good at math it might break all our cryptography, and then hack into our missiles and trigger a nuclear war (this one is obvious enough that there are multiple movies about it). If it is really good at social psychology it might manipulate us over a few years into just handing over control to it.

Just as I don't know how Kasparov will beat me but I know he will, I don't know how a sufficiently intelligent AI will beat me, but I know it will. There may be issues with how sufficiently intelligent it needs to be and whether or not an AGI will be likely undergo fast, substantial, recursive self-

improvement to get to be that intelligent is an issue of much discussion on LW (Eliezer considers it likely. Some other people such as myself consider it to be unlikely.) but the basic point about sufficient intelligent seems clear.

curi 10 April 2011 12:00:11AM -1 points

Whether you call it new knowledge or not it not relevant.

Considering that Deutsch was talking about new knowledge, and I use the same terminology as him, it is relevant.

Actually, writers sometimes start a book and find part way through that they don't want to finish,

I know that? And if I played Kasparov I might win. It's not a 100% guaranteed prediction.

@Sanderson: you understand what kind of thing he's doing pretty well. writers are a well known phenomenon. the less you know what processes he uses to write, what tradition he's following - - in general what's going on -- the less you can make any kind of useful predictions.

If it decides that human existence is non-optimal for it to carry out its goals

why would it?

Deutsch doesn't think AGI's will do fast recursive self-improvement. They can't because the first ones will already be universal and there's nothing much left to improve, besides their knowledge (not their design, besides making it faster). Improving knowledge with intelligence is the same process for AGI and humans. It won't magically get super fast.

JoshuaZ 10 April 2011 12:18:38AM * 2 points

Considering that Deutsch was talking about new knowledge, and I use the same terminology as him, it is relevant.

Then the define the term.

I know that? And if I played Kasparov I might win. It's not a 100% guaranteed prediction.

So what? How is that at all relevant. It isn't 100% guaranteed that if I jump off a tall building that I will then die. That doesn't mean I'm going to try. You [can't use the fact that something isn't definite as an argument to ignore the issue wholesale.](#)

Deutsch doesn't think AGI's will do fast recursive self-improvement. They can't because the first ones will already be universal and there's nothing much left to improve, besides their knowledge (not their design, besides making it faster).

Ok. So I'm someone who finds extreme recursive self-improvement to be unlikely and I find this to be a really unhelpful argument. Improvements in speed matter. A lot. Imagine for example, that our AI finds a proofs that $P=NP$ and that this proof gives a $O(n^2)$ algorithm for solving your favorite NP-complete problem, and that the constant in the O is really small. That means that the AI will do pretty much everything faster, and the more computing power it gets the more disparity there will be between it and the entities that don't have access to this algorithm. It wants to engineer a new virus? Oh what luck, [protein folding is under many models NP-competete](#). The AI decides to improve its memory design? Well, that involves graph coloring and the traveling salesman, also NP-complete problems. The AI decides that it really wants access to all the world's servers and add them to its computational power? Well most of those have remote access capability that is based on cryptographic problems which are much weaker than NP-complete. So, um, yeah. It got those too.

Now, this scenario seems potentially far-fetched. After all, most experts consider it to be unlikely that $P=NP$, and consider it to be extremely unlikely that there's any sort of fast algorithm for NP complete problems. So let's just assume instead that the AI tries to make itself a lot faster. Well, let's see, what can our AI do. It could give itself some nice quantum computing hardware and then use [Shor's algorithm](#) to break factoring in polynomial time and

then all AI can just take over lots of computers and have fun that way.

Improving knowledge with intelligence is the same process for AGI and humans. It won't magically get super fast

This is not at all obvious. Humans can't easily self-modify our hardware. We have no conscious access to most of our computational capability, and our computational capability is very weak. We're pathetic sacks of meat that can't even multiply four or five digits numbers in our heads. We also can't save states and swap out cognitive modules. An AGI can potentially do all of that.

Don't underestimate the dangers of a recursively self-improving entity or the value of speed.

curi 10 April 2011 01:36:02AM -3 points



Then the define the term.

See the essay on knowledge: <http://fallibleideas.com/>

Or read Deutsch's books.

It isn't 100% guaranteed that if I jump off a tall building that I will then die.

Indeed. You're the one who told me that writers sometimes don't finish books... They aren't 100% guaranteed to. I know that. Why did you say that?

Imagine for example, that our AI finds a proofs that $P=NP$ and that this proof gives a $O(n^2)$ algorithm for solving your favorite NP-complete problem, and that the constant in the O is really small.

Umm. Imagine a human does the same thing. What's your point? My/Deutsch's point is AGIs have no special advantage over non-artificial intelligences at finding a proof like that in the first place.

We're pathetic sacks of meat that can't even multiply four or five digits numbers in our heads.

That's not even close to true. First of all, I could do that if I trained a bit. Many people could. Second, many people can memorize long sequences of the digits of pi with some training. And many other things. Ever reading about Renschaw and how he trained people to see faster and more accurately?

[continue this thread »](#)

Vladimir_Nesov 10 April 2011 12:10:07AM 2 points



And if I played Kasparov I might win. It's not a 100% guaranteed prediction.

The [fallacy of gray](#)? Between zero chance of winning a lottery, and epsilon chance, there is an [order-of-epsilon difference](#). If you doubt this, let epsilon equal one over googolplex.

curi 10 April 2011 12:16:17AM -5 points



No, the fallacy of you not paying attention to the context of statements, and their purpose.

I said authors predict they will finish books.

Someone told *me* that those predictions are not 100% accurate.

I said, basically: so what? And I pointed out his same "argument" works just as well (that is, not at all) in other cases.

So, the other guy did the "fallacy of gray", not me. And you didn't read carefully.

[continue this thread »](#)

curi 09 April 2011 10:12:45PM 0 points



(Possibly an example of the halo effect: the good guys are good, the progress is good, so the good guys will make faster progress than the bad guys. Quite probably, there was better reasoning behind this argument, but Deutsch doesn't give it, and doesn't hint at its existence, probably because he considers the conclusion obvious, which is in any case a flaw of the talk.)

He doesn't consider it obvious. He considers nothing obvious in general (in a serious, not vacuous way). This in particular he has thought about, not because it is obvious but because it isn't.

The basic reason "good guys" make progress faster than "bad guys" (in the sense of: immoral guys, like prone to violence) is that they have more stable, peaceful, cooperative societies that are better suited to making progress. It's because good values are more effective in real life.

There's discussion of this stuff in his book *The Beginning of Infinity*.

JoshuaZ 09 April 2011 10:40:46PM * 4 points



The basic reason "good guys" make progress faster than "bad guys" (in the sense of: immoral guys, like prone to violence) is that they have more stable, peaceful, cooperative societies that are better suited to making progress. It's because good values are more effective in real life.

This sort of claim seems to run into historical problems. A lot of major expansionist violent empires have done quite well for themselves. In modern times, some of the most "bad" groups have done well as well. The Nazis in many ways had much better technology than the Allies. If they hadn't been ruled by an insane dictator they would have done much better. Similarly, if they had expanded just as much but waited to start the serious discrimination and genocide until after they already had won they would have likely won. Similarly, in WW2, Japan did quite well for itself, and if a handful of major battles had gone slightly differently, the outcome would have been very different.

Or to use a different, but potentially more controversial example, in North America and in Australia, the European colonizers won outright, despite having extremely violent, expansionist policies. In North America, you actually had multiple different European groups fighting amongst themselves as well and yet they still won.

Overall, this is a pleasant, optimistic claim that seems to be depressingly difficult to reconcile with actual history.

Vladimir_M 12 April 2011 01:38:54AM * 10 points



Similarly, in WW2, Japan did quite well for itself, and if a handful of major battles had gone slightly differently, the outcome would have been very different.

You are wrong about this. Even if every single American ship magically got sunk at some point in 1941 or 1942, and if every single American soldier stationed outside of the U.S. mainland magically dropped dead at the same time, it would only have taken a few years longer for the U.S. to defeat Japan. Once the American war production was up and running, the U.S. could outproduce Japan by at least two orders of magnitude and soon overwhelm the Japanese navy and air force no matter what their initial advantage. Starting the war was a suicidal move for the Japanese leadership, and even [the sane people among them knew it](#).

I think you're also overestimating the chances Germans had, and underestimating how well Hitler did given the circumstances, though that's more controversial. Also, Germany lost the technological race in pretty much all theaters of war where technology was decisive -- submarine warfare, cryptography, radars and air defense, and nuclear weapons all come to mind. The only exceptions I can think of are jet aircraft and long-range missiles, but even in these areas, they produced mostly flashy toys rather than strategically relevant weapons.

Overall, I think it's clear that the insanity of the regimes running Germany and Japan hampered their technological progress and also led to their suicidal aggressiveness. At the same time, the relative sanity of the regimes running the U.K. and the U.S. did result in significant economic and technological advantages, as well as somewhat saner strategy. Of course, that need not have been decisive -- after all, the biggest winner of the war was Stalin, who was definitely closer to the defeated sides in all the relevant respects, if not altogether in the same league with them.

JoshuaZ 12 April 2011 01:45:47AM 4 points

Ok. So all my World War 2 examples have now decisively been shown to be wrong. I don't have any other modern examples to give that go in this direction. All other modern examples go pretty strongly in the other direction. I withdraw the claim wholesale and am updating to accept the claim for post-enlightenment human societies.

Randaly 10 April 2011 01:10:46AM * 11 points

It's worth noting that most of the Nazi superiority in technology wasn't actually due to Nazi efforts, but rather due to a previous focus on technological and scientific development; for example, Germans won 14 of the first 31 Nobel Prizes in Chemistry, the vast majority of initial research into quantum mechanics was done by Germans, etc. But Nazi policies actually did actively slow down progress, by e.g. causing the emigration of free-thinking scientists like John von Neumann, Hans Bethe, Leo Szilard, Max Born, Erwin Schrodinger, and Albert Einstein, and by replacing empirically based science with [inaccurate political ideology](#). (Hitler personally believed that the stars were balls of ice, tried to avoid harmful "earth-rays" mapped out for him with a dowsing rod, and drank a toxic gun-cleaning fluid for its supposed health benefits, not to mention his bizarre racial theories.) Membership in the Society of German Natural Researchers and Physicians shrank nearly in two between 1929 and 1937; during World War II, nearly half of German artillery came from its conquered neighbors, its supply system relied in part on 700,000-2,800,000 horses, its tanks and aircraft were in many ways technologically inferior to those of many of its neighbors, etc.

"If they hadn't been ruled by an insane dictator they would have done much better. Similarly, if they had expanded just as much but waited to start the serious discrimination and genocide until after they already had won they would have likely won."

But that's Deutch's entire point- that that's what the "bad guys" do, what makes them the "bad guys". Sure if Hitler hadn't been Hitler, or somehow [not been human](#), German science wouldn't have been at a massive disadvantage. But I don't see much evidence that the "bad guys" have an advantage; at best, if you assume best case conditions and that the "bad guys" don't act like humans, you get an equal playing field.

(And we see similar things among the other "bad guys" of history- Lysenkoism, the Great Leap Forwards, etc.)

"Or to use a different, but potentially more controversial example, in North America and in Australia, the European colonizers won outright, despite having extremely violent, expansionist policies."

Conditions then no longer hold; nations are no longer isolated, the ideas of science/democracy/capitalism are fairly generally known, etc. And it's also worth noting that the colonizers have generally been transformed into "good guys".

Vladimir_M 12 April 2011 03:18:11AM 7 points

during World War II, nearly half of German artillery came from its conquered neighbors, its supply system relied in part on 7,000 horses,

According to [this article](#) published by the German Federal Archives, 2.8 million horses served in the German armed forces in WW2. The article also notes how successfully the German wartime propaganda portrayed the Wehrmacht as a high-tech motorized army, an image widely held in the public to this day, while in reality horses were its main means of transport.

Desrtopa 10 April 2011 04:05:05PM * 3 points

The first example that comes to mind for me is the collapse of the Roman empire. The Romans might have been "bad", being aggressive and expansionist, but the people they fell to were markedly worse from the perspective of truth seeking and pursuit of enlightenment, the standard Deutsch and curi are applying, and their replacements ushered in the Dark Ages.

Randaly 10 April 2011 08:25:58PM * 6 points

But different conditions hold today. The Gothic armies were virtually identical to the armies of the earlier Celts/Gauls who the Romans had crushed; even the Magyars (~1500's CE) used more or less the same tactics and organization as the Cimmerians (~ 700 BCE), though they did have stirrups, solid saddle trees, and stiff-tipped composite bows. Similarly, IIRC, the Roman armies didn't make use of any major recent technological innovations. This no longer holds today; the idea of an army using technology hundreds of years old being a serious military threat to any modern nation is frankly ludicrous. Technological and scientific development has become much, much more important than it was during Roman times.

(And, btw, it's not really accurate to say that, in practice, the barbarians were all that much much worse than the Romans in terms of development and innovation; [technological development in Europe didn't really slow down all that much during the Dark Ages](#) and the Romans had very few scientific (as opposed to engineering) advances anyways- most of their scientific knowledge (not to mention their mythology, art, architecture, etc.) was borrowed from the Greeks.)

Desrtopa 10 April 2011 08:31:54PM 0 points

Yes, but the culture of enlightenment and innovation within Greek and Roman culture had already been falling apart from within. The culture of Classical Antiquity was outcompeted by less enlightened memes.

Randaly 10 April 2011 09:27:11PM 1 point

How so? I'm not sure when, specifically, you're talking about, but the post-expansion Roman Empire still produced such noted philosophers as Marcus Aurelius, Apuleius, Boethius, St. Augustine, etc.

Desrtopa 10 April 2011 10:46:10PM 2 points

I'm thinking of the decline of Hellenist philosophy, especially the mathematical and empirical outlooks propounded by those such as [Hypatia](#).

Jayson_Virissimo 11 April 2011 07:13:01PM * 2 points

I'm thinking of the decline of Hellenist philosophy, especially the mathematical and empirical outlooks propounded by those such as Hypatia.

As far as I know, Hypatia was a [Neoplatonist](#) like [Saint Augustine](#). What evidence do you know of that she had an *empirical outlook*?

[continue this thread »](#)

Randaly 11 April 2011 04:34:50PM 1 point

Well of course the previously dominant branch of philosophy declined- that happens all the time in philosophy. But I don't think that there's grounds for proclaiming Hellenist philosophy to be significantly better than its successors: it was hardly empirical (Hypatia herself was an anti-empirical Platonist) and typically more concerned with e.g. confused explanations of the world in terms of a single property (all is fire! no, water!) or confusion regarding words (e.g. the Sorites paradox) than any kind of research valuable/relevant today.

And the group which continued the legacy of Hellenist/Roman thought, the Islamic world,

did in fact continue and, IMHO, vastly augment the level of empirical thought; for example, it's widely believed that the inventor of the Scientific Method was an Arab scientist, [Alhazen](#). Even though Europe saw a drop in learning due to the collapse of the unsustainable centralized Roman economy and the resulting wars and deurbanization, all that occurred was that its knowledge was passed onto new civilizations large/wealthy/secure enough to support science/math/philosophy. (Specifically, Persia and Byzantium, and later the Caliphates.)

[continue this thread »](#)

JoshuaZ 10 April 2011 01:15:29AM * 4 points [-]

You make a very strong case that the Nazi example does go in the other direction. I withdraw that example. If anything it goes strongly in favor of Deutsch's point.

I'm not convinced by the relevancy of your point about the historical state during the colonization of North America. The point is not whether or not someone eventually transformed, the point is that violent, expansionist groups can win over less expansionist groups.

curi 10 April 2011 01:23:50AM * 2 points [-]

Deutsch's definition of "the bad guys" is not the most expansionist groups.

He would regard the colonizers as the good guys (well, better guys) because their society was less static, more open to improvement, more tolerant of non-conformist people, more tolerant of new ideas, more free, and so on. There's a reason the natives had worse technology and their culture remained static for so long: they had a society that squashes innovation.

JoshuaZ 10 April 2011 01:27:26AM 4 points [-]

You'd have to convince me that they were more open to non-conformists. A major cause of the European colonization was flight of non-conforming groups (such as the Puritans) to North America where they then proceeded to persecute everyone who disagreed with them.

There's a reason the natives had worse technology and their culture remained static for so long: they had a society that squashes innovation.

I'm curious what you think of "Guns, Germs, and Steel" or similar works. What causes one society or another to adopt or even make innovations can be quite complicated.

Randaly 10 April 2011 09:09:15PM 7 points [-]

The Renaissance/much of modern science originated in Italy, not in England (thus, e.g. Galileo, da Vinci, etc.) And the Italian city-states of the time were fairly free: Pisa, Milan, Arezzo, Lucca, Bologna, Siena, Florence, and Venice were all at some point governed by elected officials. They were also remarkably meritocratic: as the influential Neapolitan defender of atomism [Francesco D'Andrea](#) put it, describing Naples:

There is no city in the world where merit is more recognized and where a man who has no other asset than his own worth can rise to high office and great wealth.

(Even if he's only boasting about his own city-state, it's significant that meritocracy was considered worth boasting about.)

Similarly, merchants, not priests, politicians, etc. were considered the highest status group: nobles up to and including national leaders (e.g. the Doge of Venice) dressed like merchants.

(Incidentally, the other factors you mentioned below also played a role: competition between city-states and the influence of outside science from Byzantium and the Islamic world showing what could be done. Nevertheless, Italian freedoms were also necessary: e.g. Galileo was only

able to publish his ideas because he lived in the free Republic of Venice, where Jesuits were banned and open inquiry encouraged; he was persecuted and forced to recant his theories when he moved to Tuscany.)

curi 10 April 2011 01:46:29AM * -1 points 

read *The Beginning of Infinity* by Deutsch. It discusses that Diamond book and other similar works.

Yes European society was not favorable to non-conformists. One period I've studied, which is *later* (so, i think, *better* in this regard) is around 1790 ish. At that time, to take one example, the philosopher william godwin's wife died in childbirth and he published memoirs and people got really pissed off because she had had sex out of wedlock and stuff along those lines. when godwin's daughter ran off with shelley there were rumors he had sold her. meanwhile, for example, there was lots of discrimination against irish catholics. i know some stuff about how biased and intolerant people can be.

but what i also know is a bit about static societies (again, see the book for more details, or at least check out my website, e.g. <http://fallibleideas.com/tradition>).

when a society doesn't change for thousands of years that means it's even harsher than the european society i was talking about. preventing change for such a long period is hard. stuff is done to prevent it. the non-conformists don't even get off the ground. everyone's spirits are squashed in childhood -- thoroughly -- and so the adults don't rebel at all. if there were adults who were eccentric then the society simply wouldn't stay the same so long. european society was already getting fairly near fairly rapid changes (e.g. industrial revolution) when it started colonizing the new world.

JoshuaZ 10 April 2011 02:16:33AM * 0 points 

Minor remark: Your essay about tradition is much more readable than a lot of the other material on your site. I'm not sure why but if you took a different approach to writing/thinking about it, you might want to apply that approach elsewhere.

curi 10 April 2011 03:59:28AM 1 point 

I think the difference is you. I wrote that entire site in a short time period. I regard it as all being broadly similar in style and quality. I attempted to use the same general approach to the whole site; I didn't change my mind about something midway. I think it's a subject you understand better than epistemology directly (it is about epistemology, indirectly. traditions are long lived knowledge). The response I've had from other readers has varied a lot, not matched your response.

I do know how to write in a variety of different styles, and have tried each in various places. The one I've used here in the last week is not the best in various senses. But it serves my purpose.

JoshuaZ 10 April 2011 02:01:12AM * 4 points 

when a society doesn't change for thousands of years that means it's even harsher than the european society i was talking about.

This doesn't follow. (Incidentally, I don't know why you sometimes drop back to failing to capitalize but it makes what you write much harder to read.) For example, if one doesn't have good nutrition then people won't be as smart and so won't innovate. Similarly, if one doesn't have free time people won't innovate. Some technologies and cultural norms also reinforce innovation. For example, having a written language allows a much larger body of ideas, and having market economies gives market incentives to coming up with new technologies.

Moreover, innovation can occur directly through competition. When you are convinced that your religion or tribe is the best and that you need to beat the others by any means necessary you'll do a lot better at innovating.

There's also a self-reinforcing spiral: the more you innovate the more people think that innovation is possible. If your society hasn't changed much then there's no reason to think that new technologies are easy to find.

There's no reason to think that Native American populations were systematically preventing change. There's a very large difference between having infrastructural and systemic issues that make the development of new technologies unlikely and the claim that "everyone's spirits are squashed in childhood -- thoroughly".

curi 10 April 2011 04:12:26AM -9 points



(Incidentally, I don't know why you sometimes drop back to failing to capitalize but it makes what you write much harder to read.)

I don't know either. I have noticed that I will often stop using capitals in parentheses, even if they contain multiple sentences or words that are supposed to be capitalized like "I". (you can see in the first parenthetical, and this one, missing capitalization, even though that first parenthetical in my previous comment is in a section of text where, otherwise, I was capitalizing.) I don't really care. I can capitalize when I want to impress people. Here I do not wish to impress. I want to filter people. If they can't look past some capitalization -- if they are shallow -- then let them dislike me and we'll go our separate ways quickly. You can, btw, looking through my history see that I've asked people tangential questions sometimes which might be taken as rude or aggressive. It's again for filtering purposes. I don't regard offending a portion of the people here as a bad thing, but a good thing. Then when a few people like me better and keep talking with me, my tone changes somewhat, and I'll write stuff like this which is more open, cooperative and non-confrontational. Then one thing that will happen is other people, who I didn't write this for, will jump in and find it arrogant, condescending, and so on. But I think you (JoshuaZ) might appreciate these remarks. No guarantees, but worth a try.

For example, if one doesn't good nutrition then people won't be as smart and so won't innovate. Similarly, if one doesn't have free time people won't innovate.

Where does free time come from? Where does better nutrition come from? Ideas.

Here's an example from Bol: llamas. South America had llamas. Why didn't they spread? Why didn't they get sold to distant towns, and bred to have more, and used to save tons of labor and create more free time? It's not for lack of suitable animals that people were doing more hand labor in some places than others. It's for lack of ideas.

Some technologies and cultural norms also reinforce innovation. For example, having a written language allows a much larger body of ideas, and having market economies gives market incentives to coming up with new technologies.

Yes, that's just my point. Things like written languages, technological ideas, and progress cultural norms aren't natural resources provided by Nature. They are ideas people have. And they make all the difference.

[continue this thread »](#)

curi 09 April 2011 10:50:37PM * -1 points



This sort of claim seems to run into historical problems

Athens lost to sparta. But it was a close call. Sparta excelled at nothing but war. Athens spread its efforts around and was good at everything. And it was close! That's how much more powerful Athens was: it did

tons of other stuff and nearly won the war anyway.

If Athens had had an extra 100 years to improve, it would have gotten a big lead on Sparta. Long term, that kind of society wins.

A lot of major expansionist violent empires have done quite well for themselves.

Not long term.

Or to use a different, but potentially more controversial example, in North America and in Australia, the European colonizers won outright, despite having extremely violent, expansionist policies.

They were up against closed societies that were much worse than they themselves were in pretty much every respect including morally. The natives were not non-violent philosophers.

curi 09 April 2011 10:00:04PM * 2 points



Merely pulling the trigger less often doesn't change the inevitability of doom. [...] One of the most important uses of technology is to counteract disasters and to recover from disasters, both from foreseen and unforeseen evil. Therefore, the speed of progress itself is one of the things that is a defense against catastrophe.

The idea is: if you're going to pull the trigger once every 100 years, instead of once every 5, and it's a 2% chance of doom each time, you're still doomed eventually. Any static society is doomed in that way. The delays don't help anything because nothing is changing in the mean time, so eventually doom happens.

The attitude of not making progress, but just trying to sustain a fixed lifestyle forever, cannot work. Even if the chance of doom per year is made low, there is some chance so it will have to destroy them eventually. There's nothing to stop it from doing so.

It's only in a dynamic society creating new knowledge and progress that lasting longer matters to whether you're doomed eventually, b/c in that extra time more progress is made.

curi 09 April 2011 08:55:15PM * 1 point



(If it's unknowable, how can we know that a certain prediction strategy is going to be systematically biased in a known direction? Biased with respect to what knowable standard?)

I forget how much detail there is on this later in this talk, but it is in his book. The systematic bias towards pessimism is due to the method of trying to imagine the future using today's knowledge (which is less than the future's knowledge).

Quoting Deutsch from *The Beginning of Infinity*:

Trying to know the unknowable leads inexorably to error and self-deception. Among other things, it creates a bias towards pessimism. For example, in 1894, the physicist Albert Michelson made the following prophecy about the future of physics:

The more important fundamental laws and facts of physical science have all been discovered, and these are now so firmly established that the possibility of their ever being supplanted in consequence of new discoveries is exceedingly remote. ... Our future discoveries must be looked for in the sixth place of decimals. (Albert Michelson, address at the opening of the Ryerson Physical Laboratory, University of Chicago, 1894)

What exactly was Michelson doing when he judged that there was only an 'exceedingly remote' chance that the foundations of physics as he knew them would ever be superseded? He was prophesying the future. How? On the basis of the best knowledge available at the time. But that consisted of the physics of 1894! Powerful and accurate though it was in countless applications, it was not capable of predicting the content of its successors. It was poorly suited even to imagining the changes that relativity and quantum theory would bring – which is why the physicists who did imagine them won Nobel prizes. Michelson would not have put the expansion of the universe, or the existence of parallel universes, or the non-existence of the force of gravity, on any list of possible discoveries whose probability was 'exceedingly remote'. He just didn't conceive of them at all.

FAWS 10 April 2011 01:51:08AM * 8 points [-]

It's inconsistent to expect the future to be better than one expects. If you think your probability estimates are too pessimistic adjust them until you don't know whether they are too optimistic or too pessimistic. No one stops you from assigning probability mass to outcomes like "technological solution that does away with problem X" or "scientific insight that makes the question moot". Claimed knowledge that the best possible probability estimate is biased in a particular direction cannot possibly ever be correct.

Vladimir_Nesov 09 April 2011 07:29:57PM * 1 point [-]

From the very beginning of the talk:

I don't have to persuade you that, for instance, life is better than death; and I don't have to explain exactly why knowledge is a good thing, and that the alleviation of suffering is good, and communication, and travel, and space exploration, and ever-faster computers, and excellence in art and design, all good.

One of these things is not like the others.

lukeprog 09 April 2011 07:50:33PM 3 points [-]

Ever-faster computers jumped out at me when I first heard that sentence.

Matt_Simpson 11 April 2011 12:50:28AM 0 points [-]

me too. Instrumental vs terminal values.

JoshuaZ 09 April 2011 09:14:35PM 0 points [-]

Really? The comment about art and design jumped out at me.

curi 09 April 2011 09:17:00PM * 1 point [-]

FYI DD's talk on why flowers are beautiful:

<http://193.189.74.53/~qubitor/people/david/index.php?path=Video/Why%20Are%20Flowers%20Beautiful>

That URL is weird. In case it breaks, it's on youtube in parts:

<http://www.youtube.com/watch?v=56o2n8sVvM8>

curi 09 April 2011 07:31:25PM 5 points [-]

Which?

JGWeissman 09 April 2011 07:17:00PM 0 points [-]

How was Curi able to post this without having 20 karma?

curi 09 April 2011 07:18:34PM * 3 points [-]

I had 20 karma. I don't anymore. My karma has had a lot of fluctuations.

edit: see. back to 21 now.

CarlShulman 09 April 2011 07:10:08PM * 7 points 

This should not have been made as a top-level post without some more explanation to let people evaluate whether to watch the video.

curi 09 April 2011 07:13:46PM -1 points 

I don't want to bias the reactions.

JoshuaZ 09 April 2011 09:17:14PM 5 points 

Unfortunately, this attitude and your decision to put this in main rather than the discussion section is getting it downvoted. That will likely continue. Moreover, downvotes for main section articles hurt a lot more than downvotes in the discussion section. I strongly urge you to move this into the discussion section where it will be considered a much more reasonable post.

curi 10 April 2011 08:35:08PM -11 points 

Unfortunately, this attitude and your decision to put this in main rather than the discussion section is getting it downvoted. That will likely continue.

It didn't. It made it back up to a score of 0.

Learn anything? Or since you only said "likely" will you say that your prediction isn't contradicted by the result. Is never actually being contradicted by evidence one of the main appeals of only saying stuff is likely instead of conjecturing that it's true?

Randaly 11 April 2011 04:19:03PM 0 points 

As far as I know, the minimum possible karma is zero; scores below that are, IIRC, displayed as zero.

Sniffnoy 12 April 2011 02:14:13AM * 1 point 

Scores in the negative are kept track of despite not being displayed, however. In particular, people with negative karma have a commenting frequency limit, whereas people with zero karma do not.

JoshuaZ 11 April 2011 04:22:58PM 0 points 

Scores for total karma are displayed as zero if they are negative. Scores for individual articles can be negative (and in fact it is back to -1). I have various hypotheses about why the score has moved up but I'm waiting to gather more evidence before I state them.

PhilGoetz 10 April 2011 12:44:35AM * 0 points 

Everyone who downvotes links posted in the main section because they think it's a cheap way to get karma - you can just choose not to vote for them. Thus, trying to discourage people from posting to the main page for karma reasons is trying to make karma voting decisions *for other LWers*.

Karma is supposed to indicate which articles and comments are worth reading. Karma doesn't function to tell people whose opinions to respect, so people should stop worrying that other people are getting easy karma. Trust me - I have 15,000 karma, and people don't cut me any more slack than when I had none.

JoshuaZ 10 April 2011 12:51:26AM 2 points 

Cur's karma has repeatedly dropped low enough that his posting rate is moderated. If that's going to happen then it should occur based on the quality of posts not to him being socially tone-deaf about community norms of where to post things.

(Incidentally, there's another reason to downvote short link posts and the like in the main section- some people just have the RSS feed for the main posts and don't want every little link to show up).

PhilGoetz 10 April 2011 12:56:32AM * 2 points

Curi's karma has repeatedly dropped low enough that his posting rate is moderated. If that's going to happen then it should occur based on the quality of posts not to him being socially tone-deaf about community norms of where to post things.

That's Curi's decision.

(Incidentally, there's another reason to downvote short link posts and the like in the main section- some people just have the RSS feed for the main posts and don't want every little link to show up).

Okay - a valid reason.

I would still like to say that, when considering whether to impose a social norm against posting certain things on the main page, saying that you think they're unworthy of karma is not a good reason, because (a) karma point accumulation to users beyond getting enough to post does not give them any advantage, and (b) you can choose not to vote, and therefore you can object only because you don't trust the judgement of other users on LW and so would like to deprive them of the freedom to vote for such articles.

This may not have been your reason, but this seemed like a good place to make my point.

pjebj 09 April 2011 07:30:11PM 13 points

You might want to move this to the discussion section, then; unadorned links like this are generally not considered appropriate to the main LW section.

(You can move it by editing the article, then changing where it's being published to.)

Dorikka 09 April 2011 08:33:41PM 0 points

Yep. I would downvote this, but it's already invisible on the top-level page.

Larks 09 April 2011 11:21:36PM 0 points

The main page is for things we think of sufficient quality that they're worth the time and cognitive effort of reading. Is this worth an hour of your time to read? If not, it should be downvoted to invisibility.

Dorikka 09 April 2011 11:32:38PM 1 point

As of now and when I first saw the post appear on the sidebar, it is/was invisible on the main page and visible only through the sidebar.

Larks 09 April 2011 11:34:46PM 1 point

Yup.

It's worth noting in general that the 'main page' is actually the 'promoted' page, which requires an admin to move you there. But you're right, the article is also not visible on the 'new' page either.

As usual with tradeoffs like these, one has to decide on a policy that is willing to tolerate some of one type of error to keep the error you care about to some desired level.

I think a good heuristic for deciding who is an expert and who is an idiot with references is credentialism. But credentialism has a bad brand here, due to a "love affair with amateurism" LW has. One of the consequences of this love affair is a lot of folks here make the above trade off badly (in particular they ignore legit advice to read way too frequently).

curi 20 November 2017 09:15:59PM * 0 points 

Here's a tricky example of judging authority (credentials). You say listen to SA about QM. Presumably also listen to David Deutsch (DD), who knows more about QM than SA does. But what about me? I have talked with DD about QM and other issues at great length and I have a very accurate understanding of what things I can say about QM (and other matters) that are what DD would say, and when I don't know something or disagree with DD. (I have done things like debate physics, with physicists, many times, while being advised by DD and him checking all my statements so I find out when I have his views right or not.) So my claims about QM are about as good as DD's, when I make them – and are therefore even better than SA's, even though I'm not a physicist. Sorta, not exactly. Credentials are complicated and such a bad way to judge ideas.

What I find most people do is decide what they want to believe or listen to first, and then find an expert who says it second. So if someone doesn't want to listen, credentials won't help, they'll just find some credentials that go the other way. DD has had the same experience repeatedly – people aren't persuaded due to his credentials. That's one of the main reasons I'm here instead of DD – his credentials wouldn't actually help with getting people here to listen/understand. And, as I've been demonstrating and DD and I already knew, arguments aren't very effective here either (just like elsewhere).

And I, btw, didn't take things on authority from DD – I asked questions and brought up doubts and counter-arguments. His credentials didn't matter to me, but his arguments did. Which is why he liked talking with me!

Lumifer 20 November 2017 10:15:44PM 1 point 

So my claims about QM are ... even better than SA's

ROFL

That's one of the main reasons I'm here instead of DD

And here I was, completely at loss as to why David Deutsch doesn't hang out at LW... But now we know.

Yes, the article I was looking for.

by **fblogin** on Double Crux — A Strategy for Resolving Disagreement | 0 points

Sorry, but it is. Simple test: open a
by **elharo** on LW 2.0 Open Beta Live | 0 points

[Lucifer's version]

by **roland** on The "Outside the Box" Box | 0 points

Interesting discussion. Does

by **kimberchoi** on Actions and Words: Akrasia and the Fruit of Self-Knowledge | 0 points

RECENT POSTS

LW 2.0 Open Beta Live

by Vaniver | 22v (36c)

In support of yak shaving part 2

by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag

by Raemon | 41v (106c)

European Community Weekend 2017

by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about *protecting* the epistemology

by AnnaSalamon | 73v (54c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"

by AnnaSalamon | 36v (43c)

Be secretly wrong

by Benquo | 32v (47c)

CFAR's new focus, and AI Safety

by AnnaSalamon | 30v (88c)

curi 20 November 2017 10:18:05PM * 0 points [-]

you're mean and disruptive. at least you're demonstrating why credentials are a terrible way to address things, which is my point. you just assume the status of various credentials without being willing to think about them, let alone debate them (using more credentials (regress), or perhaps arguments? but if arguments, why not just use those in the first place?). so for you, like most people, using credentials = using bias.

Lumifer 21 November 2017 03:33:50PM 0 points [-]

| you're mean and disruptive

Woo, kindergarten flashbacks!

| you just assume the status of various credentials without being willing to think about them

Am I? Please demonstrate.

| using credentials = using bias

What do you mean by bias? In statistics bias is one of those things you trade off against other things (like variance). Being unbiased is not always optimal.

gjm 21 November 2017 10:18:25PM 0 points [-]

Yeah, credentials are a poor way of judging things. But that first paragraph doesn't show remotely what you think it does.

Some of David Deutsch's credentials that establish him as a credible authority on quantum mechanics: He is a physics professor at a leading university, a Fellow of the Royal Society, is widely recognized as a founder of the field of quantum computation, and has won some big-name prizes awarded to eminent scientists.

Your credentials as a credible authority on quantum mechanics: You assure us that you've talked a lot with David Deutsch and learned a lot from him about quantum mechanics.

Fact Posts: How and Why

by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement

by Duncan_Sabien | 61v (103c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all

by vaultDweller on Rationality Quotes

April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

IlyaShpitsner (22)

alanforr (22)

gjm (19)

Elo (17)

Lumifer (16)

curi (15)

cousin_it (7)

gwern (5)

mirefek (5)

OrthernLight (5)

Oscar_Cunningham (5)

Dagon (4)

JenniferRM (4)

morganism (4)

jimrandomh (4)

RECENT KARMA AWARDS

This is *not how credentials work*. Leaving aside what useful information (if any) they impart: when it comes to quantum mechanics, David Deutsch has credentials *and you don't*.

It's not clear to me what argument you're actually making in that first paragraph. But it seems to *begin* with the claim that you have good credentials when it comes to quantum mechanics for the reasons you recite there, and that's flatly untrue.

IlyaShpitser 22 November 2017 03:24:07PM * 0 points [-]

| Yeah, credentials are a poor way of judging things.

They are not, though. It's standard "what LW calls 'Bayes' and what I call 'reasoning under uncertainty'" -- you condition on things associated with the outcome, since those things carry information. Outcome (O) -- having a clue, thing (C) -- credential. $p(O | C) > p(O)$, so your credence in O should be computed after conditioning on C, on pain of irrationality. Specifically, the type of irrationality where you leave information on the table.

You might say "oh, I heard about how argument screens authority." This is actually not true though, even by "LW Bayesian" lights, because you can never be certain you got the argument right (or the presumed authority got the argument right). It also assumes there are no other paths from C to O except through argument, which isn't true.

It is a foundational thing you do when reasoning under uncertainty to condition on everything that carries information. The more informative the thing, the worse it is not to condition on it. This is not a novel crazy thing I am proposing, this is bog standard.

The way the treatment of credentialism seems to work in practice on LW is a reflexive rejection of "experts" writ large, except for an explicitly enumerated subset (perhaps ones EY or other "recognized community thought leaders" liked).

This is a part of community DNA, starting with EY's stuff, and Luke's "philosophy is a diseased discipline."

That is crazy.

gjm 22 November 2017 05:16:07PM 0 points [-]

| They are not, though.

Actually, I somewhat agree, but being an agreeable sort of chap I'm willing to concede things *arguendo* when there's no compelling reason to do otherwise :-), which is why I said "Yeah, credentials are a poor way of judging things" rather than hedging more.

More precisely: I think credentials very much can give you useful information, and I agree with you that argument does not perfectly screen off authority. On the other hand, I agree with prevailing LW culture (perhaps with you too) that credentials typically give you *very imperfect* information and that argument does *somewhat* screen off authority. And I suggest that how much credentials tell you may vary a great deal by discipline and by type of credentials. Example: the Pope has, by definition, *excellent* credentials of a certain kind. But I don't consider him an authority on *whether any sort of gods exist* because I think the process that gave him the credentials he has isn't sufficiently responsive to that question. (On the other hand, that process is highly responsive to *what Catholic doctrine is* and I would consider the Pope a very good authority on that topic even if he didn't have the ability for control that doctrine as well as reporting it.)

It seems to me that e.g. physics has norms that tie its credentials pretty well (though not perfectly) to actual understanding and knowledge; that philosophy doesn't do this so well; that theology does it worse; that homeopathy does it worse still. (This isn't just about the moral or cognitive excellence of the disciplines in question; it's also that it's harder to tell whether someone's any good or not in some fields than in others.)

IlyaShpitsner 22 November 2017 09:25:26PM * 0 points [-]

I guess the way I would slice disciplines is like this:

(a) Makes empirical claims (credences change with evidence, or falsifiable, or [however you want to define this]), or has universally agreed rules for telling good from bad (mathematics, theoretical parts of fields, etc.)

(b) Does not make empirical claims, and has no universally agreed rules for telling good from bad.

Some philosophy is in (a) and some in (b). Most statistics is in (a), for example.

Re: (a), most folks would need a lot of study to evaluate claims, typically at the graduate level. So the best thing to do is get the lay of the land by asking experts. Experts may disagree, of course, which is valuable information.

Re: (b), why are we talking about (b) at all?

curi 20 November 2017 08:56:22PM 0 points [-]

|"The amount of energy necessary to refute bullshit is an order of magnitude bigger than to produce it."

i think this is false, and is an indication of using the wrong methods to refute bullshit – the right methods *reuse* refutations of *categories* of bad ideas. do you have some comprehensive argument that it must be true?

i find it disturbing how much people here are in favor of judging ideas by sources instead of content – credentialism. that's pretty pure irrationality. also debating which credentials are worth how much is a bad way to approach discussions, but it's totally non-obvious and controversial which credentials are how good even for standard credentials like PhDs from different universities.

phonymapercut 21 November 2017 04:00:35AM 0 points [-]

Is English your first language?

Lumifer 20 November 2017 07:40:52PM 0 points [-]

|(in particular they ignore legit advice to read way too frequently)

The context matters. If you are trying to figure out how X actually works you probably should go read or at least scan the relevant books even if no one is throwing references at you. On the other hand, if you're just procrastinating by engaging in a Yet Another Internet Argument with zero consequences for your life, going off to read the references is just a bigger waste of time.

LessWrong DISCUSSION

MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

You're looking at Less Wrong's discussion board. This includes all posts, including those that haven't been promoted to the front page yet. For more information, see [About Less Wrong](#).

[Link] Simple refutation of the 'Bayesian' philosophy of science

1 curi 01 November 2017 06:54AM

[Comments \(30\)](#)

Tags: [deutsch](#) [popper](#)

Comments (30)

Sort By: New

Enter a comment here

Comment

Show help

Dagon 01 November 2017 09:54:23PM 0 points

(1) the objective of science is, or should be, to increase our 'credence' for true theories

Well, no. Theories are maps, and are by necessity simpler than the territory (the universe is it's own best model). There is no such thing as a "true" theory. There are only theories which predict a larger or smaller subset of future states better or worse than others.

Manfred 02 November 2017 08:55:20PM 0 points

I think this neglects the idea of "physical law," which says that theories can be good when they capture the dynamics and building-blocks of the world simply, even if they are quite ignorant about the complex initial conditions of the world.

Dagon 03 November 2017 04:29:36PM 1 point

Sure. This is true of all maps and models. As simple as possible, but no simpler.

That simplicity *ALWAYS* comes with a loss of fidelity to the actual state of the universe.

curi 02 November 2017 01:24:47AM 0 points

I disagree with viewing theories as predictive. Deutsch calls that instrumentalism and refutes in his book, *The Fabric of Reality*, in chapter 1. The basic problem is predictions aren't explanations about what's going on (the causality behind the prediction) or *why*.

Yet some philosophers — and even some scientists — disparage the role of explanation in science. To them, the basic purpose of a scientific theory is not to explain anything, but to predict the outcomes of experiments: its entire content lies in its predictive formulae. They consider that any consistent explanation that a theory may give for its predictions is as good as any other — or as good as no explanation at all — so long as the predictions are true. This view is called *instrumentalism* (because it says that a theory is no more than an 'instrument' for making predictions). To instrumentalists, the idea that science can enable us to understand the underlying reality that accounts for our observations is a fallacy and a conceit. They do

Custom Search

curi

Karma Score

16

6

✉

Create new article

Submit a new link

Add new meetup

Preferences

Log out

Website <https://elliottemple.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

Create

 [Subscribe to RSS Feed](#)

NEAREST MEETUPS

San Francisco Meetup: Shallow Questions: 06 November 2017 06:15PM

San Francisco Meetup: Shallow Questions: 06 November 2017 06:15PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
[Less Wrong Study Hall](#)

RECENT COMMENTS

> [This statement indicates to me](#) by Lumifer on Questions about AGI's Importance | 0 points

>> [We disagree about e.g.](#) by curi on Questions about AGI's Importance | 0 points

> [I'm trying to discuss the](#) by Lumifer on Questions about AGI's Importance | 0 points

[I'm trying to discuss the](#) by curi on Questions about AGI's Importance | 0 points

> [When I rewrite canonical material](#) by Lumifer on Questions about AGI's Importance | 0 points

not see how anything a scientific theory may say beyond predicting the outcomes of experiments can be more than empty words. Explanations, in particular, they regard as mere psychological props: a sort of fiction which we incorporate in theories to make them more easily remembered and entertaining.

(Deutsch goes on at too much length to paste.)

Dagon 03 November 2017 11:44:06PM 0 points [-]

To instrumentalists, the idea that science can enable us to understand the underlying reality that accounts for our observations is a fallacy and a conceit

"understand" is doing a lot of work in this. What does it mean beyond "ability to make predictions conditional on future actions"?

curi 04 November 2017 01:47:30AM 0 points [-]

teaching you things like what "understand" means is a large task. are you willing to put in effort by e.g. reading a book chapter, and answering questions to identify what you do and don't already understand about the matter?

Dagon 04 November 2017 04:34:45AM 0 points [-]

Almost certainly not. I take this as confirmation that "understand" is the key misleadingly-simple word in your quote.

entirelyuseless 04 November 2017 05:45:42PM 0 points [-]

Not at all. It means the ability to explain, not just say what will happen.

Dagon 04 November 2017 07:28:54PM 0 points [-]

When you say "ability to explain", I hear "communicate a model that says what will happen (under some set of future conditions/actions)".

There is no such thing as "why" in the actual sequence of states of matter in the universe. It just is. Any causality is in the models we use to predict future states. Which is really useful but not "truth".

entirelyuseless 05 November 2017 11:28:55PM 0 points [-]

I hear "communicate a model that says what will happen (under some set of future conditions/actions)".

You're hearing wrong.

curi 04 November 2017 05:25:02PM 0 points [-]

it's not, i don't know why you're making a stink about it. i think you just wanted indirect evidence to convince yourself to stop conversing and be able to blame me in your head.

OrthernLight 01 November 2017 02:46:31PM 1 point [-]

(Epistemic status: sufficiently abstract that I can't be very confident without more familiarity with the topic)

(1) the objective of science is, or should be, to increase our 'credence' for true theories

RECENT POSTS

[\[Link\] AGI](#)
by curi | 0v (0c)

[\[Link\] Kialo -- an online discussion platform that attempts to support reasonable debates](#)
by mirefek | 1v (6c)

[\[Link\] Intent of Experimenters; Halting Procedures; Frequentists vs. Bayesians](#)
by curi | 1v (0c)

[\[Link\] Intercellular competition and the inevitability of multicellular aging](#)
by Gunnar_Zarncke | 1v (3c)

[Announcing the AI Alignment Prize](#)
by cousin_it | 5v (2c)

[Problems as dragons and papercuts](#)
by Elo | 1v (1c)

[November 2017 Media Thread](#)
by ArisKatsaris | 1v (16c)

[\[Link\] Simple refutation of the 'Bayesian' philosophy of science](#)
by curi | 1v (30c)

[\[Link\] Why Competition in The Politics Industry Is Failing America - pdf](#)
by morganism | 0v (1c)

[Questions about AGI's Importance](#)
by curi | 0v (57c)

LATEST OPEN THREAD

[By the way, is there an explanation](#)
by OrthernLight on Open thread, October 30 - November 5, 2017 | 3 points

LATEST RATIONALITY DIARY

[I learned something new. As a](#)
by Applesauce on Group Rationality Diary, February 2017 | 4 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

ingres (30)
gwern (26)
Elo (21)
turchin (18)
Lumifer (17)
IlyaShpitser (15)
SaidAchmiz (15)
Stuart_Armstrong (14)
Dagon (13)
Yosarian2 (11)
leohmarruda (10)

entirelyuseless (10)
 fowlertm (9)
 Raemon (8)
 Habryka (7)

RECENT KARMA AWARDS

I would suggest that it should also decrease our credence in false theories, and allow us to correctly estimate the likelihood of conjectures not yet proven or disproved.

However, if T is an explanatory theory (e.g. 'the sun is powered by nuclear fusion'), then its negation $\sim T$ ('the sun is not powered by nuclear fusion') is not an explanation at all.

Well, no - it's a set of explanations. A very large set, consisting of every explanation other than 'the sun is powered by nuclear fusion', but smaller than $T \mid \sim T$, and therefore somewhat useful, however slightly.

Therefore, suppose (implausibly, for the sake of argument) that one could quantify 'the property that science strives to maximise'.

Per the first line, we are supposing this property to be 'our credence in true theories'

If T had an amount q of that, then $\sim T$ would have none at all, not $1-q$ as the probability calculus would require if q were a probability.

All else being equal, if we come to believe T, our credence in true theories will be higher by $1-p$, where p is our previous credence in T. If we come to believe $\sim T$, our credence in true theories will be lower than if we were uncertain by p.

I'm not sure that it makes sense in this context to assign a value of 'the property that science strives to maximise' to a statement. It's not a property of statements alone but of our belief in them.

If you want to assign a value of q to near-absolute confidence in T, I would say that it's $1-\epsilon$. Thus, the $\sim t$ has near-zero value as far as the objective of science is concerned, and also has $1-q+\epsilon=0+\epsilon$ as the laws of probability demand.

Also, the conjunction ($T_1 \& T_2$) of two mutually inconsistent explanatory theories T_1 and T_2 (such as quantum theory and relativity) is provably false, and therefore has zero probability. Yet it embodies some understanding of the world and is definitely better than nothing.

(Assuming for the sake of example that quantum theory and relativity mutually inconsistent, but both likely,) $T_1 \& T_2$ is provably false, and indeed idea that quantum theory and relativity are both true is nonsense. $T_1 \mid T_2$, on the other hand, embodies some understanding of the world and is definitely better than nothing.

curi 01 November 2017 06:55:29PM 1 point



Well, no - it's a set of explanations. A very large set, consisting of every explanation other than 'the sun is powered by nuclear fusion', but smaller than $T \mid \sim T$, and therefore somewhat useful, however slightly.

Infinity minus one isn't smaller than infinity. That's not useful in that way.

It may be useful in some way. But just ruling a single thing out, when dealing with infinity, isn't a road to progress.

indeed idea that quantum theory and relativity are both true is nonsense

He's saying we use them both, and that has value, even though we know there must be some mistake somewhere. Saying "or" misrepresents the current situation. Both of them seem to be partly right. The situation (our current understanding which has value) looks nothing like we'll end up keeping one and rejecting the other.

Manfred 02 November 2017 09:07:28PM 0 points



Infinity minus one isn't smaller than infinity. That's not useful in that way.

The thing being added or subtracted is not the mere number of hypotheses, but a measure of the likelihood of those hypotheses. We might suppose an infinitude of mutually exclusive theories of the world, but most of them are *extremely* unlikely - for any degree of unlikeliness, there are an infinity of theories less likely than that! A randomly-chosen theory is so unlikely to be true, that if you add up the likelihoods of every single theory, they add up to a number less than infinity.

It is for this reason that it is important when we divide our hypotheses between something likely, and everything else. "Everything else" contains infinite possibilities, but only finite likelihood.

curi 02 November 2017 09:54:30PM 0 points

Well, no - it's a set of explanations. A very large set, consisting of every explanation other than 'the sun is powered by nuclear fusion', but smaller than $T \mid \sim T$, and therefore somewhat useful, however slightly.

This was talking about set sizes, which is what I replied about.

You can't quantify your fallibility in the sense of knowing how likely you are to be mistaken in an unexpected way. That's not possible.

OrthernLight 02 November 2017 12:45:06AM 0 points

He's saying we use them both, and that has value, even though we know there must be some mistake somewhere. Saying "or" misrepresents the current situation. Both of them seem to be partly right. The situation (our current understanding which has value) looks nothing like we'll end up keeping one and rejecting the other.

I haven't much knowledge of physics, and though that he was discussing the idea of two mutually exclusive theories which we use both of. From what you're saying, it sounds more like the crucial point is that they are presumably false, but still useful. Is that a good description of the situation?

As far as partly right theories that have value: if we know quantum theory is not completely right, then we've ruled out the hypothesis 'quantum mechanics' and are now dealing with the hypothesis space of theories relevantly similar to quantum theory. So I agree that I theory known to be inaccurate in some cases can be useful, but by treating it as a piece of evidence towards the truth, which is rather different than how we treated it when we thought it could be true in its own right.

curi 02 November 2017 01:35:01AM 0 points

Relativity and QM contradict but we don't know which is mistaken or why. Either one, individually, could be true in its own right.

OrthernLight 02 November 2017 05:39:49PM 0 points

Relativity and QM contradict but we don't know which is mistaken or why. Either one, individually, could be true in its own right.

The situation (our current understanding which has value) looks nothing like we'll end up keeping one and rejecting the other.

I don't see how these two statements can be consistent. If either one, individually, could be true in its own right, then why wouldn't we won't end up keeping one? If they contradict, then why wouldn't we reject the other?

curi 02 November 2017 05:48:47PM 0 points

i expect we'll keep parts of both.

OrthernLight 02 November 2017 06:19:56PM 0 points

As far as partly right theories that have value: if we know quantum theory is not completely right, then we've ruled out the hypothesis 'quantum theory' and are now dealing with the hypothesis space of theories that share some parts with quantum theory.

T in this case is not atomic; it is itself a conjunction of a lot of statements. So I agree that I theory known to be inaccurate in some cases can be useful, in that it may contain some true components as well as some untrue ones. But this is rather different than how we treated it when we thought it could be true in its own right.

In general, I agree that there are certain ideas in science that aren't propositions in Bayesian sense, and that treating them as if they were is a serious mistake. I don't think that this means that there's something wrong with the probability calculus, however.

curi 02 November 2017 06:57:03PM 0 points

As far as partly right theories that have value: if we know quantum theory is not completely right, then

But, again, we don't know that. QM could be right.

OrthernLight 02 November 2017 09:55:17PM 0 points

QM could be right.

Relativity and QM contradict

The situation (our current understanding which has value) looks nothing like we'll end up keeping one and rejecting the other.

I don't see how these statements can be consistent.

...if relativity and QM contradict, and QM turns out to be right, I'd expect us to reject relativity. Do you agree?

Lumifer 01 November 2017 07:08:04PM * 0 points

But just ruling a single thing out, when dealing with infinity, isn't a road to progress.

We don't deal with infinities. When asked "what sun is powered by?", humans formulate a **finite**, typically small, set of hypotheses, e.g.

- By nuclear fusion
- By a burning woodpile
- By elven magic
- By *something else*

Ruling even a single thing out from this small set is quite useful.

If you manage to rule out everything but *something else*, that's the most exciting time in science because you're now in uncharted territory (where every true scientist wants to be) and might be on a verge of a major breakthrough.

curi 01 November 2017 07:45:58PM 0 points

DD:

However, if T is an explanatory theory (e.g. 'the sun is powered by nuclear fusion'), then its negation $\sim T$ ('the sun is not powered by nuclear fusion') is not an explanation at all.

Ideas don't negate to all the alternatives humans are currently interested in. That isn't how logic works.

Lumifer 01 November 2017 08:04:39PM 0 points

It is not an explanation, but it is a (potentially) useful statement which leads you closer to an explanation. And I don't see any logical problems here (notice the *something else* alternative).

In any case, the underlying issue is hypothesis generation and any purely Bayesian view of science is necessarily incomplete because St.Bayes says absolutely nothing about how to generate hypotheses.

curi 01 November 2017 09:11:43PM 0 points

I agree that ruling statements like you talk about out is useful – I just don't think it's useful in the Bayesian model. The use is due to the Critical Rationalist approach.

Oscar_Cunningham 01 November 2017 02:45:56PM 2 points

However, if T is an explanatory theory (e.g. 'the sun is powered by nuclear fusion'), then its negation $\sim T$ ('the sun is not powered by nuclear fusion') is not an explanation at all.

The words "explanatory theory" seem to me to have a lot of fuzziness hiding behind them. But to the extent that "the sun is powered by nuclear fusion" is an explanatory theory I would say that the proposition $\sim T$ is just the union of many explanatory theories: "the sun is powered by oxidisation", "the sun is powered by gravitational collapse", and so on for all explanatory theories except "nuclear fusion".

Therefore, suppose (implausibly, for the sake of argument) that one could quantify 'the property that science strives to maximise'. If T had an amount q of that, then $\sim T$ would have none at all, not $1-q$ as the probability calculus would require if q were a probability.

There are lots of negative facts that are worth knowing and that scientists did good work to discover. When Michelson and Morley discovered that light did *not* travel through luminiferous aether that was a fact worth knowing, and lead to the discovery of special relativity. So even if you don't call $\sim T$ an explanatory theory it seems like it still has a lot of "the property that science strives to maximise"

Also, the conjunction (T_1 & T_2) of two mutually inconsistent explanatory theories T_1 and T_2 (such as quantum theory and relativity) is provably false, and therefore has zero probability. Yet it embodies some understanding of the world and is definitely better than nothing.

A Bayesian might instead define theories T_1' = "quantum theory leads to approximately correct results in the following circumstances ..." and T_2' "relativity leads to approximately correct results in the following circumstances ...". Then T_1' and T_2' would both have a high probability and be worth knowing, and so would their conjunction. The original conjunction, T_1 & T_2 , would mean "both quantum theory and relativity are exactly true". This of course is provably false, and so has probability 0.

Furthermore if we expect, with Popper, that all our best theories of fundamental physics are going to be superseded eventually, and we therefore believe their negations, it is still those false theories, not their true negations, that constitute all our deepest knowledge of physics.

Right, right. The statement T_1 is false; but the statement T_1' is true.

What science really seeks to 'maximise' (or rather, create) is explanatory power.

Does Deutsch write anywhere about what a precise definition of "explanation" would be?

curi 01 November 2017 07:29:26PM 0 points

Does Deutsch write anywhere about what a precise definition of "explanation" would be?

Yes, in Bol. <http://beginningofinfinity.com/books>

In short, explanations typically talk about why/how/because.

The words "explanatory theory" seem to me to have a lot of fuzziness hiding behind them. But to the extent that "the sun is powered by nuclear fusion" is an explanatory theory I would say that the proposition $\sim T$ is just the union of many explanatory theories: "the sun is powered by oxidisation", "the sun is powered by gravitational collapse", and so on for all explanatory theories except "nuclear fusion".

Unless you're claiming *non-explanatory theories don't exist at all*, then $\sim T$ includes both explanations and non-explanations. It doesn't consist of a union of many explanations.

A Bayesian might instead define theories T_1 = "quantum theory leads to approximately correct results in the following circumstances ..."

You're changed it to an instrumentalist theory which focuses on prediction instead of explanation. Deutsch refutes instrumentalism in his first book, FoR, also at the link above.

cousin_it 06 November 2017 10:30:52AM 0 points

You're changed it to an instrumentalist theory which focuses on prediction instead of explanation.

How so? I think it's still an explanatory theory, it just explains 99% of something instead of 100%.

curi 06 November 2017 06:14:07PM 0 points

Where's the explanation? What do you think an explanation is? You said the theory gets "approximately correct results" in some circumstances – doesn't that mean making approximately correct predictions?



MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

Do people think in a Bayesian or Popperian way?

-22 curi 10 April 2011 10:18AM

People think A&B is more likely than A alone, if you ask the right question. That's not very Bayesian; as far as you Bayesians can tell it's really quite stupid.

Is that maybe evidence that Bayesianism is failing to model how people actually thinking?

Popperian philosophy can make sense of this (without hating on everyone! it's not good to hate on people when there's better options available). It explains it like this: people like explanations. When you say "A happened because B happened" it sounds to them like a pretty good explanatory theory which makes sense. When you say "A alone" they don't see any explanation and they read it as "A happened for no apparent reason" which is a bad explanation, so they score it worse.

To concretize this, you could use A = economic collapse and B = nuclear war.

People are looking for good explanations. They are thinking in a Popperian fashion.

Isn't it weird how you guys talk about all these biases which basically consist of people not thinking in the way you think they should, but when someone says "hey, actually they think in this way Popper worked out" you think that's crazy cause the Bayesian model must be correct? Why did you find all these counter examples to your own theory and then never notice they mean your theory is wrong? In the cases where people don't think in a Popperian way, Popper explains why (mostly b/c of the justificationist tradition informing many mistakes since Aristotle)

More examples, from <http://wiki.lesswrong.com/wiki/Bias>

Scope Insensitivity - The human brain can't represent large quantities: an environmental measure that will save 200,000 birds doesn't conjure anywhere near a hundred times the emotional impact and willingness-to-pay of a measure that would save 2,000 birds.

Changing the number does not change most of the explanations involved, such as why helping birds is good, what the person can afford to spare, how much charity it takes the person to feel altruistic enough (or moral enough, involved enough, helpful enough, whatever), etc... Since the major explanatory factors they were considering don't change in proportion to the number of birds, their answer doesn't change proportionally either.

Correspondence Bias, also known as the fundamental attribution error, refers to the tendency to attribute the behavior of others to intrinsic dispositions, while excusing one's own behavior as the result of circumstance.

This happens because people usually know the explanations/excuses for why they did stuff, but they don't know them for others. And they have more reason to think of them for themselves.

Confirmation bias, or Positive Bias is the tendency to look for evidence that confirms a hypothesis, rather than disconfirming evidence.

People do this because of the justificationist tradition, dating back to Aristotle, which Bayesian epistemology is part of, and which Popper rejected. This is a way people really don't think in the Popperian way -- but they could and should.

Planning Fallacy - We tend to plan envisioning that everything will go as expected. Even assuming that such an estimate is accurate conditional on everything going as expected, things will not go as expected. As a result, we routinely see outcomes worse than the ex ante worst case scenario.

This is also caused by the justificationist tradition, which Bayesian epistemology is part of. It's not fallibilist enough. This is a way people really don't think in the Popperian way -- but they could and should.

Well, that's part of the issue. The other part is they come up with a good explanation of what will happen, and they go with that. That part of their thinking fits what Popper said people do. The problem is not enough criticism, which is from the popularity of justificationism.

Do We Believe Everything We're Told? - Some experiments on priming suggest that mere exposure to a view is enough to get one to passively accept it, at least until it is specifically rejected.

That's very Popperian. The Popperian way is that you can make conjectures however you want, and you only reject them if there's a criticism. No criticism, no rejection. This contrasts with the justificationist approach in which ideas are required to (impossibly) have positive support, and the focus is on positive support not criticism (thus causing, e.g., Confirmation Bias)

Illusion of Transparency - Everyone knows what their own words mean, but experiments have confirmed that we

Custom Search

curi [Create new article](#)
 Karma Score [Submit a new link](#)
 12 [Add new meetup](#)
 2 [Preferences](#)
 [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
[Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by Habryka on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by korin43 on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by Multipartite on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by trickster on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by adjuant on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

systematically overestimate how much sense we are making to others.

This one is off topic but there's several things I wanted to say. First, people don't always know what their own words mean. People talking about tricky concepts like God, qualia, or consciousness often can't explain what they mean by the words if asked. Sometimes people even use words without knowing the definition, they just heard it in a similar circumstance another time or something.

The reason others don't understand us, often, is because of the nature of communication. To communicate what has to happen is the other person creates knowledge of what idea(s) you are trying to express to him. That means he has to make guesses about what you are saying and use criticisms to improve those guesses (e.g. by ruling stuff out incompatible with the words he heard you use). In this way Popperian epistemology lets us understand communication, and why it's so hard.

Evaluability - It's difficult for humans to evaluate an option except in comparison to other options. Poor decisions result when a poor category for comparison is used. Includes an application for cheap gift-shopping.

It's because they are trying to come up with a good explanation of what to buy. And "this one is better than this other one" is a pretty simple and easily available kind of explanation to create.

The Allais Paradox (and subsequent followups) - Offered choices between gambles, people make decision-theoretically inconsistent decisions.

How do you know that kind of thing and still think people reason in a Bayesian way? They don't. They just guess at what to gamble, and the quality of the guesses is limited by what criticisms they use. If they don't know much math then they don't subject their guesses to much mathematical criticism. Hence this mistake.

[Article Navigation](#)

[Comments \(37\)](#)

Tags: [popper](#) | [deutsch](#) | [Delete](#)

Comments (37)

Sort By: **New**

Enter a comment here

Comment

Show help

JoshuaZ 10 April 2011 03:41:41PM 11 points

Others have already answered this, but there's another problem: you clearly haven't read the actual literature on the conjunction fallacy. It doesn't just occur in the form "A because of B." It connects with the representative heuristic. Thus, for suitably chosen A and B, people act like "A and B" is more likely than "A". See Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Tversky, Amos; Kahneman, Daniel Psychological Review, Vol 90(4), Oct 1983, 293-315. [doi: 10.1037/0033-295X.90.4.293](https://doi.org/10.1037/0033-295X.90.4.293)

Please stop posting and read the literature on these issues.

zaph 10 April 2011 03:18:29PM * 2 points

I think you should read up on [the conjunction fallacy](#). Your example does not address the observations made in research by Kahneman and Tversky. The questions posed in the research do not assume causal relationships, they are just two independent probabilities. I won't rewrite the whole wiki article, but the upshot of the conjunction fallacy is that people using representativeness heuristic to asses odds, instead of using the correct procedures they would have used if that heuristic isn't cued. People who would never say "Joe rolled a six and a two" is more likely than "Joe rolled a two" do say "Joe is a New Yorker who rides the subway" is more likely than "Joe is a New Yorker", when presented with information about Joe.

jimrandomh 10 April 2011 01:21:59PM 15 points

Please, stop. This has gone on long enough. You don't have to respond to everything, and you *shouldn't* respond to everything. By trying to do so, you have generated far more text than any reasonable person would be willing to read, and it's basically just repeating the same incorrect position over and over again. It is quite clear that we are not having a [rational discussion](#), so there is nothing further to say.

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- [gwern](#) (32)
- [ingres](#) (30)
- [Lumifer](#) (27)
- [turchin](#) (23)
- [Habryka](#) (22)
- [Yosarian2](#) (21)
- [IlyaShpitser](#) (19)
- [SaidAchmiz](#) (19)
- [gjm](#) (19)
- [fowlertm](#) (17)
- [SquirrelInHell](#) (15)
- [Dagon](#) (15)
- [Elo](#) (14)
- [Stuart_Armstrong](#) (13)
- [Gunnar_Zarncke](#) (12)

RECENT KARMA AWARDS

Nic_Smith 10 April 2011 07:17:59PM 9 points

Indeed. This Popperclipping of the discussion section should cease.

[deleted] 10 April 2011 10:20:51PM 4 points

This situation seems an ideal test of the karma system.

prase 10 April 2011 11:47:00PM -1 points

And it works.

[deleted] 11 April 2011 12:06:44AM 6 points

What beneficial effect have you observed? I ask because people were complaining about the forum being popperclipped. Do you disagree with these complaints? Or do you think that the karma system has trained the low-karma popperclipping participants to improve the quality of their comments? One of them recently wrote a post admitting and defending the tactic of being obnoxious - he said that his obnoxiousness was to filter out time-wasters.

prase 11 April 2011 12:28:58AM 0 points

I mean curi has now insufficient karma to post on the main page and his comments are generally heavily downvoted. People can disable viewing low karma comments, so popperclipping (whatever it means - did the old term "troll" grow out of fashion?) may not be a problem. Therefore I think that karma works.

Desrtopa 11 April 2011 02:13:12AM 10 points

Curi's karma periodically spikes despite posting no significantly upvoted comments or any improvement in his reception. I suspect he or someone else who frequents his site may be generating puppet accounts to feed his comments karma (his older comments appear to have gone through periodic blanket spikes.) He's posted main page and discussion articles multiple times after his karma has dropped to zero without first producing more comments that are upvoted, due to these spikes.

prase 11 April 2011 08:17:28AM 6 points

If this is true, it would be natural for the moderators to step in and ban him.

Alicorn 12 April 2011 08:55:15PM 1 point

I asked matt if this could be confirmed, but apparently there's only a very time-consuming method to gather anything other than circumstantial evidence for the accusation.

JoshuaZ 12 April 2011 09:08:31PM 1 point

I asked matt if this could be confirmed, but apparently there's only a very time-consuming method to gather anything other than circumstantial evidence for the accusation.

Jimrandomh had an idea for setting up a script that might help, maybe talk to him? In any event, it might be useful to have the capability to do this in general. That said, since this is only the first time we've had such a problem, it doesn't seem as of right now that this is a common enough issue to really justify investing in additional capabilities for the software.

[deleted] 11 April 2011 12:51:12AM 6 points 

popperclipping (whatever it means...)

I believe that "popperclipping" is a play on words, a joke, alluding to a popular LW topic. Explaining it more might kill the joke.

I mean curi has now insufficient karma to post on the main page

Currently, on the main page, the most recent post under "Recent Posts" is curi's The Conjunction Fallacy Does Not Exist. The comments under this are showing up in the Recent Comments column. Of the five comments I see in the recent comments column, three are comments under curi's posts. That is a majority. As of now, then, it appears that curi continues to dominate discussion, either directly or by triggering responses.

prase 11 April 2011 01:06:08AM 1 point 

I believe that "popperclipping" is a play on words, a joke, ...

Certainly. I only missed the standard name for that behaviour spelled out loud.

prase 11 April 2011 01:01:26AM * 5 points 

Damn, I thought it was in the discussion. Then, I retract my statement that karma works. Still, what's the explanation? Where did curi get enough karma to balance the blow from his heavily downvoted comments and posts? I have looked onto two pages of his recent activity where his score was -112 (-70 for the main page post, -42 for the rest). And I know he was near zero after his last but one main page post was published.

TheOtherDave 10 April 2011 02:00:25PM 6 points 

While agreeing with you completely, I'll also point out that *quite a few* people have been feeding this particular set of threads... that is, continuing to have, at enormous length, a discussion in which no progress is being made.

Emile 10 April 2011 01:34:10PM 5 points 

Seconded. When I discovered this ongoing conversation on Popperian epistemology, there were already three threads, some of them with hundreds of comments, and no signs of progress and mutual agreement, only argument. There may be some comments worth reading in the stack, but they're not worth the effort of digging.

benelliott 10 April 2011 11:01:29AM 5 points 

With the Allais Paradox, would you say that the decisions people make are consistent with Popperian philosophy? Or at any rate would you say that, as a Popperian, you would make similar decisions?

benelliott 10 April 2011 10:25:28AM 18 points 

Nobody here is claiming that people naturally reason in a Bayesian way.

We are claiming that they should.

shokwave 10 April 2011 03:29:03PM 2 points 

This, this, a million times this.

curi 10 April 2011 10:27:30AM * -1 points

If people don't reason in a Bayesian way, but they do reason, it implies there is a non-Bayesian way to reason which works (at least a fair amount, e.g. we managed to build computers and space ships). Right?

Claims that people think in an inductive way are common here. Note how my descriptions are different than that and account for the evidence.

Someone told me that humans do and must think in a bayesian way at some level b/c it's the only way that works.

Eugine_Nier 10 April 2011 04:48:54PM * 9 points

As Eliezer said in [Searching for Bayes-Structure](#):

The way you begin to grasp the Quest for the Holy Bayes is that you learn about cognitive phenomenon XYZ, which seems really useful - and there's this bunch of philosophers who've been arguing about its true nature for centuries, and they are still arguing - and there's a bunch of AI scientists trying to make a computer do it, but they can't agree on the philosophy either -

And - *Huh, that's odd!* - this cognitive phenomenon didn't look anything like Bayesian on the surface, but there's this non-obvious underlying structure that has a Bayesian interpretation - but wait, there's still some useful work getting done that can't be explained in Bayesian terms - no wait, *that's* Bayesian too - OH MY GOD this *completely different* cognitive process, that *also* didn't look Bayesian on the surface, ALSO HAS BAYESIAN STRUCTURE - hold on, are these non-Bayesian parts even *doing* anything?

- Yes: Wow, those are Bayesian too!
- No: Dear heavens, what a [stupid design](#). I could eat a bucket of amino acids and puke a better brain architecture than that.

jmmcd 10 April 2011 02:25:19PM -1 points

The core of the problem:

Someone told me that humans do and must think in a bayesian way at some level b/c it's the only way that works.

No link to that someone? If you can remember who it was, you should go and argue with them. To everyone else, this is a straw man.

curi 10 April 2011 07:14:17PM 4 points

Here's someone saying it again by quoting Yudkowsky saying it:

http://lesswrong.com/lw/56e/do_people_think_in_a_bayesian_or_popperian_way/3w7o

Some straw man... I thought people would be familiar with this kind of thing without me having to quote it.

Cyan 10 April 2011 02:49:26PM * 2 points

(Certainly there are researchers [looking for Bayes structure](#) in low-level neural processing, but those investigations focus on tasks far below human cognition.)

RichardKennaway 10 April 2011 10:48:40AM 5 points

If people don't reason in a Bayesian way, but they do reason, it implies there is a non-Bayesian way to reason which works (at least a fair amount, e.g. we managed to build computers and space ships).

There is. That does not mean that it is without error, or that errors are not errors. A&B is, everywhere and always, no more likely than A. Any method of concluding otherwise is wrong. If the form of reasoning that Popper advocates endorses this error, it is wrong.

Someone told me that humans do and must think in a bayesian way at some level b/c it's the only way that works.

Whoever that was is wrong.

curi 10 April 2011 08:09:15PM 2 points

Here's someone saying it again by quoting Yudkowsky saying it:

http://lesswrong.com/lw/56e/do_people_think_in_a_bayesian_or_popperian_way/3w7o

No doubt Yudkowsky is wrong, as you say.

RichardKennaway 11 April 2011 06:27:46AM * 0 points

See [my other response](#) to Oscar_Cunningham, who cited the same article.

Oscar_Cunningham 10 April 2011 03:14:28PM 6 points

Someone told me that humans do and must think in a bayesian way at some level b/c >>it's the only way that works.

Whoever that was is wrong.

[Eliezer?](#)

RichardKennaway 10 April 2011 05:53:36PM * 6 points

Eliezer can say whether curi's view is a correct reading of that article, but it seems to me that if Bayesian reasoning is the core that works, but humans do a lot of other stuff as well that is all either useless or harmful, and don't even know the gold from the dross, then this is not in contradiction with demonstrating that the other stuff is due to Popperian reasoning. It rather counts against Popper though. Or at least, Popperianism.

Oscar_Cunningham 10 April 2011 07:33:38PM 1 point

Agreed.

benelliott 10 April 2011 10:41:57AM 7 points

Someone told me that humans do and must think in a bayesian way at some level b/c it's the only way that works.

Humans think in an approximately Bayesian way. The biases are the places where the approximation breaks down, and human thinking starts to fail.

Claims that people think in an inductive way are common here. Note how my descriptions are different than that and account for the evidence.

You have not given one example of non-inductive thinking. I really do not see how you could get through the day without induction.

I am riding my bike to college after it rained during the night, and I notice that the rain has caused a path I use to become a muddy swamp, meaning I have to take a detour and arrive late. Next time it rains, I leave home early because I expect to encounter mud again.

If you wish to claim that most people are non-inductive you must either:

1) Show that I am unusual for thinking in this way

or

2) Show how someone else could come to the same conclusion without induction.

If you choose 1) then you must also show why this freakishness puts me at a disadvantage, or concede that other people should be inductive.

curi 10 April 2011 07:23:22PM -2 points

You have not given one example of non-inductive thinking. I really do not see how you could get through the day without induction.

I get hungry. So I guess some things I might like to eat. I criticize my guesses. I eat.

benelliott posts on less wrong. I guess what idea he's trying to communicate. With criticism and further guessing I figure it out. I reply.

Most of this is done subconsciously.

Now how about an example of induction?

In order to evaluate if it is an example of induction, you'll need to start with a statement of the method of induction. This is not b/c I'm unfamiliar with such a thing but because we will disagree about it and we better have one to get us on the same page more (inductivists vary a lot. I know many different statements of how it works.)

In the example you give, you don't give any explanation of what you think it has to do with induction. Do you think it's inductive because you learned a new idea? Do you think it's inductive because it's impossible to conjecture that you should do that next time it rains? Do you think it's inductive because you learned something from a single instance? (Normally people giving examples of induction will have multiple data points they learn from, not one. Your example is not typical at all.)

benelliott 10 April 2011 09:39:39PM 2 points

In order to evaluate if it is an example of induction, you'll need to start with a statement of the method of induction. This is not b/c I'm unfamiliar with such a thing but because we will disagree about it and we better have one to get us on the same page more (inductivists vary a lot. I know many different statements of how it works.)

I'm tempted just to point to my example and say 'there, that's what I call induction', but I doubt that will satisfy you so I will try to give a more rigorous explanation.

I view induction as Bayesian updating/decision theory with an inductive prior. To clarify what I mean, suppose I am faced with a opaque jar, containing ten beads, each of which is either red or white. What is my prior for the contents of the jar? It depends on my background knowledge.

1) I may know that someone carefully put 5 red beads and 5 white beads in the jar

2) I may know that each ball was chosen randomly with probability p , where p is a parameter which is (as far as I know) equally likely to be anywhere between 0 and 1

3) I may know that each ball was tossed in by a monkey which was drawing randomly from two barrels, one containing red balls, one containing white balls.

I may also have many other states of knowledge, but I give just three examples for simplicity.

1) is anti-inductive. If I have drawn N balls, R of which have been red, then $P(\text{the next ball is red}) = (5-R)/(10-N)$, so every red I draw decreases my anticipation of red, while every white increases it.

2) is inductive. If I have drawn N balls, R of which have been red, then $P(\text{the next ball is red}) = (R+1)/(N+2)$ (this is a theorem due to Laplace, the proof is not quite trivial). Every red ball increases my anticipation of red, while every white increases it. Notice how it takes many reds to provide strong

evidence, but even one red is sufficient for a fairly large update, from 0.5 to 0.67.

3) is neither inductive nor anti-inductive. $P(\text{the next ball is red}) = 0.5$ regardless of what I have drawn. Past observations do not influence expectation of future observations.

With the mud, neither of the three examples perfectly describes my prior, but 2) comes closest. Most proposals for universal priors are to some extent inductive, for examples Solomonoff assigns a much higher probability to '1000 0s' than '999 0s followed by a 1'.

Brief note: Human induction, and Solomonoff Induction are more sophisticated than 2) mainly because they have better pattern spotting abilities, and so the process is not quite analogous.

Alexandros 10 April 2011 10:24:51AM * 7 points

Are you implying human thinking should be used as some sort of benchmark? Why in the space of all possible thought processes would the human family of thought processes, hacked together by evolution to work just barely well enough, represent the ideal? Also, are you applying the 'popperian' label to human thinking? If I prove human thinking to be wrong by its own standards, have I falsified the popperian process of approaching truth?

I am not well versed (or much invested) in bayes but this is not making much sense.

radical_negative_one 10 April 2011 10:32:05AM 4 points

To clarify/rephrase/expand on this, i think Alexandros is suggesting that questions "how do humans think", and "what is a rational way to think" are separate questions, and if we are discussing the first of these two questions then perhaps we have been sidetracked.

In fact, this is nicely highlighted by your very first sentence:

People think A&B is more likely than A alone, if you ask the right question. That's not very Bayesian; as far as you Bayesians can tell it's really quite stupid.

That *is* a quite stupid way to think, and if we want to think rationally we should desire to not think that way, regardless of whether it is in fact a common way of thinking.

curi 10 April 2011 10:27:54AM -12 points

No. What?



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

some posts on instrumentalism and empiricism in less wrong **Draft**

2 alanforr 13 November 2017 01:06PM

Elliot Temple has written a couple of blog posts about empiricism and instrumentalism that criticise some ideas that are common on Less Wrong:

<http://curi.us/2067-empiricism-and-instrumentalism>

<http://curi.us/2066-replies-to-gyrodidiot-about-fallible-ideas-critical-rationalism-and-paths-forward>

Article Navigation

[Comments \(4\)](#)

[Delete](#) | [Ban](#)

Comments (4)

Sort By: **Best**

Enter a comment here

Comment

Show help

curi 12 November 2017 09:02:13PM * 0 points



You can find more discussion of Paths Forward in the comments here. It's higher quality than a lot of the discussion that happened at LW:

<http://curi.us/2065-open-letter-to-machine-intelligence-research-institute>

Custom Search

alanforr

Create new article

Karma Score

Submit a new link

17

Add new meetup

25

Preferences



Log out

Subscribe to RSS Feed

NEAREST MEETUPS

Brussels - We meet every month: 12 January 2019 02:00PM

LW Netherlands: 03 September 2122 04:52PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
Less Wrong Study Hall

RECENT COMMENTS

That's really broadening the term
by **pmw7070** on "Flinching away from truth" is often about *protecting* the

Also I made a video with some thoughts on my visit to Less Wrong (which I fear is wrapping up). It's informal and I don't expect it to convince anyone who was unconvinced by the written blog posts. But it has animoji! It's my first recording using animoji! They're neat! 🐱🐶🐼🐻🐾

https://www.youtube.com/watch?v=hyh_i7rzt8

[Ban](#)

Yosarian2 13 November 2017 10:28:51AM 0 points [\[-\]](#)

I was watching part of your video, and I'm really surprised that you think that LessWrong doesn't have what you call "paths forward", that is, ways for people who disagree to find a path towards considering where they may be wrong and trying to hear the other person's point of view. In fact, that's actually a huge focus around here, and a lot has been written about ways to do that.

[Ban](#)

Gyrodidiot 12 November 2017 08:19:54PM * 0 points [\[-\]](#)

As I'm the one being answered to, a bit of context. A long discussion started on the #philosophy channel of the Slack group. For reasons irrelevant to the present discussion, I'm continuing the exchange on the website linked above.

I'm currently writing an answer to this. I do not claim to represent the LW community, though I'm trying my best to reflect the broad concepts and reasoning outlined in the Sequences, notably. Please correct any blatant inaccuracies in my prose, if you think it worthwhile.

[Ban](#)

curi 12 November 2017 08:50:43PM 0 points [\[-\]](#)

FYI to everyone: The comments on empiricism and instrumentalism (the first link) are basically impersonal and apply to views of many LW participants. The material in the second link is more mixed, but probably some of the comments apply to you, dear reader.

[Ban](#)

epistemology | 0 points

Great post, obviously. You argue
by Caspar42 on Intellectual Hipsters
and Meta-Contrarianism | 0 points

Exactly. "The reality is undecidation"
by entirelyuseless on Reductionism |
0 points

That observation runs headlong
by TheAncientGeek on Reductionism
| 2 points

If it requires a round-trip of human
by AndyC on Positive Bias: Look Into
the Dark | 0 points

RECENT POSTS

LW 2.0 Open Beta Live
by Vaniver | 22v (34c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

**European Community Weekend
2017**
by DreamFlasher | 16v (16c)

**"Flinching away from truth" is often
about *protecting* the epistemology**
by AnnaSalamon | 73v (54c)

**Further discussion of CFAR's focus
on AI safety, and the good things
folks wanted from "cause
neutrality"**
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

Popperian Decision making

-1 curi 07 April 2011 06:42AM

Branching from: http://lesswrong.com/lw/54u/bayesian_epistemology_vs_popper/3uta?context=4

The question is: how do you make decisions without justifying decisions, and without foundations?

If you can do that, I claim the regress problem is solved. Whereas induction, for example, is refuted by the regress problem (no, arbitrary foundations or circular arguments are not solutions).

OK stepping back a bit, and explaining less briefly:

Infinite regresses are nasty problems for epistemologies.

All justificationist epistemologies have an infinite regress.

That means they are false. They don't work. End of story.

There's options of course. Don't want a regress? No problem. Have an arbitrary foundation. Have an unjustified proposition. Have a circular argument. Or have something else even sillier.

The regress goes like this, and the details of the justification don't matter.

If you want to justify a theory, T0, you have to justify it with another theory, T1. Then T1 needs justify by T2. Which needs justifying by T3. Forever. And if T25 turns out wrong, then T24 loses it's justification. And with T24 unjustified, T23 loses its justification. And it cascades all the way back to the start.

I'll give one more example. Consider probabilistic justification. You assign T0 a probability, say 99.999%. Never mind how or why, the probability people aren't big on explanations like that. Just do your best. It doesn't matter. Moving on, what we have to wonder if that 99.999% figure is correct. If it's not correct then it could be anything such at 90% or 1% or whatever. So it better be correct. So we better justify that it's a good theory. How? Simple. We'll use our whim to assign it a probability of 99.99999%. OK! Now we're getting somewhere. I put a lot of 9s so we're almost certain to be correct! Except, what if I had that figure wrong? If it's wrong it could be anything such as 2% or 0.0001%. Uh oh. I better justify my second probability estimate. How? Well we're trying to defend this probabilistic justification method. Let's not give up yet and do something totally differently, instead we'll give it another probability. How about 80%? OK! Next I ask: is that 80% figure correct? If it's not correct, the probability could be anything, such as 5%. So we better justify it. So it goes on and on forever. Now there's two problems. First it goes forever, and you can't ever stop, you've got an infinite regress. Second, suppose you stopped have some very large but finite number of steps. Then the probability the first theory is correct is arbitrarily small. Because remember that at each step we didn't even have a guarantee, only a high probability. And if you roll the dice a lot of times, even with very good odds, eventually you lose. And you only have to lose once for the whole thing to fail.

OK so regresses are a nasty problem. They totally ruin all justificationist epistemologies. That's basically every epistemology anyone cares about except skepticism and Popperian epistemology. And forget about skepticism, that's more of an anti-epistemology than an epistemology: skepticism consists of giving up on knowledge.

Now we'll take a look at Popper and Deutsch's solution. In my words, with minor improvements.

Regresses all go away if we drop justification. Don't justify anything, ever. Simple.

But justification had a purpose.

The purpose of justification is to sort out good ideas from bad ideas. How do we know which ideas are any good? Which should we believe are true? Which should we act on?

Custom Search

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by Habryka on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by korin43 on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by Multipartite on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by trickster on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by adjuant on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Popperian Decision making - Less Wrong

-1 curi 07 April 2011 06:42AM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(100\)](#)

Tags: [popper](#) | [deutsch](#)

Comments (100)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 09:43:55PM 1 point

Many starting points work fine.

In theory, could you get stuck? I don't have a proof either way.

I don't mind too much. Humans already have standards of criticism which don't get stuck. We have made scientific progress. Our standards we already have allow self-modification and thereby unbounded progress. So it doesn't matter what would have happened if we had started with a bad standard once a upon a time, we're past that (it does matter if we want to create an AI).

[deleted] 07 April 2011 09:49:28PM 1 point

You would definitely get stuck. The problem Khoth pointed out is that your method can't distinguish between good criticism and bad criticism. Thus, you could criticize any standard that you come up with, but you'd have know way of knowing which criticisms are legitimate, so you wouldn't know which standards are better than others.

I agree that in practice we don't get stuck, but that's because we don't use the method or the assumptions you are defending.

curi 07 April 2011 09:51:31PM 1 point

Thus, you could criticize any standard

I meant stuck in the sense of *couldn't* get out of. Not in the sense of *could* optionally remain stuck.

I agree that in practice we don't get stuck, but that's because we don't use the method or the assumptions you are defending.

What's the argument for that?

We have knowledge about standards of criticism. We use it. Objections about starting points aren't very relevant because Popperians never said they were justified by their starting points. What's wrong with this?

[deleted] 07 April 2011 09:59:59PM 1 point

I meant stuck in the sense of couldn't get out of. Not in the sense of could optionally remain stuck.

I don't think there's a way out if your method doesn't eventually bottom out somewhere. If you don't have a

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

reliable or objective way of distinguishing good criticism from bad, the act of criticism can't help you in any way, *including* trying to fix this standard.

We have knowledge about standards of criticism. We use it. Objections about starting points aren't very relevant because Popperians never said they were justified by their starting points. What's wrong with this?

If you don't have objective knowledge of standards of criticism and you are unwilling to take one as an axiom, then what are you justified by?

curi 08 April 2011 12:34:26AM 0 points

If you don't have objective knowledge of standards of criticism and you are unwilling to take one as an axiom, then what are you justified by?

Nothing. Justification is a mistake. The request that theories be justified is a mistake. They can't be. They don't need to be.

If you don't have a reliable or objective way of distinguishing good criticism from bad, the act of criticism can't help you in any way, including trying to fix this standard.

Using the best ideas we know of so far is a partially reliable, partially objective way which allows for progress.

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all' by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (32)

ingres (30)

Lumifer (27)

turchin (23)

Habryka (22)

Yosarian2 (21)

IlyaShpitser (19)

SaidAchmiz (19)

gjm (19)

fowlertm (17)

SquirrelInHell (15)

Dagon (15)

Elo (14)

Stuart_Armstrong (13)

Gunnar_Zarncke (12)

RECENT KARMA AWARDS

BTW that's the same general problem that induction was trying to address. And induction is false. So that's another reason we need a solution to this issue.

The method of addressing this issue has several steps, so try to follow along.

Step 1) You can suggest any ideas you want. There's no rules, just anything you have the slightest suspicion might be useful. The source of the ideas, and the method of coming up with them, doesn't matter to anything. This part is easy.

Step 2) You can criticize any idea you want. There's no rules again. If you don't understand it, that's a criticism -- it should have been easier to understand. If you find it confusing, that's a criticism -- it should have been clearer. If you think you see something wrong with it, that's a criticism -- it shouldn't have been wrong in that way, *or* it should have included an explanation so you wouldn't make a mistaken criticism. This step is easy too.

Step 3) All criticized ideas are rejected. They're flawed. They're not good enough. Let's do better. This is easy too. Only the *exact* ideas criticized are rejected. Any idea with at least one difference is deemed a new idea. It's OK to suggest new ideas which are similar to old ideas (in fact it's a good idea: when you find something wrong with an idea you should try to work out a way to change it so it won't have that flaw anymore).

Step 4) If we have exactly one idea remaining to address some problem or question, and no one wants to revisit the previous steps at this time, then we're done for now (you can always change your mind and go back to the previous steps later if you want to). Use that idea. Why? Because it's the only one. It has no rivals, no known alternatives. It stands alone as the only non-refuted idea. We have sorted out the good ideas from the bad -- as best we know how -- and come to a definite answer, so use that answer. This step is easy too!

Step 5) What if we have a different number of ideas left over which is not exactly one? We'll divide that into two cases:

Case 1) What if we have two or more ideas? This one is easy. There is a particular criticism you can use to refute all the remaining theories. It's the same every time so there's not much to remember. It goes like this: idea A ought to tell me why B and C and D are wrong. If it doesn't, it could be better! So that's a flaw. Bye bye A. On to idea B: if B is so great, why hasn't it explained to me what's wrong with A, C and D? Sorry B, you didn't answer all my questions, you're not good enough. Then we come to idea C and we complain that it should have been more help and it wasn't. And D is gone too since it didn't settle the matter either. And that's it. Each idea should have settled the matter by giving us criticisms of all its rivals. They didn't. So they lose. So whenever there is a stalemate or a tie with two or more ideas then they all fail.

Case 2) What if we have zero ideas? This is crucial because case one always turns into this! The answer comes in two main parts. The first part is: think of more ideas. I know, I know, that sounds hard. What if you get stuck? But the second part makes it easier. And you can use the second part over and over and it keeps making it easier every time. So you just use the second part until it's easy enough, then you think of more ideas when you can. And that's all there is to it.

OK so the second part is this: be less ambitious. You might worry: but what about advanced science with its cutting edge breakthroughs? Well, this part is optional. If you can wait for an answer, don't do it. If there's no hurry, then work on the other steps more. Make more guesses and think of more criticisms and thus learn more and improve your knowledge. It might not be easy, but hey, the problem we were looking at is how to sort out good ideas from bad ideas. If you want to solve hard problems then it's not easy. Sorry. But you've got a method, just keep at it.

But if you have a decision to make then you need an answer now so you can make your decision. So in that case, if you actually want to reach a state of having exactly one theory which you can use now, then the trick is when you get stuck be less ambitious. I think how you can see how that would work in general terms. Basically if human knowledge isn't good enough to give you an answer of a certain quality right now, then your choices are either to work on it more and not have an answer now, or accept a lower quality answer. You can see why there isn't really any way around that. There's no magic way to always get a top quality answer now. If you want a cure for cancer, well I can't tell you how to come up with one in the next five minutes, sorry.

This is a bit vague so far. How does lowering your standards address the problem. So what you do is propose a new idea like this, "I need to do something, so I will do..." and then you put whatever you want (idea A, idea B, some combination, whatever).

This new idea is not refuted by any of the existing criticisms. So now you have one idea, it isn't refuted, and you might be done. If you're happy with it, great. But you might not be. Maybe you see something wrong with it, or you have another proposal. That's fine; just go back to the first three steps and do them more. Then you'll get to step 4 or 5 again.

What if we get back here? What do we do the second time? The third time? We simply get less ambitious each time. The harder a time we're having, the less we should expect. And so we can start criticizing any ideas that aim too high.

BTW it's explained on my website here, including an example:

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

[gwern](#) (32)

[ingres](#) (30)

[Lumifer](#) (27)

[turchin](#) (23)

[Habryka](#) (22)

[Yosarian2](#) (21)

[IlyaShpitser](#) (19)

[SaidAchmiz](#) (19)

[gjm](#) (19)

[fowlertm](#) (17)

[SquirrelInHell](#) (15)

[Dagon](#) (15)

[Elo](#) (14)

[Stuart_Armstrong](#) (13)

[Gunnar_Zarncke](#) (12)

RECENT KARMA AWARDS

<http://fallibleideas.com/avoiding-coercion>

Read that essay, keeping in mind what what I've been saying, and hopefully everything will click. Just bear in mind that when it talks about cooperation between people, and disagreements between people, and coming up with solutions for people -- when it discusses ideas in two or more separate minds -- everything applies exactly the same if the two or more conflicting ideas are all in the same mind.

What if you get real stuck? Well why not do the first thing that pops into your head? You don't want to? Why not? Got a criticism of it? It's better than nothing, right? No? If it's not better than nothing, do nothing! You think it's silly or dumb? Well so what? If it's the best idea you have then it doesn't matter if it's dumb. You can't magically instantly become super smart. You have to use your best idea even if you'd like to have better ideas.

Now you may be wondering whether this approach is truth-seeking. It is, but it doesn't always find the truth immediately. If you want a resolution to a question immediately then its quality cannot exceed today's knowledge (plus whatever you can learn in the time allotted). It can't do better than the best that is known how to do. But as far as long term progress, the truth seeking came in those first three steps. You come up with ideas. You criticize those ideas. Thereby you eliminate flaws. Every time you find a mistake and point it out you are making progress towards the truth. That's how we approach the truth: not by justifying but by identify mistakes and learning better. This is evolution, it's the solution to Paley's problem, it's discussed in BoI and on my Fallible Ideas website. And it's not too hard to understand: improve stuff, keep at it, and you get closer to the truth. Mistake correcting -- criticism -- is a truth-seeking method. That's where the truth-seeking comes from.

Article Navigation

[Comments \(100\)](#)

Tags: [popper](#) | [deutsch](#)

Comments (100)

Sort By: [New](#)

Enter a comment here

Comment

Show help

zntneo 19 May 2011 06:15:14AM 0 points [\[x\]](#)

It seems you have completely talked about interalist versions of epistemology. What about Relaiablism? It does not fall into either of your categories (its one i'm pretty sympathetic towards).

Aso to make sure i understand you correctly this is arguing about getting rid of justified in the standard true justified belief (also including getier part to) am i right? or are you saying something is "justified" when it can no longer be criticized (due to not being about to come up with a criticism)? I also agree with Yvain that it seems this "criticize the idea" needs to be taken apart more.

I might have more comments but need to think about it more

Architectonic 18 April 2011 04:01:05PM 0 points [\[x\]](#)

Many of the criticisms mentioned in the above comments have in fact been addressed by Bartley in his conception of pan-critical rationalism. See his book "The Retreat to Commitment".

Bayesian methods can be considered useful within such an epistemological system, however one cannot justify that one fact is more true than another merely based on Bayesian probabilities.

Both justificationist and falsificationist outlooks are stated with respect to something else. That is why philosophers played all those language games. They soon realised that you couldn't reduce everything down to language, be it natural or symbolic without losing something. Axiomatic systems don't make any sense on their own. It is sad that many commit to a justificationist position without realising that they are doing so.

Yvain 10 April 2011 01:49:53PM * 10 points [\[x\]](#)

I am very inexperienced in epistemology, so forgive me if I'm making a simple error.

But it sounds like everything *important* in your theory is stuck into a black box in the words "criticize the idea".

Suppose we had a computer program designed to print the words "I like this idea" to any idea represented as a string with exactly 5 instances of the letter 'e' in it, and the words "I dislike this idea because it has the wrong number of 'e's in it" to any other idea.

And suppose we had a second computer program designed to print "I like this idea" to any idea printed on blue paper, and "I dislike this idea because it is on the wrong color paper" to any idea printed on any other color of paper.

These two computers could run through your decision making process of generating and criticizing ideas, and eventually would settle on the first idea generated which was written on blue paper and which used the letter 'e' exactly five times.

So it would seem that for this process to capture what we mean by "truth", you have to start out with some reasoners who *already* have a pretty good set of internal reasoning processes kind of like our own that they use when criticizing an idea.

But everything that's interesting and difficult about epistemology is captured in that idea of "a pretty good set of internal reasoning processes kind of like our own that they use when criticizing an idea", so really this decision-making process only works for entities that are already running a different epistemology that's doing all the work.

It almost seems like a [detached lever fallacy](#), where the lever is the ability to criticize ideas, and the machinery the lever is activating is the actual epistemology the agent is using.

drethelin 07 April 2011 06:39:51PM 2 points 

I fail to see how, in practical terms, this is at all better than using induction based reasoning. It may make you feel and look smarter to tell someone they can't prove or disprove anything with certainty, but that's not exactly a stunning endorsement. You can't actually ACT as if nothing is ever conclusively true. I would like to see a short description as to WHY this is a better way to view the world and update your beliefs about it.

curi 07 April 2011 07:49:26PM 2 points 

You can't actually ACT as if nothing is ever conclusively true.

But I do act that way. I am a fallibilist. Are you denying fallibilism? Some people here endorsed it. Is there a Bayesian consensus on it?

Why do you think I don't act like that?

I would like to see a short description as to WHY this is a better way to view the world and update your beliefs about it.

Because it works and makes sense. If you want applications to real life fields you can find applications to parenting, relationships and capitalism here:

<http://fallibleideas.com/>

GuySrinivasan 07 April 2011 08:23:00PM 0 points 

I was interested in applications to capitalism. Is there a place on that site other than the one titled "Capitalism" which shows applications to capitalism? I saw nothing there involving fallibilism or acting as if nothing is ever conclusively true.

curi 07 April 2011 08:32:46PM -2 points 

I'll just quickly write something for you:

Capitalism is a part of liberalism. It applies liberal ideas, such as individual freedom, to economic issues, and thus advocates, for example, free trade.

What might we consider instead of freedom? Force.

Liberalism hates force. It wants all disputes to be resolved without the use of force. This leads to capitalist ideas (taking capitalism seriously) like that taxes are a use of force which should be improved on, that people don't have a right to bread (provided by someone else, who becomes in a small way their slave), etc...

The best argument against force comes from fallibilism. This was first discussed by the liberal philosopher William Godwin.

It is: in any disagreement, we might be wrong. The other guy might be right. Therefore, we should not impose our will on him. That isn't truth seeking, and truth seeking is needed because we don't know who is right and shouldn't assume it's us.

Force is inherently irrational because it assumes who is right based on the source of the ideas in question (or, if you prefer, denies the other guy has an idea, or something like that).

Why is it *initiating* force in particular that is bad, but defense is OK? Because defense does not sabotage truth seeking. The outcome *already wasn't going to be decided based on reason* when the first guy initiated force. Defense doesn't cause any new problem.

Capitalist values allow for all voluntary interaction, which is compatible with correcting our mistakes (does not require it, but allows it) and bans non-voluntary interaction in which some party is acting contrary to fallibilism.

Get the idea?

Desrtopa 08 April 2011 09:12:06PM 2 points

This only demonstrates that you can argue in a fallibilist framework for something you can argue for in practically any other philosophical framework as well. Simply showing that your epistemology allows you to do things *as well* as people who don't even know what an epistemology is isn't a rousing argument for its usefulness.

curi 08 April 2011 09:49:49PM * -3 points

What are the other arguments for liberalism, of this quality and just as fundamental?

I read some other philosophies and wasn't able to find great liberal arguments like this. I'd like to hear them.

As an example, Mises has very good arguments for liberalism, but none of them are as fundamental as this. They are all higher level stuff.

Desrtopa 08 April 2011 09:52:17PM 0 points

I'm not convinced that this is an argument of exceptional quality, or fundamental at all, so I'm just going to have to say "most of them."

JoshuaZ 09 April 2011 03:58:41PM 0 points

For what it is worth, I don't think this argument is fundamental in any useful way but it is an argument that I find quite strong and I don't agree with curi's basic philosophy.

curi 08 April 2011 09:54:05PM -4 points

Give one that's better.

Desrtopa 08 April 2011 10:13:22PM 2 points

I'm not convinced that *any* moral argument is fundamental. What do you want one for? Is

that just an "I'd like to see you do better?" challenge? If so, I'm not going to bother, because I doubt it would serve a useful purpose in furthering the conversation, and I don't think anyone was particularly swayed by yours in the first place. Those of us who are already liberals have our own, and those who aren't weren't compelled to change our minds.

If this is a "If I can see an argument for liberalism *that is more convincing to an average person not already entrenched in my philosophy*," then I'll try to provide one, but I'll stipulate that the persuasiveness should be decided by a poll elsewhere, not by either of us. If you can provide a response to my request in my other comment that convinces me that I should continue to be interested in this conversation at all, I would be amenable to that.

Desrtopa 07 April 2011 03:24:50PM * 5 points

I'm not all that sure that this is going anywhere helpful, but since curi has asked for objections to Critical Rationalism, I might as well make mine known.

My first objection is that it attempts to "solve" the problem of induction by doing away with an axiom. Yes, if you try to prove induction, you get circularity or infinite regress. That's what happens when you attempt to prove axioms. The Problem of Induction essentially amounts to noticing that we *have* axioms on which inductive reasoning rests.

Popperian reasoning could be similarly "refuted" simply by refusing to accept any or all of its axioms. The axiom of induction that Critical Rationalism rejects is one that we have every reason to suspect is true, save that we cannot prove it, which is as good as axioms get, so Critical Rationalism is not any more secure.

Second, with respect to the principle of criticism, it gives too much leeway to a mere [clever arguer](#). Philosophy as a discipline is testament to the fact that humans can criticize each others' ideas endlessly without coming to meaningful consensuses. Wittgenstein criticized Popper. Was that the end of Popper? No, Popper and supporters countered criticisms, which were met with counter-counters, and so on till today and almost certainly beyond. For humans operating in natural language, following this principle does not have a good track record of getting people to promote good ideas over bad ones, when compared to distinction based on plausibility and evidence.

curi 07 April 2011 08:11:01PM 2 points

My first objection is that it attempts to "solve" the problem of induction by doing away with an axiom.

Proposed axioms can be mistakes. Do we need that axiom? Popper says the argument that we need it is mistaken. That could be an important and valid insight if true, right?

Popperian reasoning could be similarly "refuted" simply by refusing to accept any or all of its axioms.

You are applying foundationalist and justificationist criticism to a philosophy which has, as one of its big ideas, that those ways of thinking are mistaken. That is not a good answer to Popper's system.

The axiom of induction that Critical Rationalism rejects is one that we have every reason to suspect is true

No, it's worse than that. Try specifying the axiom. How are ideas justified?

If the answer you give is "they are justified by other ideas, which themselves must be justified" then that isn't merely not proven but wrong. That doesn't work.

If the answer you give is, "they are justified by other ideas which are themselves justified, or the following set of foundational ideas ..." then the problem is not merely that you can't prove your foundations are correct, but that *this method of thinking is anti-critical* and not sufficiently fallibilist.

Fallibilism teaches that people make mistakes. A lot. It is thus a bad idea to try to find the perfect truth and set it up to rule your thinking, as the unquestionable foundations. You will set up mistaken ideas in the role of foundations, and that will be bad. What we should do instead is to try set up institutions which are good at (for example):

1) making errors easy to find and question -- highlighting rather than hiding them 2) making ideas easy to change when we find errors

Popper's solution to the problem of justification isn't merely dropping an axiom but involves advocating an approach along these lines (with far more detail than I can type in). It's substantive. And rejecting it on the grounds that "you can't prove it" or "i refuse to accept it" is silly. You should give a proper argument, or reconsider rejecting it.

Second, with respect to the principle of criticism, it gives too much leeway to a mere clever arguer.

It is non-obvious how it doesn't. There is a legitimate issue here. But the problem can be and is dealt with. I do not have this problem in practice, and I know of no argument that I really have it in theory. You give Wittgenstein as an example, and suggest things go back and forth kind of indefinitely. Some replies:

1) stuff gets resolved or figured out. broad consensus get reached. there isn't a rule to *force* people not to be idiots, but being an idiot isn't a fulfilling life and people tire of it when they understand a better way. this isn't easy but it happens.

2) adding a rule to require people to listen to your conception of reason would not solve the problem. they might refuse your premises.

3) I don't agree with your pessimism. most people don't try hard enough to organize their thinking and reach conclusions. but it can be done, in practice. you can look at issues in depth, and actually sort through all the arguments and counter arguments. it takes patience, persistence and effort to do high quality learning, but progress is possible.

following this principle does not have a good track record

As I see it, it is responsible for the enlightenment, in broad terms. i'm not trying to claim all the credit but the things done in the enlightenment were basically compatible with what i'm advocating. Also for the golden age of athens (plato and aristotle are *not* representative of the height of athens, they lived just after. it's the pre-socratics i have in mind, such as xenophanes.)

one aspect of the enlightenment was rebelling against authority (in particular religious authority, and in particular religious authority as applied to science. also in particular, political authority). this is broadly in line with my philosophy. and i think not so much in line with an attempt to set up foundational axioms for everyone to obey to prevent them from thinking wrong.

That said, Popper's insights beyond description of how people already learn have little track record. What they do have is pretty positive. Popper and his philosophy of science in particular has been respected by many scientists great and small, like Einstein, Deutsch, Feynman, Wheeler, Medawar, Eccles, Monod.

Anyway, back to those clever people. They can be dealt with. Wittgenstein is easy to argue with, he's so bad. Do you have an example of some argument you don't know how to resolve and want me to resolve?

Desrtopa 07 April 2011 09:16:19PM * 2 points 

Edit:

I wrote a response up, but I deleted it because I think this is getting too confrontational to be useful. I have plenty of standing objections to Critical Rationalism, but I don't think I can pose them without creating an attitude too adversarial to be conducive to changing either of our minds. I hate to be the one to start bringing this in *again*, but I think perhaps if you want to continue this discussion, you should read the Sequences, at which point you should hopefully have some understanding of how we are convinced that Bayesianism improves people's information processing and decisionmaking from a practical standpoint. I will be similarly open to any explanations of how you feel Critical Rationalism improves these things (let me be *very* clear that I'm not asking for more examples of people who approved of Popper or things you feel Critical Rationalism can take credit for, show me how people who can only be *narrowly* interpreted as Popperian outperform people who are not Popperian.) I have standing objections to Popper's critiques of induction, but this is what I actually care about and am amenable to changing my mind on the basis of.

curi 07 April 2011 09:34:03PM 0 points 

The reason I'm not very interested in carefully reading your Sequences is that I feel they miss the point and

aren't useful (that is, useful to *philosophy*. lots of your math is nice). In my discussions here, I have not found any reason to think otherwise.

show me how people who can only be narrowly interpreted as Popperian outperform people who are not Popperian

Show it how? I can conjecture it. Got a criticism?

Peterdjones 14 April 2011 04:46:49PM 1 point

Which exposes one of the problems with Popperianism..it leads to the burden of proof being shifted to the wrong place. The burden should be with who proposes a claim, or whoever makes the most extraordinary claim. Popperianism turns it into a burden of disproof on the refuter. All you have to do is "get in first" with your conjecture, and you can sit back and relax. "Why, you have to show Barack Obama is NOT an alien".

curi 14 April 2011 05:14:59PM -1 points

In Popperism, there is no burden of proof.

How did you find me here, Peter?

JoshuaZ 10 April 2011 04:11:01AM 0 points

Show it how? I can conjecture it. Got a criticism?

I'm replying a second time to this remark because thinking about it more it illustrates a major problem you are having. You are using a specific set of epistemological tools and notation when that is one of the things that is in fact in dispute. That's unproductive and is going to get people annoyed.

It is all the more severe because many of these situations are cases where the specific epistemology doesn't even matter. For example, the claim under discussion is the claim that " people who can only be narrowly interpreted as Popperian outperform people who are not Popperian" That's something that could be tested regardless of epistemology. To use a similar example, if someone is arguing for Christianity and they claim that Christians have a longer lifespan on average then I don't need to go into a detailed discussion about epistemology to examine the data. If I read a paper, in whatever branch of science, I could be a Popperian, a Bayesian, completely uncommitted, or something else, and still read the paper and come to essentially the same results.

Trying to discuss claims using exactly the framework in question is at best unproductive, and is in general unhelpful.

curi 10 April 2011 04:26:45AM * 0 points

I'm replying a second time to this remark because thinking about it more it illustrates a major problem you are having. You are using a specific set of epistemological tools and notation when that is one of the things that is in fact in dispute. That's unproductive and is going to get people annoyed.

Sure I am. But so are you! We can't help but use epistemological tools. I use the ones I regard as actually working. As do you. I'm not sure what you're suggesting I do instead.

If you want me to recognize what I'm doing, I do. If you want me to consider other toolsets, I have. In depth. (Note, btw, that I am here voluntarily choosing to learn more about how Bayesians think. I like to visit various communities. Simultaneously I'm talking to Objectivists (and reading their books) who have a different way of approach epistemology.) The primary reason some people have accused me (and Brian) of not understanding Bayesian views and other views not our own isn't our unfamiliarity but because we disagree and choose to think differently than them and not to accept or go along with various ideas.

When people do stuff like link me to <http://yudkowsky.net/rational/bayes> it is their mistake to think I

haven't read it. I have. They think if I read it I would change my mind; they are just plain empirically wrong; I did read it and did not change my mind. They ought to learn something from their mistake, such as that their literature is less convincing than they think it is.

On the other hand, no one here has noticeably familiar with Popper. And no one has pointed me to any rigorous criticism of Popper by any Bayesian. Nor any rigorous rebuttal to Popper's criticisms of Bayesianism (especially the most important ones, that is the philosophical not mathematical ones).

The situation is that Popper read up on Bayesian stuff, and many other ideas, engaged with and criticized other ideas, formulated his own ideas, rebutted criticisms of his views, and so on. That is all good stuff. Bayesians do it some too. But they haven't done it with Popper; they've chosen to disregard him based on things like his reputation, and misleading summaries of his work. At best some people here have read his first book, which is not the right one to start with if you want to understand what he's about, and gotten a very unrepresentative picture of what Popper is about. This disregarding of Popper without engaging with the bulk of his work is no good.

The same thing can be found, btw, in Objectivist circles. Here's what happened when my friend asked Harry Binswanger (a big shot who knew Rand personally for a long time) about Popper: Binswanger gave Popper quotes attributed to the wrong book and briefly stated a few myths about Popper. And in one of the quotes he inserted a clarifying word. It was roughly, at the start of the quote: "That doctrine [realism]" when Popper wasn't talking about realism, Binswanger hadn't read (or had misread, but I think hadn't read since he didn't know what book the quotes were from) the context. When confronted with his mistakes he basically ignored them and said "I'm right anyway" except, amazingly, without the "anyway" part. I think you may be happy to jump on the dumb Objectivists. But from my perspective, the reception here hasn't been better. In some ways the Objectivists were superior. They provided some relevant published material on the matter (it was badly wrong, but at least they had something).

It is all the more severe because many of these situations are cases where the specific epistemology doesn't even matter.

As I was talking about in other comments recently (edit: oops, I actually wrote a different post first but haven't managed to post it yet due to the rate limit. it's a reply to you, so you can find it in your inbox in 10 minutes. i'll post it next), *all mistakes matter*. It doesn't work to ignore mistakes thinking they aren't relevant and just keeping going and hope they won't bite you. You know, I've gotten a bunch of flak where people say Popper isn't rigorous enough and Bayesian stuff is more rigorous. But it's not quite like that. Popper thought that certain kinds of formalness used by philosophers were mistakes, said why, and didn't do them (especially in his later works). But for other issues, the Popperian attitude is more rigorous. We don't gloss over small mistakes. We think they all matter! Is that not being more rigorous in a way? Maybe you think the wrong way. But that's a substantive disagreement.

BTW your entire question asking for empirical evidence that a non-empirical philosophy produces better results is itself a product of your epistemological tools. Popperians regard that as something of a bad question. That's why you don't get the direct answer you expect. It's not evasion but disagreement about your premises. Large parts of philosophy are not empirical and can't really be judged empirically. And there's so many issues that make a rigorous answer to what you want impracticable. No philosophy can answer it because there's too many uncontrolled factors. People always have lots of ideas of a variety of types, and imperfect understanding of the philosophy they are associated with.

hairyfigment 10 April 2011 04:44:59AM 0 points

[...]

I'm not sure what you're suggesting I do instead.

Tell us what you think your tool does better, in some area where we see a problem with Bayes. (And I do mean Bayes' Theorem as a whole, not a part of it taken out of context.)

Show it how? I can conjecture it. Got a criticism?

Seems like any process that leads to harmful priors can also produce a criticism of your position *as you've explained it so far*. As I mentioned before, the Consistent Gambler's Fallacy would lead us to criticize any theory that has worked in the past.

JoshuaZ 08 April 2011 02:25:19PM * 0 points [-]

show me how people who can only be narrowly interpreted as Popperian outp

Show it how? I can conjecture it. Got a criticism?

Yes. There's no reason to conjecture this other than your own personal preference. I could conjecture that people with red hair perform better and that would have almost as much basis.

(The mere act of asserting a hypothesis is not a reason to take it seriously.)

Desrtopa 07 April 2011 09:43:09PM 1 point [-]

Responses to criticisms are not interesting to me; proponents of *any* philosophy can respond to criticisms in ways that they are convinced are satisfying, and I'm not impressed that supporters of Critical Rationalism are doing a better job. If you cannot yourself come up with a convincing way to demonstrate that Critical Rationalism results in improved success in ways that supporters of other philosophies cannot, why should I take it seriously?

curi 07 April 2011 09:45:40PM 2 points [-]

What would you find convincing? What convinced you of Bayes' or whatever you believe?

Desrtopa 07 April 2011 09:53:35PM * 1 point [-]

Examples of mistakes in processing evidence people make in real life which lead to bad results, and how Bayesian reasoning resolves them, followed by concrete applications such as the review of the Amanda Knox trial.

Have you already looked at the review of the Amanda Knox trial? If you haven't, it might be a useful point for us to examine.

It doesn't help anyone to point out an example of inductive reasoning, say "this is a mistake" because you reject the foundations of inductive reasoning, but *not* demonstrate how rejecting it leads to better results than accepting it. So far the examples you have given of the supposed benefits of Critical Rationalism have been achievements of people who can only be loosely associated with Critical Rationalism, or arguments in a frame of Critical Rationalism for things that have already been argued for outside a frame of Critical Rationalism.

curi 08 April 2011 12:40:20AM * -5 points [-]

Inductive reasoning doesn't lead to any results, ever.

No one has ever used it.

The theory they have is a mistake.

This cannot be demonstrated in the way you request. It can only be argued. e.g. by beginning with the question: what *precisely* does induction say to do? (which has never been successfully answered.)

have people done stuff similar to induction, and did it work OK? well that depends on philosophical understanding of what is and isn't similar to something that doesn't make sense. i'm not very inclined to start calling any coherent things similar to any incoherent ones.

Many Popperian insights are of this type: they are philosophical ideas.

So far the examples you have given of the supposed benefits of Critical Rationalism have been achievements of people who can only be loosely associated with Critical Rationalism

What are you talking about? I don't think you know much about the philosophies of the people I

listed. They aren't all just loosely associated.

Desrtopa 08 April 2011 08:47:09PM 1 point

This doesn't address my requests at all.

So what if nobody has ever used induction? I'm convinced that Popper is wrong, but without any evidence that following his epistemology produces improved results, I don't see why I should be interested in the possibility that he's right. Even supposing induction is merely an approximation of how we really gain knowledge, it's a computable approximation which produces results that are at least as viable, so there's no reason why it not being the "real" method of knowledge production should matter, for AI or for humans.

What are you talking about? I don't think you know much about the philosophies of the people I listed. They aren't all just loosely associated.

Then explain specifically what each of them have achieved that *could not* have been achieved equally well had they not been Critical Rationalists, and *why* these achievements are due to Critical Rationalism. Or hell, explain what *any* of them have achieved that's unambiguously due to critical rationalism.

curi 08 April 2011 09:25:29PM -1 points

Even supposing induction is merely an approximation

Popper says it's not.

Does that matter to you?

Then explain specifically what each of them have achieved that could not have been achieved equally well had they not been Critical Rationalists, and why these achievements are due to Critical Rationalism. Or hell, explain what any of them have achieved that's unambiguously due to critical rationalism.

You are challenging me to explain things to you which you could learn about on your own if you wanted. You want me to answer questions you chose not to research. That is OK, but...

Before that you were dismissive. So I'm not sure if I want to help answer your questions about scientists. Are you a person with intellectual integrity who is worth talking to? Help me decide. You made a statement about the people I had listed, without knowing much about the people I had listed, and in particular without knowing if they all only have a loose association with CR or not. You falsely asserted they did all have a loose association only. You were mistaken to speak from ignorance about scientists -- assuming I was wrong without even asking -- and now you would like to learn better and change your mind. Is that correct?

Note for example that Deutsch has published two books advocating Popperian philosophy and talking extensively about Popper. That isn't a loose association. Even wikipedia level knowledge of these people would be sufficient not to make the mistake you did. You had less than that level of knowledge and posted anyway. Do you want to apologize, retract your statements, or anything? Or do you want to get mad at me now? I want to test your reaction.

[continue this thread »](#)

JoshuaZ 08 April 2011 01:35:48AM 0 points

Inductive reasoning doesn't lead to any results, ever.

No one has ever used it.

This is so empirically false that I don't know how to approach it. Do you actually think that

when people are saying that they are using induction they really aren't? Note that this isn't the same claim that people shouldn't be using induction or that their induction is unjustified. But claiming they are not using it is just wrong unless you are using some very non-standard terminology under which one could say things like "No one has ever used homeopathy." This seems like an abuse of language.

curi 08 April 2011 09:53:35AM -2 points

Post the method of induction, step by step, in sufficient detail that a reasonable person could do it without having to ask any questions.

When you fail -- in particular by having large unspecified parts -- it will be because you are wrong about the issue in question.

When you respond to this failure by making ad hoc additions that still don't provide followable instructions, then I will stop talking to you.

OK, go ahead.

[continue this thread »](#)

jimrandomh 08 April 2011 01:05:48AM * 1 point

Inductive reasoning doesn't lead to any results, ever. No one has ever used it.

Go read Jaynes' defense of Laplace's rule of succession (which is an example of inductive reasoning) in Chapter 18.

[deleted] 07 April 2011 02:44:13PM 2 points

) What if we have two or more ideas? This one is easy. There is a particular criticism you can use to refute all the remaining theories. It's the same every time so there's not much to remember. It goes like this: idea A ought to tell me why B and C and D are wrong. If it doesn't, it could be better! So that's a flaw. Bye bye A. On to idea B: if B is so great, why hasn't it explained to me what's wrong with A, C and D? Sorry B, you didn't answer all my questions, you're not good enough. Then we come to idea C and we complain that it should have been more help and it wasn't. And D is gone too since it didn't settle the matter either. And that's it. Each idea should have settled the matter by giving us criticisms of all its rivals. They didn't. So they lose. So whenever there is a stalemate or a tie with two or more ideas then they all fail.

This seems absurd, since an explanation like "Phlogiston!", which can "explain" everything because it is a [mysterious answer](#), would pass your test but a [legitimate explanation](#) wouldn't.

curi 07 April 2011 07:50:26PM 2 points

If something can explain everything (by not being adapted to addressing any particular problem) we can criticize it for doing just that. So we dispense with it.

[deleted] 07 April 2011 08:33:19PM 1 point

In that case, you seem to be saying "dispense with a hypothesis if it can't explain everything, and also dispense with it if it does explain everything." How *both* of these can be legitimate reasons for dismissal?

curi 07 April 2011 08:33:56PM * 2 points

If it doesn't explain everything (relevant to some problem you are trying to address), improve it.

If it explains everything *vacuously*, reject it.

Khoth 07 April 2011 10:50:44AM 5 points

You can criticize any idea you want. There's no rules again. If you don't understand it, that's a criticism -- it should have been easier to understand. If you find it confusing, that's a criticism -- it should have been clearer. If you think you see something wrong with it, that's a criticism -- it shouldn't have been wrong it that way, or it should have included an explanation so you wouldn't make a mistaken criticism. This step is easy too Step 3) All criticized ideas are rejected. They're flawed. They're not good enough.

t!;dr

I've just Criticised your idea. Your idea is not good enough. You have to come up with a new one.

Or is there some level of criticism that doesn't count because it's not good enough either?

curi 07 April 2011 07:51:23PM 2 points

Your criticism is generic. It would work on all ideas equally well. It thus fails to differentiate between ideas or highlight a flaw in the sense of something which could possibly be improved on.

So, now that I've criticized your criticism (and the entire category of criticisms like it), we can reject it and move on.

Khoth 07 April 2011 08:25:00PM 3 points

My criticism is not generic. It would not work on an idea which consisted of cute cat pictures. Therefore, your criticism of my criticism does not apply.

I can continue providing specious counter-counter-...counter-criticisms until the cows come home. I don't see how your scheme lets sensible ideas get in edgeways against that sort of thing.

Anyhow, criticism of criticisms wasn't in your original method.

curi 07 April 2011 08:38:51PM 2 points

I can continue providing specious counter-counter-...counter-criticisms

If you understand they are specious, then you have a criticism of it.

Anyhow, criticism of criticisms wasn't in your original method.

Criticisms are themselves ideas/conjectures and should themselves be criticized. And I'm not saying this ad hoc, I had this idea before posting here.

Khoth 07 April 2011 08:42:05PM 0 points

I understand they are specious, but I'm not using your epistemology to determine that. What basis do you have for saying that they are specious?

curi 07 April 2011 08:54:35PM 1 point

It doesn't engage with the substance of my idea. It does not explain what it regards as a flaw in the idea.

Unless you meant the t!;dr as your generic criticism and the flaw you are trying to explain is that all good idea should be short and simple. Do you want me to criticize *that?* :-)

Khoth 07 April 2011 09:18:50PM 3 points 

What I'm trying to get at is: By your system, the idea to be accepted is the one without an uncountered criticism. What matters isn't any external standard of whether the criticism is good or bad, just whether it has been countered. But any criticism, good or bad, can be countered by a (probably bad) criticism, so your system doesn't offer a way to distinguish between good criticism and bad criticism.

curi 07 April 2011 09:36:43PM 2 points 

You have to conjecture standards of criticism (or start with cultural ones). Then improve them by criticism, and perhaps by conjecturing new standards.

If you want to discuss some specific idea, say gardening, you can't discuss *only* gardening in a very isolated way. You'll need to at least implicitly refer to a lot of background knowledge, including standards of criticism.

One way this differs from foundations is if you think a standard of criticism reaches the wrong conclusion about gardening, you can argue from your knowledge of gardening *backwards* (as some would see it) to criticize the standard of criticism for getting a wrong answer.

[deleted] 07 April 2011 09:40:08PM 1 point 

How can you expect that criticizing your standards of criticism will be productive if you don't *have* a good standard of criticism in the first place?

curi 07 April 2011 09:43:55PM 1 point 

Many starting points work fine.

In theory, could you get stuck? I don't have a proof either way.

I don't mind too much. Humans already have standards of criticism which don't get stuck. We have made scientific progress. Our standards we already have allow self-modification and thereby unbounded progress. So it doesn't matter what would have happened if we had started with a bad standard once upon a time, we're past that (it does matter if we want to create an AI).

[continue this thread »](#)

prase 07 April 2011 08:02:32PM 0 points 

Doesn't this create an infinite regress of criticisms, if you try hard enough? (Your counter-criticism is also generic, when it applies to the whole category.)

curi 07 April 2011 08:19:21PM 2 points 

If you try hard enough you can refuse to think at all.

Popperian epistemology helps people learn who want to. It doesn't provide a set of rules that, if you follow them exactly while trying your best not to make progress, then you will learn anyway. We only learn much when we seriously try to, with good intentions.

You can always create trivial regresses, e.g. by asking "why?" infinitely many times. But that's different than the following regress:

If you assert "theories should be justified, or they are crap"

and you assert "theories are justified in one way: when they are supported by a theory which is itself

justified"

Then you have a serious problem to deal with which is not the same type as asking "why?" forever.

Things which reject entire categories is not a precise way to state what theories should be rejected. You are correct that the version I wrote can be improved to be clearer and more precise. One of the issues is whether a criticism engages with the substance of the idea it is criticizing, or not. "All ideas are wrong" (for example) doesn't engage with any of the explanations that the ideas it rejects give, it doesn't point out flaws in them, it doesn't help us learn. Criticisms which don't help us learn better are no good -- the whole purpose and meaning of criticism, as we conceive it, is you explain a flaw so we can learn better.

One issue this brings up is that communication is never 100% precise. There is always ambiguity. If a person wants to, he can interpret everything you say in the worst possible way. If he does so, he will sabotage your discussion. But if he follows Popper's (not unique or original) advice to try to interpret ideas he hears as the *best* version they could mean -- to try to figure out good ideas -- then the conversation can work better.

cousin_it 07 April 2011 08:26:45AM * 10 points

Regarding the technical side of your post, if a Bayesian computer program assigns probability 0.87 to proposition X, then obviously it ought to assign probability 1 to the fact that it assigns probability 0.87 to proposition X. (If you don't trust your own transistors, add error-correction at the level of transistors, don't contaminate the software.) But it's hard to think of a situation where the program will need to make use of the latter probability.

Regarding the substance, I think you disagree with popular opinion on LW because there are two possible meanings of "epistemology":

- 1) If a human wants to have rational beliefs, what rulebook should they follow?
- 2) If we want to write a computer program that arrives at rational beliefs, what algorithms should we use?

From your posts and comments it looks like you're promoting Popperianism as an answer to (1). The problem is, it's pretty hard to determine whether a given answer to (1) is right, wrong or meaningless, when it's composed of mere words (cognitive black boxes) and doesn't automatically translate to an answer for (2). So most LWers think that (2) is really the right question to ask, and any non-confused answer to (2) ought to dissolve any leftover confusion about (1).

PS: this might be slightly off topic, but any discussion of anti-Bayesianism ought to contain links to [two texts](#) by Cosma Shalizi, the most interesting anti-Bayesian that I know of.

timtyler 07 April 2011 04:14:03PM * -4 points

Regarding the technical side of your post, if a Bayesian computer program assigns probability 0.87 to proposition X, then obviously it ought to assign probability 1 to the fact that it assigns probability 0.87 to proposition X.

I am pretty sure that is wrong. For one thing it would be overconfident. For another [0 and 1 are not probabilities](#).

But it's hard to think of a situation where the program will need to make use of the latter probability.

It's a measure of how much confidence there is in the estimate, so it could be used when updating in response to evidence. High confidence there would mean that it takes a lot of new evidence to shift the 0.87 estimate.

cousin_it 07 April 2011 04:47:01PM * 5 points

Your last paragraph is wrong. Here's an excruciatingly detailed explanation.

Let's say I am a perfect Bayesian flipping a possibly biased coin. At the outset I have a uniform prior over all possible biases of the coin between 0 and 1. Marginalizing (integrating) that prior, I assign 50% probability to the event of seeing heads on the first throw. Knowing my own neurons perfectly, I believe all the above statements with probability 100%.

The first flip of the coin will still make me update the prior to a posterior, which will have a different mean. Perfect knowledge of myself doesn't stop me from that.

Now skip forward. I have flipped the coin a million times, and about half the results were heads. My current probability assignment for the next throw (obtained by integrating my current prior) is 50% heads and 50% tails. I have monitored my neurons diligently throughout the process, and am 100% confident of their current state.

But it will take much more evidence now to change the 50% assignment to something like 51%, because my prior is very concentrated after seeing a million throws.

The statement "I have perfect knowledge of the current state of my prior" (and its integral, etc.) does not in any way imply that "my current prior is very concentrated around a certain value". It is the latter, not the former, that controls my sensitivity to evidence.

Sniffnoy 07 April 2011 10:45:59PM 0 points

Upvoted for giving a good explanation where I failed earlier...

timtyler 07 April 2011 05:38:25PM 0 points

Your last paragraph is wrong. Here's an excruciatingly detailed explanation.

That does clarify what you originally meant. However, this still seems "rather suspicious" - due to the 1.0:

if a Bayesian computer program assigns probability 0.87 to proposition X, then obviously it ought to assign probability 1 to the fact that it assigns probability 0.87 to proposition X.

cousin_it 07 April 2011 05:43:03PM * 2 points

I'm willing to bite the bullet here because all hell breaks loose if I don't. We don't know how a Bayesian agent can ever function if it's allowed (and therefore required) to doubt arbitrary mathematical statements, including statements about its own algorithm, current contents of memory, arithmetic, etc. It seems easier to just say 1.0 as a stopgap. Wei Dai, paulfchristiano and I have been thinking about this issue for some time, with no results.

curi 07 April 2011 08:59:14AM -1 points

I don't agree with either meaning of epistemology. The traditional meaning of epistemology, which I accept, is the study of knowledge, and in particular questions like What is knowledge? and How do we sort out good ideas from bad ideas? and How is knowledge created?

Both of your definitions of the field have bayesian ways of thinking already built in to them. They are biased.

If you don't want to be an epistemology, that would be OK with me. But for example Yudkowsky claimed that Bayesianism was dethroning Popperism. To do that it has to be an epistemology and deal with the same questions Popper addresses.

Popperian epistemology does not offer any rulebook. It says rulebooks are an authoritarian and foundationalist mistake, which comes out of the attempt to find a source of justification. (Well, the psychological claims are not important and not epistemology. But Popper did occasionally say things like that, and I think it's true)

I will take a look at your links, thanks. I respect that author a lot for this post on why heritability studies are wrong:

<http://cscs.umich.edu/~crshalizi/weblog/520.html>

(1). The problem is, it's pretty hard to determine whether a given answer to (1) is right, wrong or meaningless, when it's composed of mere words (cognitive black boxes) and doesn't automatically translate to an answer for (2). So most LWers think that (2) is really the right question to ask, and any non-confused answer to (2) ought to dissolve any leftover confusion about (1).

Note that Popperians think there is no algorithm that automatically arrives at rational beliefs. There's no privileged road to truth. AIs will not be more rational than people. OK they usually won't have a few uniquely human flaws (like, umm, caring if they are fat). But there is no particular reason to expect this stuff will be replaced with correct ideas. Whatever AIs think of instead will have its own mistakes. It's the same kind of issue as if some children were left on a deserted island to form their own culture. They'll avoid various mistakes from our culture, but they will also make new ones. The rationality of AIs, just like the rationality of the next generation, depends primarily on the rationality of the educational techniques used (education is closely connected to epistemology in my view, because it's about learning, i.e. creating knowledge. Popperian epistemology has close connections to educational theory which led to the philosophy "Taking Children Seriously" by David Deutsch).

[deleted] 07 April 2011 01:15:52PM * 3 points

Note that Popperians think there is no algorithm that automatically arrives at rational beliefs. There's no privileged road to truth. AIs will not be more rational than people. OK they usually won't have a few uniquely human flaws (like, umm, caring if they are fat). But there is no particular reason to expect this stuff will be replaced with correct ideas. Whatever AIs think of instead will have its own mistakes. It's the same kind of issue as if some children were left on a deserted island to form their own culture. They'll avoid various mistakes from our culture, but they will also make new ones. The rationality of AIs, just like the rationality of the next generation, depends primarily on the rationality of the educational techniques used (education is closely connected to epistemology in my view, because it's about learning, i.e. creating knowledge).

This is mostly irrelevant to your main point, but I'm going to talk about it because it bothered me. I don't think anyone on LessWrong would agree with this paragraph, since it assumes a whole bunch of things about AI that we have good reasons to *not* assume. The rationality of an AI will depend on its mind design--whether it has biases built into its hardware or not us up to us. In other words, you can't assert that AIs will make their own mistakes because this assumes things about the mind design of the AI, things that we *can't* assume because we haven't built it yet. Also, even if an AI does have its own cognitive biases, it still might be orders of magnitude more rational than a human being.

JoshuaZ 07 April 2011 07:50:13PM 0 points

Also, even if an AI does have its own cognitive biases, it still might be orders of magnitude more rational than a human being.

Or orders of magnitude less rational. This isn't terribly germane to your original point but it seemed worth pointing out. We really have no good idea what the minimum amount of rationality actually is for an intelligent entity.

[deleted] 07 April 2011 08:35:44PM 0 points

Oh, I definitely agree with that. It's certainly possible to conceive of a really, really, really suboptimal mind that is still "intelligent" in the sense that it can attempt to solve problems.

curi 07 April 2011 07:13:38PM 2 points

I'm not assuming stuff by accident. There is serious theory for this. AI people ought to learn these ideas and engage with them, IMO, since they contradict some of your ideas. If we're right, then you need to make some changes to how you approach AI design.

So for example:

The rationality of an AI will depend on its mind design--whether it has biases built into its hardware or not us up to us.

If an AI is a universal knowledge creator, in what sense can it have a built in bias?

timtyler 07 April 2011 07:22:09PM 2 points

I'm not assuming stuff by accident. There is serious theory for this. AI people ought to learn these ideas and engage with them, IMO, since they contradict some of your ideas.

Astrology also conflicts with "our ideas". That is not in itself a compelling reason to brush up on our astrology.

[deleted] 07 April 2011 07:19:50PM 0 points

If an AI is a universal knowledge creator, in what sense can it have a built in bias?

I don't understand this sentence. Let me make my view of things clearer: An AI's mind can be described by a point in [mind design space](#). Certain minds (most of them, I imagine) have cognitive biases built into their hardware. That is, they function in suboptimal ways because of the algorithms and heuristics they use. For example: [human beings](#). That said, what is a "universal knowledge creator?" Or, to frame the question in the terms I just gave, what is its mind design?

curi 07 April 2011 07:27:51PM 2 points

Certain minds (most of them, I imagine) have cognitive biases built into their hardware.

That's not what mind design space looks like. It looks something like this:

You have a bunch of stuff that isn't a mind at all. It's simple and it's not there yet. Then you have a bunch of stuff that is a fully complete mind capable of anything that any mind can do. There's also some special cases (you could have a very long program that hard codes how to deal with every possible input, situation or idea). AIs we create won't be special cases of that type which are a bad kind of design.

This is similar to the computer design space, which has no half-computers.

what is a "universal knowledge creator?"

A knowledge creator can create knowledge in some repertoire/set. A universal can do any knowledge creation that any other knowledge creator can do. There is nothing in the repertoire of some other knowledge creator, but not its own.

Human beings are universal knowledge creators.

Are you familiar with universality of computers? And how very simple computers can be universal? There's a lot of parallel issues.

[deleted] 07 April 2011 08:45:52PM 1 point

You have a bunch of stuff that isn't a mind at all. It's simple and it's not there yet. Then you have a bunch of stuff that is a fully complete mind capable of anything that any mind can do. There's also some special cases (you could have a very long program that hard codes how to deal with every possible input, situation or idea). AIs we create won't be special cases of that type which are a bad kind of design. This is similar to the computer design space, which has no half-computers.

I'm somewhat skeptical of this claim--I can design a mind that has the functions $O(n)$ (zero function), $S(n)$ (successor function), and $P(x_0, x_1, \dots, x_n)$ (projection function) but not primitive recursion, I can compute most but not all functions. So I'm skeptical of this "all or little" description of mind space and computer space.

However, I suspect it ultimately doesn't matter because your claims don't directly contradict my original point. If your categorization is correct and human beings are indeed universal knowledge creators, that doesn't preclude the possibility of us having cognitive biases (which it had better not do!). Nor does it contradict the larger point, which is that cognitive biases come from cognitive architecture, i.e. where one is located in mind design space.

Are you familiar with universality of computers? And how very simple computers can be

universal? There's a lot of parallel issues.

If you're referring to Turing-completeness, then yes I am familiar with it.

curi 07 April 2011 09:03:38PM 3 points

I'm somewhat skeptical of this claim--I can design a mind that has the functions $O(n)$ (zero function), $S(n)$ (successor function), and $P(x_0, x_1, \dots, x_n)$ (projection function) but not primitive recursion, I can compute most but not all functions. So I'm skeptical of this "all or little" description of mind space and computer space.

How is that a mind? Maybe we are defining it differently. A mind is something that can create knowledge. And a lot, not just a few special cases. Like people who can think about all kinds of topics such as engineering or art. When you give a few simple functions and don't even have recursion, I don't think it meets my conception of a mind, and I'm not sure what good it is.

If your categorization is correct and human beings are indeed universal knowledge creators, that doesn't preclude the possibility of us having cognitive biases (which it had better not do!).

In what sense can a bias be very important (in the long term), if we are universal? We can change it. We can learn better. So the implementation details aren't such a big deal to the result, you get the same kind of thing regardless.

Temporary mistakes in starting points should be expected. Thinking needs to be mistake tolerant.

cousin_it 07 April 2011 09:16:50AM * 6 points

I'm willing to reformulate like this:

- 1) How can a human sort out good ideas from bad ideas?
- 2) How can a computer program sort out good ideas from bad ideas?

and the subsequent paragraph can stay unchanged. Whatever recipe you're proposing to improve human understanding, it ought to be "reductionist" and apply to programs too, otherwise it doesn't meet the LW standard. Whether AIs can be more rational than people is beside the point.

curi 07 April 2011 09:43:32AM 1 point

I don't think you understood the word "reductionist". Reductionism doesn't mean that things *can* be reduced to lower levels but that they *should* -- it actually objections to high level statements and considers them worse. There's no need for reductionism of that kind for ideas to be applicable to low level issues like being programmable.

Yes Popperian epistemology can be used for an AI with the reformulations (at least: I don't know any argument that it couldn't).

Why aren't we there yet? There aren't a lot of Popperians, Popperian philosophy does not seek to be formal which makes it harder to translate into code, and most effort has been directed at human problems (including criticizing large mistakes plaguing the field of philosophy, and which also affect regular people and permeate our culture). The epistemology problems important to humans are not all the same as the ones important to writing an AI. For an AI you need to worry about what information to start it with. Humans are born with information, we don't yet have the science to control that, so there's is only limited reason to worry about it. Similarly there is the issue of how to educate a very young child. No one knows the answer to that *in words* -- they can do it by following cultural traditions but they can't explain it. But for AIs, how to deal with the very young stages is important.

Broadly an AI will need a conjecture generator, a criticism generator, and a criticism evaluator. Humans

have these built in. So again the problems for AI are somewhat different than what's important for, e.g., explaining epistemology to human adults.

You may think the details of these things in humans are crucially important. The reason they aren't is that they are universal, so implementation details don't affect anything much about our lives.

It's still interesting to think about. I do sometimes. I'll try to present a few issues. In abstract terms we would be content with a random conjecture generator, and with sorting through infinitely many conjectures. But we can't program it like that -- too slow. You need shortcuts. A big one is you generate new conjectures by taking old conjectures and making random but limited changes to them. How limited is a good idea? I don't know how to quantify that. Moving on, there is an issue of: do you wait until conjectures are created and then criticize them afterwards? Or do you program it in such a way that conjectures which would be refuted by a criticism can sometimes not be generated in the first place, as a kind of optimization? I lean towards the second view, but I don't know how to code it. I'm partial to the notion of using criticisms as filters on the set of possible conjectures. There's no danger of getting stuck, or losing universality if the filters can be disabled as desired, and modified as desired, and they don't prevent conjectures that would want to modify them. That raises another issue which is: can people think themselves into a bad state they can't get out of? I don't know if that's *impossible* or not. I don't think it happens in practice (yes people can be really dumb, but i don't think they are even close to *impossible* to get out of). If it was technically possible for an AI to get stuck, would that be a big deal? You can see here perhaps some of the ways I don't care for rulebooks.

BTW one of the things our theory tells us is you can never build half an AI. It will jump straight from very minimal functionality to universal functionality, just as computer programming languages do. (The "jump to universality" is discussed by David Deutsch in *The Beginning of Infinity*). One thing this means is there is no way to know how far along we are -- the jump could come at any time with one new insight.

Whether AIs can be more rational than people is beside the point.

Is it? What good are they, then? I have some answers to that, but nothing really huge. If they aren't assumed to be super rational geniuses then they can't be expected to quickly bring about the singularity or that kind of thing.

timtyler 07 April 2011 04:19:11PM 2 points



Reductionism doesn't mean that things can be reduced to lower levels but that they should -- it actually objections to high level statements and considers them worse.

Conventionally, and confusingly, the word *reductionism* has [two meanings](#):

Reductionism can either mean (a) an approach to understanding the nature of complex things by reducing them to the interactions of their parts, or to simpler or more fundamental things or (b) a philosophical position that a complex system is nothing but the sum of its parts, and that an account of it can be reduced to accounts of individual constituents.

timtyler 07 April 2011 01:38:37PM * 6 points



BTW one of the things our theory tells us is you can never build half an AI. It will jump straight from very minimal functionality to universal functionality, just as computer programming languages do. (The "jump to universality" is discussed by David Deutsch in *The Beginning of Infinity*). One thing this means is there is no way to know how far along we are - the jump could come at any time with one new insight.

That sounds pretty bizarre. So much for the idea of progress via better and better compression and modeling. However, it seems pretty unlikely to me that you *actually* know what you are talking about here.

curi 07 April 2011 07:10:30PM 1 point



Insulting my expertise is not an argument. (And given you know nothing about my expertise, it's a silly too. Concluding that people aren't experts because you disagree with them is biased and closed minded.)

Are you familiar with the topic? Do you want me to give you a lecture on it? Will you read about it?

cousin_it 07 April 2011 09:55:15AM * 0 points

Is it? What good are they, then?

I didn't say it was false, just irrelevant to the current discussion of what we want from a theory of knowledge.

You could use math instead of code. To take a Bayesian example, the [Solomonoff prior](#) is uncomputable, but well-defined mathematically and you can write computable approximations to it, so it counts as progress in my book. To take a non-Bayesian example, [fuzzy logic](#) is formalized enough to be useful in applications.

Anyway, I think I understand where you're coming from, and maybe it's unfair to demand new LW-style insights from you. But hopefully you also understand why we like Bayesianism, and that we don't even think of it at the level you're discussing.

curi 07 April 2011 10:33:06AM * 1 point

I understand some. But I think you're mistaken and I don't see a lot to like when judged by the standards of good philosophy. Philosophy is important. Your projects, like inventing an AI, will run into obstacles you did not foresee if your philosophy is mistaken.

Of course I have the same criticism about people in all sorts of other fields. Architects or physicists or economists who don't know philosophy run into problems too. But claiming to have an epistemology, and claiming to replace Popper, those are things most fields don't do. So I try to ask about it. Shrug.

I think I figured out the main idea of Bayesian epistemology. It is: Bayes' theorem is the source of justification (this is intended as the solution to the problem of justification, which is a bad problem).

But when you start doing math, it's ignored, and you get stuff right (at least given the premises, which are often not realistic, following the proud tradition of game theory and economics). So I should clarify: that's the main *philosophical* claim. It's not very interesting. Oh well.

[deleted] 07 April 2011 02:46:11PM 1 point

I think I figured out the main idea of Bayesian epistemology. It is: Bayes' theorem is the source of justification (this is intended as the solution to the problem of justification, which is a bad problem).

No. See [here](#), where Eliezer specifically says that this is not the case. ("But first, let it be clearly admitted that the rules of Bayesian updating, do not of themselves solve the problem of induction.")

curi 07 April 2011 07:05:28PM 1 point

I had already seen that.

Note that I said justification not induction.

I don't want to argue about this. If you like the idea, enjoy it. If you don't, just forget about it and reply to something else I said.

Matt_Simpson 07 April 2011 08:08:00AM * 2 points

Popperian epistemology still relies on deductive logic. Why is deductive logic trustworthy? (Serious question, I think

it illuminates the nature of foundations)

You might argue that we conjecture that deductive logic, as we know it, is true/valid/correct and nothing that we've come up with seems to refute it - yet that doesn't mean we've "proved" that deductive logic is correct. I would go further with some positive arguments, but we'll leave it at that for now.

A Bayesian might argue that the basic assumptions that go into Bayesian epistemology (The assumptions for Cox's theorem + some assumptions that yield prior distributions) have the same status as the rules of logic - we conjecture that they're true and they stand up to criticism, yet we don't think we've proven these assumptions.

This is my understanding of Bayesian epistemology - something like (sophisticated) Popperian falsificationism / critical rationalism provides support for the *assumptions* of Bayesian epistemology. I would argue that you can go farther than falsificationism and actually give some positive arguments for the foundations, but that's irrelevant, really. Even without the positive arguments, the assumptions of Cox's theorem seem to stand up - as in, not get knocked down.

timtyler 07 April 2011 04:27:21PM 0 points



A Bayesian might argue that the basic assumptions that go into Bayesian epistemology (The assumptions for Cox's theorem + some assumptions that yield prior distributions) have the same status as the rules of logic - we conjecture that they're true and they stand up to criticism, yet we don't think we've proven these assumptions.

I don't think I have heard that argued. The problem of the reference machine in Occam's razor leads to a million slightly-different variations. That seems much more dubious than deduction does.

curi 07 April 2011 08:36:21AM 2 points



Popperian epistemology still relies on deductive logic.

Uses, especially in science. Doesn't rely on in any fundamental way for the main philosophical ideas.

Why is deductive logic trustworthy? (Serious question, I think it illuminates the nature of foundations)

It is not "trustworthy". But I don't have a criticism of it. I don't reject ideas for no reason. Lacks of positive justification isn't a reason (since nothing has justification, lack of it cannot differentiate between theories, so it isn't a criticism to lack it).

What is "trustworthy" in an everyday sense, but not a strict philosophical sense, is *knowledge*. The higher quality an idea (i.e. the more it's been improved up to now), the more trustworthy since the improvements consist of getting rid of mistakes, so there's less mistakes to bite you. But how many mistakes are left, and how big are they? Unknown. So, it's not trustworthy in any kind of definitive way.

You might argue that we conjecture that deductive logic, as we know it, is true/valid/correct and nothing that we've come up with seems to refute it - yet that doesn't mean we've "proved" that deductive logic is correct.

Yes. We do conjecture it. And it's not proved. So what?

A Bayesian might argue that the basic assumptions that go into Bayesian epistemology (The assumptions for Cox's theorem + some assumptions that yield prior distributions) have the same status as the rules of logic.

A difference between Popperian views and justificationist ones is Popperian views don't say anything should be justified and then fail to do it. But if you do say things should be justified, and then fail to do it, that is bad.

When Bayesians or inductivists set out to justify theories (and we mean that word broadly, e.g. "support" is a type of justification. So is "having high probability" when it is applied to ideas rather than to events.), they are proposing something rather different from logic. A difference is: justificationism has been criticized while logic hasn't. The criticism is: if you say that theories should be justified, and you say that they gain their justification from other ideas which are themselves justified, then you get a regress. And if you don't say that, you face the question: how are ideas to be justified? And you better have an answer. But no known answers work.

So justification has a different kind of status than logic. And also, if you accept justificationism then you face the problem of justifying logic. But if you don't, then you don't. So that's why you have to, but we don't.

So you might wonder what a non-justificationist Bayesian epistemology would look like. If you're interested, maybe you could tell me. I certainly do think that Bayes' theorem itself is correct, but I'm not convinced it has any important applications to epistemology. I think that trying to have it play the role of justifying ideas is a mistake.

This is my understanding of Bayesian epistemology - something like (sophisticated) Popperian falsificationism / critical rationalism provides support for the assumptions of Bayesian epistemology. Then once you have the assumptions, well the rest falls out.

Popperians aren't overly attached to any particular idea. Our favorites are things like fallibilism, not deduction. But we don't have the structure of having some foundational ideas and then "the rest falls out". We regard that kind of structure as fragile and bad. Popper said knowledge is like a "woven web". There's no up and down, no foundations and derivative parts, no preferred directions, and no simple structure like $A \rightarrow B \rightarrow C$. Everything is interconnected in messy fashion (much more so than real spider webs, which actually have relatively simple geometric patterns). And you are permitted to start in the middle, or anywhere, even in mid air. Whenever you want. It doesn't matter. You can conjecture *anything* with no support, not just foundational ideas.

One of the problems with trying to use Popperian ideas for your foundations, then forgetting about them, is that *they say you shouldn't*. If you let them in as your foundations, they will immediately tell you what to do next, and it isn't Bayesian epistemology! They will tell you to be Popperians, and also to stop being foundationalists, they will not endorse your attempt to have "the rest fall out". If Popper was right enough to serve as a foundation, why is he wrong about all the rest?

Another aspect of your approach is *reductionism*. You treat low level theories are more important. We consider that a mistake. There's nothing wrong with emergent properties. There is nothing wrong with arguing from a higher level idea to a lower level one. Higher level ideas are just as valid as any others.

JoshuaZ 08 April 2011 02:48:27AM 0 points

It is not "trustworthy". But I don't have a criticism of it. I don't reject ideas for no reason.

Um, there's a lot of criticism out there of deductive logic. For one thing, humans often make mistakes in deductive logic so one doesn't know if something is correct. For another, some philosophers have rejected the law of the excluded middle. Yet others have proposed logical systems which try to localize contradictions and prevent explosions (under the sensible argument that when a person is presented with two contradictory logical arguments that look valid to them they don't immediately decide that the moon is made of green cheese). There's a lot to criticize about deductive logic.

[deleted] 07 April 2011 06:57:40PM * 1 point

Another aspect of your approach is reductionism. You treat low level theories are more important. We consider that a mistake. There's nothing wrong with emergent properties. There is nothing wrong with arguing from a higher level idea to a lower level one. Higher level ideas are just as valid as any others.

No, that's incorrect. That may be how other philosophers use the term, but [that's not what it means here](#).

Edit: To clarify, I mean that LessWrong doesn't define reductionism the same way you just did, so your argument doesn't apply.

prase 07 April 2011 01:46:49PM * 2 points

Why is deductive logic trustworthy? (Serious question, I think it illuminates the nature of foundations)

It is not "trustworthy". But I don't have a criticism of it.

I second here [Khoth's comment](#). How do you decide about validity of a criticism? There are certainly people who don't understand logic, and since you have said

You can criticize any idea you want. There's no rules again. If you don't understand it, that's a criticism -- it should have been easier to understand.

doesn't it mean that you actually have a criticism of logic? Or does only count that you personally don't criticise it? If so, how this approach is different from accepting any idea at your wish? What's the point of having an epistemology when it actually doesn't constrain your beliefs in any way?

A technical question: how do I make nested quotes?

curi 07 April 2011 06:42:20PM * 2 points

How do you decide about validity of a criticism?

You conjecture standards of criticism, and use them. If you think they aren't working well, you can criticize them within the system and change the, or you can conjecture new standards of criticism and use those. Note: this has already been done, and we already have standards of criticism which work pretty well and which allow themselves to be improved. (They are largely *not* uniquely Popperian, but well known.)

Different aspect: in general, all criticisms always have some valid point. If someone is making a criticism, and it's wrong, then why wasn't he helped enough not to do that? Theories should be clear and help people understand the world. If someone doesn't get it then there is room for improvement.

doesn't it mean that you actually have a criticism of logic?

I don't regard logic as 'rules', in this context. But terminology is not important. The way logic figures into Popperian critical discussions is: if an idea violates logic you can criticize it for having done so. It would then in theory be possible to defend it by saying why this idea is out of the domain of logic or something (and of course you can point out if it doesn't actually violate logic) -- there's no rule against that. But no one has ever come up with a good argument of that type.

prase 07 April 2011 08:07:14PM * 1 point

Isn't this

all criticisms always have some valid point

contradicting this

no one has ever come up with a good argument of that type

?

I mean, if you can judge arguments and say whether they are good, doesn't it mean that there are bad arguments which don't have a valid point?

curi 07 April 2011 08:13:14PM 2 points

All criticisms have some kind of point, e.g. they might highlight a need for something to be explained better. This is compatible with saying no one ever came up with a good argument (good in the context of modern knowledge) for the Earth being flat, or something. If someone thinks the Earth is flat, then this is quite a good criticism of something -- and I suspect that something is his own background knowledge. We could discuss the matter. If he had some argument which addresses my round-earth views, I'd be interested. Or he might not know what they are. Shrug.

Matt_Simpson 07 April 2011 02:34:25PM * 0 points

If this quote is nested, put two >'s in front of the part you want to be quoted twice.

prase 07 April 2011 07:55:00PM * 0 points

This works for me. However, I want to quote something inside a quote *and then continue on the first level*, such as

inner quote outer quote

The text in italic should be one quoting level deeper.

jimrandomh 07 April 2011 07:59:53PM 1 point

>> Inner quote
>
> Outer quote

Yields

| Inner quote
| Outer quote

prase 07 April 2011 08:03:15PM 0 points

Thanks!

FAWS 07 April 2011 07:28:02AM * 5 points

You assign T0 a probability, say 99.999%. Never mind how or why, the probability people aren't big on explanations like that. Just do your best. It doesn't matter. Moving on, what we have to wonder if that 99.999% figure is correct.

Subjective probabilities don't work like that. Your subjective probability just is what it is. In Bayesian terms the closest thing to a "real" probability is whatever probability estimation is the best you can do with the available data. There is no "correct" or "incorrect" subjective probability, just predictably doing worse than possible to different degrees.

[deleted] 07 April 2011 08:06:00AM * 0 points

Are you saying that there is no regress problem? [Yudkowsky disagrees](#). And so do other commenters here, one of whom called it a "necessary flaw".

Manfred 07 April 2011 03:24:20PM * 0 points

Yeah, you're making a flawed argument by analogy. "There's an infinite regress in deductive logic, so therefore any attempt at justification using probability will also lead to an infinite regress." The reason that probabilistic justification doesn't run into this (or at least, not the exact analogous thing) is that "being wrong" is a *definite state with known properties*, that is *taken into account when you make your estimate*. This is very unlike deductive logic.

timtyler 07 April 2011 01:35:31PM 0 points

That essay seems pretty yuck to me.

Agent beliefs don't normally regress to before they were conceived. They get assigned some priors around when they are born - usually by an evolutionary process.

[deleted] 08 April 2011 12:09:18AM 1 point

I'm not clear on what you are saying.

FAWS 07 April 2011 08:16:39AM 3 points

Are you saying that there is no regress problem?

No, just that it doesn't manifest itself in the form of a pyramid of probabilities of probabilities being "correct". There certainly is the problem of priors, and the justification for reasoning that way in the first place (which were sketched by others in the other thread).

Matt_Simpson 07 April 2011 07:56:33AM * 1 point

There is no "correct" or "incorrect" subjective probability, just predictably doing worse than possible to different degrees.

There is a correct $P(T_0|X)$ where X is your entire state of information. Probabilities aren't strictly speaking subjective, they're [subjectively objective](#).

FAWS 07 April 2011 08:35:06AM * 1 point

"Subjectively objective" just means that trying to do the best you can doesn't leave any room for choice. You can argue that you aren't really talking about probabilities if you knowingly do worse than you could, but that's just a matter of semantics.



MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

Bayesian Epistemology vs Popper

-1 curi 06 April 2011 11:50PM

I was directed to this book (<http://www-biba.inrialpes.fr/Jaynes/prob.html>) in conversation here:

http://lesswrong.com/lw/3ox/bayesianism_versus_critical_rationalism/3ug7?context=1#3ug7

I was told it had a proof of Bayesian epistemology in the first two chapters. One of the things we were discussing is Popper's epistemology.

Here are those chapters:

<http://www-biba.inrialpes.fr/Jaynes/cc01p.pdf>

<http://www-biba.inrialpes.fr/Jaynes/cc02m.pdf>

I have not found any proof here that Bayesian epistemology is correct. There is not even an attempt to prove it. Various things are assumed in the first chapter. In the second chapter, some things are proven given those assumptions.

Some first chapter assumptions are incorrect or unargued. It begins with an example with a policeman, and says his conclusion is not a logical deduction because the evidence is logically consistent with his conclusion being false. I agree so far. Next it says "we will grant that it had a certain degree of validity". But I will not grant that. Popper's epistemology explains that *this is a mistake* (and Jaynes makes no attempt at all to address Popper's arguments). In any case, simply assuming his readers will grant his substantive claims is no way to argue.

The next sentences blithely assert that we all reason in this way. Jaynes' is basically presenting the issues of this kind of reasoning as his topic. This simply ignores Popper and makes no attempt to prove Jaynes' approach is correct.

Jaynes goes on to give syllogisms, which he calls "weaker" than deduction, which he acknowledges are not deductively correct. And then he just says we use that kind of reasoning all the time. That sort of assertion only appeals to the already converted. Jaynes starts with arguments which appeal to the "intuition" of his readers, not on arguments which could persuade someone who disagreed with him (that is, good rational arguments). Later when he gets into more mathematical stuff which doesn't (directly) rest on appeals to intuition, it does rest on the ideas he (supposedly) established early on with his appeals to intuition.

The outline of the approach here is to quickly gloss over substantive philosophical assumptions, never provide serious arguments for them, take them as common sense, do not detail them, and then later provide arguments which are rigorous *given the assumptions glossed over earlier*. This is a mistake.

So we get, e.g., a section on Boolean Algebra which says it will state previous ideas more formally. This briefly acknowledges that the rigorous parts depend on the non-rigorous parts. Also the very important problem of carefully detailing how the mathematical objects discussed correspond to the real world things they are supposed to help us understand does not receive adequate attention.

Chapter 2 begins by saying we've now formulated our problem and the rest is just math. What I take from that is that the early assumptions won't be revisited but simply used as premises. So the rest is pointless if those early assumptions are mistaken, and Bayesian Epistemology cannot be proven in this way to anyone who doesn't grant the assumptions (such as a Popperian).

Moving on to Popper, Jaynes is ignorant of the topic and unscholarly. He writes:

<http://www-biba.inrialpes.fr/Jaynes/crefsy.pdf>

Custom Search

curi
 Karma Score
 12
 2

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
[Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

> Karl Popper is famous mostly through making a career out of the doctrine that theories may not be proved true, only false

This is pure fiction. Popper is a fallibilist and said (repeatedly) that theories cannot be proved false (or anything else).

It's important to criticize unscholarly books promoting myths about rival philosophers rather than addressing their actual arguments. That's a major flaw not just in a particular paragraph but in the author's way of thinking. It's especially relevant in this case since the author of the books tries to tell us about how to think.

Note that Yudkowsky made a similar unscholarly mistake, about the same rival philosopher, here:

<http://yudkowsky.net/rational/bayes>

> Previously, the most popular philosophy of science was probably Karl Popper's falsificationism - this is the old philosophy that the Bayesian revolution is currently dethroning. Karl Popper's idea that theories can be definitely falsified, but never definitely confirmed

Popper's philosophy is not falsificationism, it was never the most popular, and it is fallibilist: it says ideas cannot be definitely falsified. It's bad to make this kind of mistake about what a rival's basic claims are when claiming to be dethroning him. The correct method of dethroning a rival philosophy involves understanding what it does say and criticizing that.

If Bayesians wish to challenge Popper they should learn his ideas and address his arguments. For example he questioned the concept of positive support for ideas. Part of this argument involves asking the questions: "What is support?" (This is not asking for its essential nature or a perfect definition, just to explain clearly and precisely what the support idea actually says) and "What is the difference between "X supports Y" and "X is consistent with Y"?" If anyone has the answer, please tell me.

Article Navigation

Comments (226)

Tags: epistemology | popper | jaynes | Delete

Comments (226)

Sort By: New

Enter a comment here

Comment

Show help

Peterdjones 18 July 2011 12:03:38AM * 2 points [-]
If anyone can bear more of this, Poppers argument against induction using Bayes is being discussed [here](#)

Peterdjones 12 April 2011 08:23:30PM * 2 points [-]
""What is support?" (This is not asking for its essential nature or a perfect definition, just to explain clearly and precisely what the support idea actually says) and "What is the difference between "X supports Y" and "X is consistent with Y"?" If anyone has the answer, please tell me."
Bayesians appear to have answers to these questions. Moreovoer, far from wishing to refute Popper, they can actually incorporate a fomr of Popperianism.
"On the other hand, Popper's idea that there is only falsification and no such thing as confirmation turns out to be incorrect. Bayes' Theorem shows that falsification is very strong evidence compared to confirmation, but falsification is still probabilistic in nature; it is not governed by fundamentally different rules from confirmation, as Popper argued."
But of course Popper was a fallibilist as well as a falsificationist, so his falsifications aren't absolute and certain anyway. Bayes just brings out that where you don't have absolute falsification, you can't have absolute lack of positive support. Falsification of T has to support not-T. But the support gets spread thinly...

Peterdjones 12 April 2011 08:11:16PM 2 points [-]

LW 2.0 Open Beta Live
by Vaniver | 21v (33c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

European Community Weekend 2017
by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about "protecting" the epistemology
by AnnaSalamon | 73v (53c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)

Fact Posts: How and Why
by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement
by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- gwern (32)
ingres (30)
Lumifer (27)
turchin (23)
Habryka (22)
Yosarian2 (21)
IlyaShpitser (19)
SaidAchmiz (19)
gjm (19)
fowlertm (17)
SquirrelInHell (15)
Dagon (15)
Elo (14)
Stuart_Armstrong (13)
Gunnar_Zarncke (12)

RECENT KARMA AWARDS

Curi,

"Some first chapter assumptions are incorrect or unargued. It begins with an example with a policeman, and says his conclusion is not a logical deduction because the evidence is logically consistent with his conclusion being false."

Popper's epistemology doesn't explain that the conclusion of the argument has no validity, in the sense of being certainly false. In fact, it requires that the conclusion is not certainly false. No conjecture is certainly false.

Perhaps you meant he shows that the argument is invalid in the sense of being a non sequitur. (A non sequitur can still have a plausible or true conclusion). Of course it is not valid in the sense of traditional, necessitarian deduction. The whole point is that it is something different. And the argument that this non-traditional, plausibility based deduction works is just the informal observation that we use it all the time and it seems to work. What else could it be? If were valid by traditional deduction it would BE traditional deduction.

" Later when he gets into more mathematical stuff which doesn't (directly) rest on appeals to intuition, it does rest on the ideas he (supposedly) established early on with his appeals to intuition."

The Popperian argument against probabilistic reasoning is that it can't be shown how it works. If Jaynes maths shows how it works, that objection is removed.

"This is pure fiction. Popper is a fallibilist and said (repeatedly) that theories cannot be proved false (or anything else)."

Of course he has to believe in some FAPP refutation. or he ends up saying nothing at all.

PeterdJones 14 April 2011 02:27:20PM 3 points

[...]

"Science, philosophy and rational thought must all start from common sense". KRP, Objective Knowledge, p33.

Starting with common sense is exactly what Jaynes is doing. (Popper says that what is important is not to take common sense as irrefutable).

prase 07 April 2011 01:30:14PM 19 points

[...]

I have skimmed through the comments here and smelled a weak odour of a flame war. Well, the discussion is still rather civil and far from a flame war as understood on most internet forums, but it somehow doesn't fit well within what I am used to see here on LW.

The main problem I have is that you (i.e. curi) have repeatedly asserted that the Bayesians, including most of LW users, don't understand Popperianism and that Bayesianism is in fact worse, without properly explaining your position. It is entirely possible, even probable, that most people here don't actually get all subtleties of Popper's worldview. But then, a better strategy may be to first write a post which explains these subtleties and tells why they are important. On the other hand, you don't need to tell us explicitly "you are unscholarly and misinterpret Popper". If you actually explain what you ought to (and if you are right about the issue), people here will likely understand that they were previously wrong, and they will do it without feeling that you seek confrontation rather than truth - which I mildly have.

curi 07 April 2011 07:46:41PM 0 points

[...]

By "properly explaining my position" I'm not sure what you want. Properly understanding it takes reading, say, 20 books (plus asking questions about them as you go, and having critical discussions about them, and so on). If I summarize, lots of precision is lost. I have tried to summarize.

I can't write "a (one) post" that explains the subtleties of Popper. It took Popper a career and many books.

Bayesianism has a regress/foundations problem. Yudkowsky acknowledges that. Popperism doesn't. So Popperism is better in a pretty straightforward way.

On the other hand, you don't need to tell us explicitly "you are unscholarly and misinterpret Popper".

But they were propagating myths about Popper. They were unscholarly. They didn't know wtf they were talking about, not even the basics. Basically *all* of Popper's books contradict those myths. It's really not cool to *attribute*

positions to someone he never advocated. This mistake is easy to avoid by the method: *don't publish about people you haven't read.* Bad scholarship is a big deal, IMO.

Desrtopa 08 April 2011 09:34:47PM * 4 points

Bayesianism has a regress/foundations problem. Yudkowsky acknowledges that. Popperism doesn't. So Popperism is better in a pretty straightforward way.

Any system with axioms can be infinitely regressed or rendered circular if you demand that it justify the axioms. Critical Rationalism has axioms, and can be infinitely regressed.

You were upvoted in the beginning for pointing out gaps in scholarship and raising ideas not in common circulation here. You yourself, however, have demonstrated a clear lack of understanding of Bayesianism, and have attracted frustration with your *own* lack of scholarship and confused arguments, along with failure to provide good reasons for us to be *interested* in the prospect of doing this large amount of reading you insist is necessary to properly understand Popper. If doing this reading were worthwhile, we would expect you to be able to give a better demonstration of why.

curi 08 April 2011 09:44:47PM -5 points

Any system with axioms can be infinitely regressed or rendered circular if you demand that it justify the axioms. Critical Rationalism has axioms, and can be infinitely regressed.

You haven't understood the basic point that this only works *if you accept that ideas should be justify.*

If you reject the demand for justification -- as CR does -- then then regress is gone. Hence no regress in CR.

Peterdjones 15 April 2011 03:04:54PM 1 point

How do you know which forms of criticism are valid? Do you justify them, or attempt to criticise them? Either way looks regressive to me.

curi 15 April 2011 05:44:19PM * -4 points

Tell me how you found me on Less Wrong or I'm definitely not answering any of your questions.

Desrtopa 08 April 2011 09:58:13PM * 0 points

I reject this position as vacuous.

The position might be self consistent if you accept its premises, and one of its premises may be that you can introduce any idea without justification, but it's not *reality consistent*.

curi 08 April 2011 10:00:50PM -5 points

b/c?

Desrtopa 08 April 2011 10:19:35PM * 3 points

It's no less founded on unproven axioms, and it's no less arbitrary than induction, it just contains the tenet "we don't reject propositions for being arbitrary" and pats itself on the back for being self consistent. This doesn't do a better job helping people generate true information, it does a worse one.

[deleted] 09 April 2011 07:42:37AM -1 points

Being arbitrary is a criticism, so a critical rationalist can and does reject propositions for being

arbitrary. Rejecting the idea of justification does not mean accepting any old arbitrary thing. If your idea doesn't stand up to criticism, including criticism that it is just arbitrary, then it is gone.

prase 07 April 2011 08:51:35PM 2 points

I have tried to summarize.

I acknowledge that, although I would have preferred if you did that before you have written this post.

I can't write "a (one) post" that explains the subtitles of Popper. It took Popper a career and many books.

Could be five posts.

Even if such a defense can be sometimes valid, it is too often used to defend confused positions (think about theology) to be much credible.

curi 07 April 2011 08:52:52PM -1 points

It would need to be 500 posts.

But anyway, they are written and published. By Popper not me. They already exist and they don't need to be published on this particular website.

prase 07 April 2011 08:57:01PM 3 points

Following your advice expressed elsewhere, isn't the fact that the basics of Popperianism cannot be explained in five posts a valid criticism of Popperianism, which should be therefore rejected?

curi 07 April 2011 08:59:26PM 1 point

Why is that a criticism? What's wrong with that?

Also maybe it could be. But I don't know how.

And the *basics* could be explained quickly, to someone who didn't have a bunch of anti-Popperian biases, but people do have those b/c they are built into our culture. And without the details and precision then people complain about 1) not understanding how to do it, what it says 2) it not having enough precision and rigor

prase 07 April 2011 09:06:08PM * 2 points

Why is that a criticism?

Actually I don't know what constitutes a criticism in your book (since you never specified), but you have also said that there are no rules for criticism, so I suppose that it is a criticism. If not, then please say why it is *not* a criticism.

I am not going to engage in a discussion about my and your biases, since such debates rarely lead to an agreement.

curi 07 April 2011 09:11:10PM * 0 points

You can conjecture standards of criticism, or use the ones from your culture. If you find a problem with them, you can change them or conjecture different ones.

For many purposes I'm pretty happy with common sense notions of standards of criticism, which I think you understand, but which are hard to explain in words. If you have a *relevant* problem with the, you can say it.

[deleted] 07 April 2011 08:55:41PM 4 points [\[-\]](#)

One thing you could do is write a post highlighting a specific example where Bayes is wrong and Popper is right. A lot of people have asked for specific examples in this thread; if you could give a detailed discussion of one, that would move the discussion to more fertile ground.

curi 07 April 2011 08:57:01PM * 1 point [\[-\]](#)

Can you give me a link to a canonical essay on Bayesian *epistemology/philosophy*, and I'll pick from there?

Induction and justificationism are examples but I've been talking about them. I think you want something else. Not entirely sure what.

[deleted] 07 April 2011 09:04:46PM 1 point [\[-\]](#)

It's not at all canonical, but a paper that neatly summarizes Bayesian epistemology is "Bayesian Epistemology" by Stephan Hartmann and Jan Sprenger.

curi 07 April 2011 09:09:44PM 1 point [\[-\]](#)

Found it.

http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf

Will take a look in a bit.

[deleted] 07 April 2011 09:14:16PM 1 point [\[-\]](#)

Excellent, thanks.

curi 08 April 2011 09:18:03AM -7 points [\[-\]](#)

http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf

Bayesian epistemology therefore complements traditional epistemology; it does not replace it or aim at replacing it.

Since Popper refuted traditional epistemology (source: his books, and the failure of anyone to come up with any good criticisms of his main ideas), and Bayesian Epistemology retains it, then Bayesian Epistemology is refuted too. And discussing this issue can be done without mentioning probability, Bayes' theorem, or Solomonoff induction. Bringing those up cannot be a relevant defense since traditional epistemology, which doesn't use them, is retained.

Bayesian epistemology is, in the first place, a philosophical project, and that it is its ambition to further progress in philosophy.

Why are most Less Wrong people anti-philosophy then? There's so much instrumentalism, empiricism, reductionism and borderline postivism. Not much interest in philosophy.

Section 2 introduces the probability calculus and explains why degrees of belief obey the probability calculus. Section 3 applies the formal machinery to an analysis of the notion of evidence, and highlights potential application. Section 4 discusses Bayesian models of coherence and testimony, and section 5 ends this essay with a comparison of traditional epistemology and Bayesian epistemology.

Sections 2-4 are irrelevant. They are already assuming mistakes from traditional

epistemology. Moving on to 5, which is only one page.

Bayesian epistemology, on the other hand, draws much of its power from the mathematical machinery of probability theory. It starts with a mathematical intuition.

Advocating intuitionism is very silly.

traditional epistemology inspires Bayesian accounts.

So Bayesians should care about criticisms of traditional epistemology, and be willing to engage with them directly without even mentioning any Bayesian stuff.

Both Bayesian epistemology and traditional epistemology do not much consider empirical data. Both are based on intuitions,

That's not even close to what most Less Wrong people told me. They mostly are very focussed on empirical data.

This might be a problem as privilege is given to the philosopher's intuitions.

Might be? lol... What a hedge. They know it's a problem and equivocate.

non-philosophers may have different intuitions.

Also Popperian philosophers, and all other types that don't agree with you.

While it is debatable how serious these intuitions should be taken (maybe people are simply wrong!)

But not traditional philosophers, who have reliable intuitions? This is just plain silly.

It is therefore advisable that philosophers also keep on paying attention to other formal frameworks

But not informal frameworks, because your intuition says that formality is next to Godliness?

[continue this thread »](#)

Desrtopa 07 April 2011 02:49:52PM * 3 points



Upvoted and agreed. I feel at this point like further addressing the discussion on present terms would be simply irresponsible, more likely to become adversarial than productive. If curi wrote up such a post, it would hopefully give a meaningful place to continue from.

Edit: It seems that curi has [created such a post](#). I'm not entirely convinced that continuing the discussion is a good idea, but perhaps it's worth humoring the effort.

TheOtherDave 07 April 2011 02:24:22PM 2 points



For what it's worth, I have that feeling more than mildly and consequently stopped paying attention to the curi-exchange a while ago. Too much heat, not enough light.

I've been considering downvoting the whole thread on the grounds that I want less of it, but haven't yet, roughly on the grounds that I consider it irresponsible to do so without paying more careful attention to it and don't currently consider it worth paying more attention to.

David_Gerard 07 April 2011 12:15:23PM * 2 points



It has occurred to me before that the lack of a proper explanation on LessWrong of Bayesian *epistemology* (and not just saying "Here's Bayes' theorem and how it works, with a neat Java applet") is a serious lack. I've been

reduced to linking the Stanford Encyclopedia of Philosophy article, which is really not well written at all.

It is also clear from the comments on this post that people are talking about it without citable sources, and are downvoting as a mark of disagreement rather than anything else. This is bad as it directly discourages thought or engagement on the topic from those trying to disagree in good faith, as curi is here.

Is there a decent explanation of Bayesian epistemology per se (not the theorem, the epistemology) that doesn't start by talking about Popper or something else, that the Bayesian epistemology advocates here could link to? This would lead to a much more productive discussion, as everyone might at least start on approximately the same page.

benelliott 07 April 2011 12:40:43PM * 0 points

I don't know if these are what you're looking for but:

[Probability Theory: The Logic of Science](#) by Jaynes, spends its first chapter explaining why we need a 'calculus of plausibility' and what such a calculus should hope to achieve. The rest of the book is mostly about setting it up and showing what it can do. (The link does not contain the whole book, only the first few chapters, you may need to buy or borrow it to get the rest).

[Yudkowsky's Technical explanation](#), which assumes the reader is already familiar with the theorem, explains some of its implications for scientific thinking in general.

David_Gerard 07 April 2011 01:29:13PM 1 point

See [here](#) for what I see the absence of. There's a hole that needs filling here.

benelliott 07 April 2011 06:29:42AM * 5 points

I gave a description of how a Bayesian sees the difference between "X supports Y" and "X is consistent with Y" in our previous discussion. I don't know if you saw it, you haven't responded to it and you aren't acting like you accepted it so I'll give it again here:

"X is consistent with Y" is not really a Bayesian way of putting things, I can see two ways of interpreting it. One is as $P(X \& Y) > 0$, meaning it is at least theoretically possible that both X and Y are true. The other is that $P(X|Y)$ is reasonably large, i.e. that X is plausible if we assume Y.

"X supports Y" means $P(Y|X) > P(Y)$, X supports Y if and only if Y becomes more plausible when we learn of X. Bayes tells us that this is equivalent to $P(X|Y) > P(X)$, i.e. if Y would suggest that X is more likely that we might think otherwise then X is support of Y.

Suppose we make X the statement "the first swan I see today is white" and Y the statement "all swans are white". $P(X|Y)$ is very close to 1, $P(X|\sim Y)$ is less than 1 so $P(X|Y) > P(X)$, so seeing a white swan offers support for the view that all swans are white. Very, very weak support, but support nonetheless.

For a Popperian definition, you guys are allowed to criticise something right? In that case could we say that support for a proposition is logically equivalent to a criticism of its negation?

The whole 'there is no positive support' thing seems like an overreaction to the whole Cartesian 'I can prove ideas with certainty thing'. I agree that certain support is a flawed concept, but you seem to be throwing the baby out with the bathwater by saying uncertain support is guilty by association and should be rejected as well.

Also, I'm a little incredulous here, do you really reject the policeman's syllogism? Would you say he is wrong to chase the man down the road? If you encountered such a person, would you genuinely treat them as you would treat anyone else?

curi 07 April 2011 07:07:21AM * 0 points

I missed your comment. I found it now. I will reply there.

http://lesswrong.com/lw/3ox/bayesianism_versus_critical_rationalism/3uld?context=1#3uld

could we say that support for a proposition is logically equivalent to a criticism of its negation?

No. The negation of a universal theory is not universal, and the negation of an explanatory theory is not explanatory. So, the interesting theories would still be criticism only, and the uninteresting ones (e.g. "there is a cat") support only. And the meaning of "support" is rather circumscribed there.

If you want to say theories of the type "the following explanation isn't true:" get "supported" it doesn't contribute anything useful to epistemology. the support idea, as it is normally conceived, is still wrong, and this rescues none of the substance.

The other issue is that criticism isn't the same kind of thing as support. It's not in the same category of concept.

Yes I really reject the policeman's syllogism. In the sense of: I don't think the argument in the book is any good. There are other arguments which are OK for reaching the conclusion (but which rely on things the book left unstated, e.g. background knowledge and context. Without adding anything at all, no cultural biases or assumptions or hidden claims, and even doing our best to not use the biases and assumptions built into the English language, then no there isn't any way to guess what's more likely).

Peterdjones 15 April 2011 03:08:22PM 1 point [-]

If the Policeman's argument is only valid in the light of background assumptions, why would they need to be stated? Surely we would only need to make the same tacit assumptions to agree with the conclusions. Everyday reasoning differs from formal logic in various ways, and mainly because it takes short cuts. I don't think that invalidates it.

JGWeissman 07 April 2011 06:19:50AM 4 points [-]

A huge strength of Bayesian epistemology is that it tells me how to [program computers to form accurate beliefs](#). Has Popperian epistemology guided the development of any computer program as awesome as Gmail's spam filter?

curi 07 April 2011 06:59:03AM -2 points [-]

Bayesian epistemology didn't do that. Bayes' theorem did. See the difference?

JGWeissman 07 April 2011 04:40:10PM 3 points [-]

Bayesian epistemology didn't do that. Bayes' theorem did.

Bayes' theorem is part of probability theory. Bayesian epistemology essentially says to take probability theory seriously as a normative description of degrees of belief.

If you don't buy that and really want to split the hair, then I am willing to modify my question to: Has the math behind Popperian epistemology guided the development of any computer program as awesome as Gmail's spam filter? (Is there math behind Popperian epistemology?)

curi 07 April 2011 05:59:18PM -1 points [-]

gmail's spam filter does not have degrees of belief or belief.

It has things which you could call by those words if you really wanted to. But it wouldn't make them into the same things those words mean when referring to people.

JGWeissman 07 April 2011 06:14:51PM 4 points [-]

But it wouldn't make them into the same things those words mean when referring to people.

I want the program to find the correct belief, and then take good actions based on that correct belief. I don't care if lacks the conscious experience of believing.

You are [disputing definitions](#) and ignoring my actual question. Your next reply should answer the question, or admit that you do not know of an answer.

Alicorn 07 April 2011 06:12:49PM 4 points 

gmail's spam filter does not have degrees of belief or belief.

It has things which you could call by those words if you really wanted to. But it wouldn't make them into the same things those words mean when referring to people.

Augh, this reminded me of a quote that I can't seem to find based on my tentative memory of its component words... it was something to the effect that we anthropomorphize computers and talk about them "knowing" things or "communicating" with each other, and some people think that's wrong and they don't really do those things, and the quote-ee was of the opinion that computers were clarifying what we meant by those concepts all along. Anybody know what I'm talking about?

curi 07 April 2011 06:37:10PM 1 point 

To be clear, I think computers *can* do those things and AIs *will*, and *that* will help clarify the concepts a lot.

But I don't think that microsoft word does it. Nor any game "AI" today. Nor gmail's spam filter which just does mindlessly math.

jimrandomh 07 April 2011 01:02:59AM 12 points 

The assumptions behind Cox's theorem are:

1. Representation of degrees of plausibility by real numbers
2. Qualitative correspondence with common sense
3. Consistency

Would you please clearly state which of these you disagree with, and why? And if you disagree with (1), is it because you don't think degrees of plausibility should be represented, or because you think they should be represented by something other than real numbers, and if so, then what? (Please do not give an answer that cannot be defined precisely by mapping it to a mathematical set. And please do not suggest a representation that is obviously inadequate, such as booleans.)

curi 07 April 2011 03:00:06AM 1 point 

Could you explain what you're talking about a bit more? For example you state "consistency" as an assumption. What are you assuming is (should be?) consistent with what?

Larks 08 April 2011 01:10:11AM 3 points 

If only [Jaynes](#) had clearly listed them on page 114!

JoshuaZ 07 April 2011 03:25:19AM 11 points 

You may have valid points to make but it might help in getting people to listen to you if you don't exhibit apparent double standards. In particular, your main criticism seems to be that people aren't reading Popper's texts and related texts enough. Yet, at the same time, you are apparently unaware of the basic philosophical arguments for Bayesianism. This doesn't reduce the validity of anything you have to say but as an issue of trying to get people to listen, it isn't going to work well with fallible humans.

curi 07 April 2011 03:33:31AM -6 points 

It's fine if most people haven't read Popper. But they should be able to point to some Bayesian somewhere who did, or they should know at least one good argument against a major Popperian idea. or they should be interested and ask more about him instead of posting incorrect arguments about why his basic claims are false.

I do know, offhand, several arguments against Bayesian epistemology (e.g. it's inability to create moral knowledge, and i know many arguments against induction, each decisive). And anyway I came here to learn more about it. One particular thing I would be interested in is a Bayesian criticism of Popper. Are there any? By contrast (maybe), Popper did criticize Bayesian epistemology in LScD and elsewhere. And I am familiar with those criticisms.

Learning enough Bayesian stuff to sound like a Bayesian so people want to listen to me more sounds to me like more trouble than it's worth, no offense. I'm perfectly willing to read more things when I make a mistake and there is a specific thing which explains the issue. I have been reading various things people refer me to. If you wanted me to study Bayesian stuff for a month before speaking, well, I'd get bored because I would see flaws and then see them repeated, and then read arguments which depend on them. I did read the whole HP fic if that helps.

One thing that interests me, which I posted about in the initial post, is how unscholarly some Bayesian scholars are. Can anyone correct that? Are there any with higher scholarly standards? I would like there to be. I don't want to just read stuff until I happen to find something good, I want to be pointed to something considerably better than the unscholarly stuff I criticized. I don't know where to find that.

JoshuaZ 07 April 2011 04:55:06AM * 17 points



It's fine if most people haven't read Popper. But they should be able to point to some Bayesian somewhere who did, or they should know at least one good argument against a major Popperian idea. or they should be interested and ask more about him instead of posting incorrect arguments about why his basic claims are false.

Really? Much of that seems questionable. There are many different ideas out there and practically speaking, there are too many ideas out there for people to have to deal with every single one. Sure, making incorrect arguments is bad. And making arguments against strawmen is very bad. But people don't have time actually research every single idea out there or even know which one's to look at. Now, I think that Popper is important enough and has relevant enough points that he should be on the short list of philosophers that people can grapple with at least to some limited extent. But frankly, speaking as someone who is convinced of that point, you are making a very weak case for it.

I do know, offhand, several arguments against Bayesian epistemology (e.g. it's inability to create moral knowledge, and i know many arguments against induction, each decisive).

This paragraph seems to reflect a general problem you are having here in making assertions without providing any information other than vague claims of existence. I am for example aware of a large variety of arguments against induction (the consistency of anti-induction frameworks seems to be a major argument) but calling them "decisive" is a very strong claim, and isn't terribly relevant in so far as Bayesianism is not an inductive system in many senses of the term.

You've also referred to before to this claim that Popperian system can lead to moral knowledge and that's a claim I'd be very curious to hear expanded with a short summary of how that works. Generally when I see a claim that an epistemological system can create moral knowledge my initial guess is that someone has managed to bury the naturalistic fallacy somewhere or has managed to smuggle in additional moral premises that aren't really part of the epistemology. I'd be pleasantly surprised to see something that didn't function that way.

One particular thing I would be interested in is a Bayesian criticism of Popper. Are there any? By contrast (maybe), Popper did criticize Bayesian epistemology in LScD and elsewhere.

I haven't read it myself but I've been told that Earman's "Bayes or Bust" deals with a lot of the philosophical criticisms of Bayesianism as well as giving a lot of useful references. It should do a decent job in regards to the scholarly concerns.

As to Popper's criticism of Bayesianism, the discussion of it in LScD is quite small, which is understandable in that Bayesianism was not nearly as developed in that time as it is now. (You may incidentally be engaging in a classical philosophical fallacy here in focusing on a specific philosopher's

personal work rather than the general framework of ideas that followed from it. There's a lot of criticism of Bayesianism that is not in Popper that is potentially strong. Not everything is about Popper.)

Learning enough Bayesian stuff to sound like a Bayesian so people want to listen to me more sounds to me like more trouble than it's worth, no offense.

As a non-Bayesian, offense taken. You can't expect to go to a room full of people with a specific set of viewpoints offer a contrary view, act like the onus is on them to translate into your notation and terminology, and then be shocked when they don't listen to you. Moreover, knowing the basics of Cox's theorem is not asking you to "sound like a Bayesian" anyhow.

If you wanted me to study Bayesian stuff for a month before speaking, well, I'd get bored because I would see flaws and then see them repeated, and then read arguments which depend on them. I did read the whole HP fic if that helps.

What? I don't know how to respond to that. I'm not sure an exclamation exists in standard English to express my response to that last sentence. I'm thinking of saying "By every deity in the full Tegmark ensemble" but maybe I should wait for a better time to use it. You are repeatedly complaining about people not knowing much about Popper while your baseline for Bayesianism is that you've *read an incomplete Harry Potter fanfic*? This fanfic hasn't even addressed Bayesianism other than in passing. This seems akin to someone thinking they understand rocketry because they've watched "Apollo 13".

endoself 29 May 2011 09:32:12AM 2 points

By every deity in the full Tegmark ensemble

Can I steal this?

JoshuaZ 29 May 2011 06:57:37PM 0 points

Yes, by all means feel free.

curi 07 April 2011 09:21:30AM 1 point

Really? Much of that seems questionable. There are many different ideas out there and practically speaking, there are too many ideas out there for people to have to deal with every single one.

The number of major ideas in epistemology is not very large. After Aristotle, there wasn't very much innovation for a long time. It's a small enough field you can actually trace ideas all the way back to the start of written history. Any professional can look at everything important. Some Bayesian should have. Maybe some did, but I haven't seen anything of decent quality.

You've also referred to before to this claim that Popperian system can lead to moral knowledge and that's a claim I'd be very curious to hear expanded with a short summary of how that works.

It works exactly identically to how Popperian epistemology creates any other kind of knowledge. There's nothing special for morality.

Knowledge is created by an evolutionary process involving *conjecture* and *refutation*. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

You may incidentally be engaging in a classical philosophical fallacy here in focusing on a specific philosopher's personal work rather than the general framework of ideas that followed from it.

I have a lot of familiarity with the other Popperians. But Popper and Deutsch are by far the best. There isn't really anything non-Popperian that draws on Popper much. Everyone who has understood Popper is a Popperian, IMO. If you disagree, do tell.

As to Popper's criticism of Bayesianism, the discussion of it in LScD is quite small

Small is not a criticism; substance matters not length. Do you have a criticism of his arguments in LScD or not? Also he dealt with it elsewhere, as I stated.

JoshuaZ 07 April 2011 03:40:44PM * 2 points

The number of major ideas in epistemology is not very large. After Aristotle, there wasn't very much innovation for a long time. It's a small enough field you can actually trace ideas all the way back to the start of written history. Any professional can look at everything important. Some Bayesian should have. Maybe some did, but I haven't seen anything of decent quality.

As to a professional, I already referred you to Earman. Incidentally, you seem to be narrowing the claim somewhat. Note that I didn't say that the set of major ideas in epistemology isn't small, I referred to the much larger class of philosophical ideas (although I can see how that might not be clear from my wording). And the set is indeed very large. However, I think that your claim about "after Aristotle" is both wrong and misleading. There's a lot of what thought about epistemological issues in both the Islamic and Christian worlds during the Middle Ages. Now, you might argue that that's not helpful or relevant since it gets tangled up in theology and involves bad assumptions. But that's not to say that material doesn't exist. And that's before we get to non-Western stuff (which admittedly I don't know much about at all).

(I agree when you restrict to professionals, and have already recommended Earman to you.)

It works exactly identically to how Popperian epistemology creates any other kind of knowledge. There's nothing special for morality.

Knowledge is created by an evolutionary process involving conjecture and refutation. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

This is a deeply puzzling set of claims. First of all, a major point of his epistemological system is falsifiability based on data (at least as I understand it from LScD). How that would at all interact with moral issues is unclear to me. Indeed, the semi-canonical example of a non-falsifiable claim in the Popperian sense is Marxism, a set of ideas that has a large set of attached moral claims.

I also don't see how this works given that moral claims can always be criticized by the essential sociopathic argument "I don't care. Why should you?" Obviously, that line of thinking can be/should be expanded. To use your earlier example, how would you discuss "murder is wrong" in a Popperian framework? I would suggest that this isn't going to be any different than simply discussing moral ideas based on shared intuitions with particular attention to the edge cases. You're welcome to expand on these claims, but right now, nothing you've said in this regard is remotely convincing or even helpful since it amounts to just saying "well, do the same thing."

I have a lot of familiarity with the other Popperians. But Popper and Deutsch are by far the best. There isn't really anything non-Popperian that draws on Popper much. Everyone who has understood Popper is a Popperian, IMO. If you disagree, do tell.

I'm going to be obnoxious and quote a friend of mine "Everyone who understands Christianity is a Christian." I don't have any deep examples of other individuals although I would tentatively say that I understood Popper's views in Logic of Scientific Discovery just fine.

Do you have a criticism of his arguments in LScD or not?

Sure. The most obvious one is when he is discussing the law of large numbers and frequentist v. Bayesian interpretations (incidentally to understand those passages it is helpful to note that he uses the term "subjective" to describe Bayesians rather than Bayesian which is consistent with the language of the time, but in modern terminology has a very different meaning (used to distinguish between subject and objective Bayesians)). In that section he argues that (I don't have the page number unfortunately since I'm using my Kindle edition. I have a hard copy somewhere but I don't know where) that "it must be inadmissible to give *after* the deduction of Bernoulli's theorem a meaning to p different from the one which was given to it before the deduction." This is, simply put, wrong. Mathematicians all the time prove something in one framework and then interpret it in another framework. You just need to show that all the properties of the relevant frameworks overlap in sufficiently non-pathological cases. If someone wrote this as a complaint about say using the complex exponential to understand the symmetries of the Euclidean plane, we'd

immediately see this as a bad claim. There's an associated issue in this section which also turns up but it is more subtle; Popper doesn't appreciate what you can do with measure theory and L_p spaces and related ideas to move back and forth between different notions of probability and different metrics on spaces. That's ok, it was a very new idea when he wrote LScD (although the connections were to some extent definitely there). But it does render a lot of what he says simply irrelevant or outright wrong.

curi 07 April 2011 06:58:34PM * 0 points

[...]

As to a professional, I already referred you to Earman.

Which you stated you had not read. I have rather low standards for recommendations of things to read, but "I never read it myself" isn't good enough.

I don't agree with "restrict to professionals". How is it to be determined who is a professional? I don't want to set up arbitrary, authoritative criteria for dismissing ideas based on their source.

First of all, a major point of his epistemological system is falsifiability based on data (at least as I understand it from LScD).

That is a major point *for scientific research* where the problem "how do we use evidence?" is important. And the answer is "criticisms can refer to evidence". Note by "science" here I mean any empirical field. What do you do in non-scientific fields? You simply make criticisms that don't refer to evidence. Same method, just missing one type of criticism which is rather useful in science but not fundamental to the methodology.

Indeed, the semi-canonical example of a non-falsifiable claim in the Popperian sense is Marxism, a set of ideas that has a large set of attached moral claims.

It is not *empirically falsifiable*. It is criticizable. For example Popper criticized Marx in *The Open Society and its Enemies*..

I also don't see how this works given that moral claims can always be criticized by the essential sociopathic argument "I don't care. Why should you?"

Any argument which works against everything fails at the task of differentiating better and worse ideas. So it is a bad argument. So we can reject it and all other things in that category, by this criticism.

To use your earlier example, how would you discuss "murder is wrong" in a Popperian framework?

The short answer is: since we don't care to have justified foundations, you can discuss it any way you like. You can say it's bad because it hurts people. You can say it's good because it prevents overpopulation. You can say it's bad because it's mean. These kinds of normal arguments, made by normal people, are not deemed automatically invalid and ignored. Many of them are indeed mistakes. But some make good points.

For more on morality, please join this discussion:

http://lesswrong.com/lw/552/reply_to_benelliott_about_popper_issues/3uv7

I would tentatively say that I understood Popper's views in Logic of Scientific Discovery just fine.

He has like 20 books. There's way more to it. When one reads a lot of them, a whole worldview comes across that is very hard to understand from just a couple books. And I wasn't trying to *argue* with that statement, I was just commenting. I mentioned it because of a comment to do with whether I had studied results of non-Popperians using Popperian ideas.

"it must be inadmissible to give after the deduction of Bernoulli's theorem a meaning to p different from the one which was given to it before the deduction." This is, simply put, wrong.

Are you really telling me that you can prove something, then take the conclusion, redefine a term, and work with that, and consider it still proven? You could only do that *if* you created a

second proof that the change doesn't break anything, you can't just do it. I'm not sure you took what Popper was saying literally enough; I don't think your examples later actually do what he criticized. Changing the meaning of a term in a conclusion statement, and considering a conclusion from a different perspective, are different.

Popper doesn't appreciate what you can do with measure theory and L_p spaces

Would you understand if I said this has no relevance at all to 99.99% of Popper's philosophy? Note that his later books generally have considerably less mention of math or logic.

JoshuaZ 08 April 2011 02:43:45AM * 3 points

Which you stated you had not read. I have rather low standards for recommendations of things to read, but "I never read it myself" isn't good enough.

Earman is a philosopher and the book has gotten positive reviews from other philosophers. I don't know what else to say in that regard.

I don't agree with "restrict to professionals". How is it to be determined who is a professional? I don't want to set up arbitrary, authoritative criteria for dismissing ideas based on their source.

Hrrm? You mentioned professionals first. I'm not sure why you are now objecting to the use of professionals as a relevant category.

That is a major point for scientific research where the problem "how do we use evidence?" is important. And the answer is "criticisms can refer to evidence". Note by "science" here I mean any empirical field. What do you do in non-scientific fields? You simply make criticisms that don't refer to evidence. Same method, just missing one type of criticism which is rather useful in science but not fundamental to the methodology

I'm not at all convinced that this is what Popper intended (but again I've only read LScD) but if this is accurate then Popper isn't just wrong in an interesting way but is just wrong. Does one mean for example to claim that pure mathematics works off of criticism? I'm a mathematician. We don't do this. Moreover, it isn't clear what it would even mean for us to try to do this as our primary method of inquiry. Are we supposed to spend all our time going through pre-existing proofs trying to find holes in them?

He has like 20 books. There's way more to it. When one reads a lot of them, a whole worldview comes across that is very hard to understand from just a couple books.

Yes, and I'm quite sure that I get much more of a worldview if I read all of Hegel rather than just some of it. That doesn't mean I need to read all of it. Similar remarks would apply to Aquinas or more starkly the New Testament. Do you need to read all of the New Testament to decide that Christianity is bunk? Do you need to read the entire Talmud to decide that Judaism is incorrect? But you get a whole worldview that you don't obtain from just reading the major texts.

The short answer is: since we don't care to have justified foundations, you can discuss it any way you like. You can say it's bad because it hurts people. You can say it's good because it prevents overpopulation. You can say it's bad because it's mean. These kinds of normal arguments, made by normal people, are not deemed automatically invalid and ignored. Many of them are indeed mistakes. But some make good points

Right, and then we just the criticism "why bother" or "and how does that maximize the number of paperclip in the universe?" Or one can say "mean" "good" bad" are all hideously ill-defined. In any event, does it not bother you that you are essentially claiming that your moral discussion with your great epistemological system looks just like a discussion about morality by a bunch of random individuals? There's nothing in the above that uses your epistemology in any substantial way.

Are you really telling me that you can prove something, then take the conclusion, redefine a term, and work with that, and consider it still proven? You could only do that if you created a second proof that the change doesn't break anything, you can't just do it.

Right! And conveniently in the case Popper cares about you can prove that.

Popper doesn't appreciate what you can do with measure theory and

Would you understand if I said this has no relevance at all to 99.99% of Popper's philosophy? Note that his later books generally have considerably less mention of math or logic.

Do you mean understand or do you mean care? I don't understand why you are making this statement given that my remark was addressing the question you asked of whether I had specific problems with Popper's handling of Bayesianism in LScD. This is a specific problem there.

AlephNeil 08 April 2011 06:12:28PM * 1 point

Does one mean for example to claim that pure mathematics works off of criticism? I'm a mathematician. We don't do this.

I don't know what Popper himself would say, but one of his more insightful followers, namely Lakatos, argues for [exactly that position](#).

I read Proofs and Refutations too many years ago to say anything precise about it. I remember finding it interesting but also frustrating. Lakatos seems determined to ignore/deny/downplay the fact of mathematical practice that we only call something a 'theorem' when we've got a proof, and we only call something a 'proof' when it's logically watertight in such a way that no 'refutations' are possible. Still, it's well-researched (in its use of a historical case-study) and he comes up with some decent ideas along the way (e.g. about "monster barring" and "proof-oriented definitions".)

[continue this thread »](#)

curi 08 April 2011 09:47:58AM -7 points

Does one mean for example to claim that pure mathematics works off of criticism?

yes

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

[continue this thread »](#)

Randaly 07 April 2011 01:39:46PM 2 points

Err, Bayesian probability doesn't have anything special for morality either. People on LW tend to be moral non-realists, ie people who deny that there is objective moral knowledge, if that's what you're talking about (not sure- sorry!), but that's completely orthogonal to this discussion: there's nothing in Bayesianism that leads inevitably to non-realism. (Also, I'm not convinced that moral realism is right, so saying "Bayesianism leads to moral non-realism" isn't a very effective argument.)

curi 07 April 2011 07:08:51PM 0 points

Bayesian epistemology *doesn't create moral knowledge* because it only functions when fed in *observation data* (or assumptions). I get a lot of conflicting statements here, but some people tell me they *only care about prediction*, they are *instrumentalists*, and that is what Bayes stuff is for, and they don't regard it as a bad thing that it doesn't address morality at all.

Now what you have in mind, I think, is that if you make a ton of assumptions you could *then* talk about morality using Bayes. Popperism doesn't require a bunch of arbitrary starting

assumptions to create moral knowledge, it just can deal with it directly.

If I'm wrong, explain how you can deal with figuring out, e.g., what are good moral values to have (without assuming a utility function or something).

Randy 09 April 2011 04:22:53AM 2 points

As I tried to say (and probably explained really poorly- sorry!), the LW consensus is that morality is not objective. Therefore, the idea of figuring out what good moral values would be is, according to moral non-realism, impossible: any decision about what a good moral value is must rely on your pre-existing values, if an objective morality is not out there to be discovered. Using this as a criticism of Bayesianism is sorta like criticizing thermodynamics because it claims it's impossible to exactly specify the position and velocity of each particle: not only is the criticism unrelated to the subject matter, but satisfying it would require the theory to do something that is to the best of our knowledge incorrect.

[deleted] 07 April 2011 11:19:12AM * 2 points

Knowledge is created by an evolutionary process involving conjecture and refutation. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

I'm inclined to take this formula seriously, but I'd like to start by applying it to innate knowledge, knowledge we are born with, because here we are definitely talking about an evolutionary process involving mutation and natural selection. Some mutations add what amounts to a new innate conjecture (hypothesis, belief) into the cognitive architecture of the creature.

However, what occurs at this point is not that a creature with a false innate conjecture is eliminated. The creature isn't being tested purely against reality in isolation. It's being tested against other members of its species. The creature with the least-false, or least-perilously-false conjecture will tend to do better than the competitors. The competition for survival amounts to a competition between rival conjectures. The truest, or most-usefully-true, or least-wrong, or least-dangerously-wrong innate belief will tend to outdo its competitors and ultimately spread through the species. (With the odd usefully-wrong belief surviving.)

The occasional appearance of new innate conjectures resembles the conjecture part of Popperian conjecture and refutation. However, the contest between rival innate conjectures that occurs as the members of the species struggle against each other for survival seems less Popperian than Bayesian.

The relative success of the members of the species who carry the more successful hypothesis vaguely resembles Bayesian updating, because the winners increase their relative numbers and the losers decrease their relative numbers, which resembles the shift in the probabilities assigned to rival hypotheses that occurs in Bayesian updating. Consider the following substitutions applied to Bayes' formula:

$$P(H|E) = P(E|H)P(H) / P(E)$$

- $P(H|E)$ is the new proportion (i.e. in the next generation) of the species carrying the hypothesis H , given that event E occurred (E is "everything that happened to the generation")
- $P(E|H)$ is the degree to which H predicts and thus prepares the individual to handle E (measured in expected number of offspring given E)
- $P(H)$ is the old proportion (i.e. in the previous generation) of the species carrying H
- $P(E)$ is the degree to which the average member of the species predicts and thus is prepared to handle E (measured in expected number of offspring given E)

With these assignments, what the equation means is:

The new proportion of the species with H is equal to the old proportion of the species with H , times the expected number of offspring of members with H , divided by the expected number of offspring of the average member of the species.

One difference between this process and Bayesian updating is that this process allows the occasional introduction of new hypotheses over time, with what amounts to a modest but not vanishing initial prior.

curi 07 April 2011 07:19:31PM 0 points



I'm not sure if we're interested in the same stuff. But taking up one topic:

I think you regard innate/genetic ideas as important. I do not. Because people are universal knowledge creators, and can change any idea they start with, it doesn't matter very much.

The reason people are so biased is not in their genes but their memes.

There are two major replication strategies that memes use.

- 1) a meme can be useful and rational. it spreads because of its value
- 2) a meme can sabotage its holders creativity to prevent him from criticizing it, and to take away his choice not to spread it

The second type dominated all cultures on Earth for a long time. The transition to the first type is incomplete.

More details one memes and universality can be found in *The Beginning of Infinity* by David Deutsch

[deleted] 07 April 2011 10:10:10PM 2 points



I think you regard innate/genetic ideas as important. I do not. Because people are universal knowledge creators, and can change any idea they start with, it doesn't matter very much.

You misunderstand. I bring it up as a model of learning, and my choice was based on your own remarks. You said that knowledge is created by an evolutionary process. That way of putting it suggests an analogy with Darwin's theory of evolution as proceeding by random variation and natural selection. And indeed there is an analogy between Popper's conjectures and refutations and variation and natural selection, and it is this: a conjecture is something like variation (mutation), and refutation is something like natural selection.

However, what I found was that the closer I looked at knowledge which is actually acquired through natural selection - what we might call innate knowledge or instinctive knowledge - the more the process of acquisition resembled Bayesian updating rather than Popperian conjecture and refutation. I explained why.

In Bayesian updating, there are competing hypotheses, and the one for which actual events are less of a surprise (i.e., the hypothesis H_i for which $P(E|H_i)$ is higher) is strengthened relative to the one for which events are more of a surprise. I find a parallel to this in competition among alleles under natural selection, which I described.

Essential to Bayesian updating is the coexistence of competing hypotheses, and essential to natural selection is the coexistence of competing variants in a species. In contrast, Popper talks about conjecture and refutation, which is a more lonely process that need not involve more than one conjecture and a set of observations which have the potential to falsify it. Popper talks about improving the conjecture in response to refutation, but this process more resembles Lamarckian evolution than Darwinian evolution, because in Lamarckian evolution the individuals improve themselves in response to environmental challenges, much as Popper would have us improve our conjectures in response to observational challenges. Also, in Lamarckian evolution, as in the Popperian process of conjecture and refutation, competing variants (compare: competing hypotheses) do not play an *essential* role (though I'm sure they could be introduced). Rather, the picture is of a *single* animal (compare: a single hypothesis) facing existential environmental challenges (compare: facing the potential for falsification) improving itself in response (which improvement is passed to offspring).

The Popperian process of conjecture, refutation, and improvement of the conjecture, can as

it happens be understood from a Bayesian standpoint. It does implement Bayesian updating in a certain way. Specifically, when a particular conjecture is refuted and the scientist modifies the conjecture - at that point, there are two competing hypotheses. So at that point, the process of choosing between these two competing hypotheses can be characterized as Bayesian updating. The less successful hypothesis is weakened, and the more successful hypothesis is strengthened.

In short, if you want to take seriously the analogy that does exist between evolution through natural selection and knowledge acquisition of whatever type, then you may want to take a closer look at Bayesian updating as conforming more closely to the Darwinian model.

curi 07 April 2011 10:12:27PM 1 point

In short, if you want to take seriously the analogy

I wasn't talking about an analogy.

Evolution is a theory which applies to any type of replicator. Not by analogy by literally applies.

Make sense so far?

[continue this thread »](#)

jimrandomh 07 April 2011 03:18:22AM * 4 points

Cox's theorem is a proof of Bayes rule, from the conditions above. "Consistency" in this context means (Jaynes 19): If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result; we always take into account all of the evidence we have relevant to a question; and we always represent equivalent states of knowledge by equivalent plausibility assignments. By "reason in more than one way", we specifically mean adding the same pieces of evidence in different orders.

(Edit: It's page 114 in the PDF you linked. That seems to be the same text as my printed copy, but with the numbering starting in a different place for some reason.)

[deleted] 07 April 2011 03:43:28AM -5 points

Assigning degrees of plausibility to theories is an attempt to justify them. Cox's theorem just assumes you can do this. Popper argued that justification, including probabilistic justification, is impossible. How does just assuming something that Popper refuted show anything?

timtyler 07 April 2011 12:36:50PM * 0 points

Assigning degrees of plausibility to theories is an attempt to justify them. Cox's theorem just assumes you can do this.

Do you think that [grue and bleen](#) are as plausible as blue and green? Would you like to bet?

benelliott 07 April 2011 12:44:50PM 1 point

Nitpicking here, grue and bleen aren't statements and thus can't be assigned probabilities. "This object is grue" and "this object is bleen" are statements.

timtyler 07 April 2011 01:21:50PM * 0 points

Yes, I left making up more specific examples as an exercise for the reader.

[deleted] 07 April 2011 11:55:24PM 0 points

Assuming that the object in question is an emerald, then grue is in conflict with our best explanations about emeralds whereas there are no known problems with the idea that the emerald is green. So I go with green, but not because I have assigned degrees of plausibility but because I see no problem with green.

benelliott 07 April 2011 06:16:17AM 5 points

One argument for plausibility would be this.

At some point you may be called on to base a decision on whether something is true or false. The simplest of these decisions can be reduced to betting for or against something, and you cannot always choose not to bet. There must be some odds at which you switch from betting on falsity to betting on truth, and those can be taken to demonstrate your plausibility assignment.

[deleted] 07 April 2011 07:41:01AM 0 points

How does betting on the truth of a universal theory work? I can't see a bookie ever paying out on that, although it would be good business to get punters to take such bets.

timtyler 07 April 2011 12:31:50PM 2 points

So: just bet on things the theory predicts instead.

[deleted] 07 April 2011 11:25:40PM 0 points

Having the prediction turn out doesn't make the theory true or more likely, it is just consistent evidence. There are an infinitude of other theories that the same evidence is consistent with.

timtyler 08 April 2011 12:16:34PM 4 points

To give a simple example, consider flipping a coin. You observe HHH. Is this a fair coin? or a double-headed one? or a biased coin? Different theories describe these situations, and you could be asked to bet on them. Imagine you then further observe HHHH - making a total of HHHHHHH. This makes your estimate of the chances of the "double-headed coin" hypothesis go up. Other hypotheses may increase in probability too - but we are not troubled by there being an infinity of them, since we give extra weight to the simpler ones, using Occam's razor.

benelliott 07 April 2011 11:16:26AM * 4 points

The usual way on Less Wrong is to bring in Omega, the all powerful all knowing entity who spends his free time playing games with us mortals, and for some reason most of his games illustrate some point of probability or decision theory. With Omega acting as the bookie you can be forced to assign a probability to any meaningful statement. Some people just respond to such scenarios by asserting that Omega is impossible, I don't know if you're one of those people but I'll try a different approach anyway.

Imagine that in 2050 the physicists have narrowed down all their candidates for a Theory of Everything to just two possibilities, creatively named X-theory and Y-theory.

An engineer who is a passionate supporter of X-theory has designed and built a new power plant. If X-theory is correct, his power plant will produce a limitless supply of free energy and bring humanity into a post-scarcity era.

However, a number of physicists have had a look at his designs, and have shown that if Y-theory is correct his power plant will create a black hole and wipe out humanity as soon as it is turned on. Somehow, it has ended up being your decision whether or not it goes on.

This is one such 'bet', it may not a very likely scenario but you should still be able to handle it. If we combine it with many slightly altered dilemma we can figure out your probability estimate of theory X being correct, whether you admit to having one or not.

[deleted] 07 April 2011 11:18:05PM * 0 points

You've presented this as a scenario in which you have to make a choice between two conflicting theories. But the problem you face isn't should I choose X or should I choose Y; the problem you face is that given you have this conflict, what should I do now. This problem is objective, it is different to the problem of whether X is right or Y is right, and it is solvable. Given that this is the year 2050 and humanity won't in fact be wanting, the best solution to the problem may be to wait, pending further research to resolve the conflict. This isn't an implicit bet against X and for Y, it is a solution to a different problem to the problems X and Y address.

benelliott 08 April 2011 06:56:28AM * 2 points

For sake of argument we say that the plant requires a rare and unstable isotope to get started. Earth's entire supply is contained in the plant and will decay in 24 hours.

I could also ask you a similar dilemma, but this time there is only one theory which acknowledges that whether the plant works or creates a black hole depends on a single quantum event, which has a 50% chance of going either way. What do you do? If you wouldn't launch I can ask the same question but now there's only a 25% chance of a black hole, and so on until I learn the ratio of the utility values that you assign to "post scarcity future" and "extinction of humanity". This might for example tell me that the chance of a black hole has to be less than 30% for you to press the button.

Then I ask you the original dilemma, and learn whether the probability you assign to theory X is above or below 70%. If I have far too much time on my hands I can keep modifying the dilemma with slightly altered pay-offs until I pinpoint your estimate.

wedrifid 08 April 2011 07:07:57AM 1 point

| creates a black whole

I suppose you get that when the container containing the black dye explodes....

[continue this thread »](#)

JoshuaZ 08 April 2011 03:32:23AM 1 point

This avoids the question. If it helps, try to construct a version of this in [the least convenient possible world](#). For example, one obvious thing to do would be that something about theory X means the plant can only be turned on at a certain celestial conjunction (yes, this is silly but it gets the point across. That's why it is a least convenient world) and otherwise would need to wait a thousand years.

One can vary the situation. For example, it might be that under theory X, medicine A will save a terminally ill cancer patient, and under theory Y, medicine B will save them. And A and B together will kill the patient according to both theories.

JoshuaZ 07 April 2011 01:00:33AM 5 points

There's an associated problem here that may be getting ignored: Popper isn't a terribly good writer. "The Logic of Scientific Discovery" was one of the first phil-sci books I ever read and it almost turned me off of phil-sci. This is in contrast for example with Lakatos or Kuhn who are very readable. Some of the difficulty with reading Popper and understanding his viewpoints is that he's just tough to read.

That said, I think that chapter 3 of that books makes clear that Popper's notion of falsification is more subtle than what I would call "naive Popperism". But Popper never fully gave an explanation of how to distinguish between strict falsification theory and his notions.

There's an associated important issue: many people claim to support naive Popperism as an epistemological position, either as a demarcation between science and non-science or as a general epistemological approach. In so far as both are somewhat popular viewpoints (especially among scientists) responding to and explaining what is wrong with that approach is important even as one should acknowledge that Popper's own views were arguably more nuanced.

curi 07 April 2011 03:03:25AM 0 points

I do not find Popper hard to read.

Popper never fully gave an explanation of how to distinguish between strict falsification theory and his notions.

Did you read his later books? He does explain his position. One distinguishing difference is that Popper is not a justificationist and they are. Tell me if you don't know what that means.

falenas108 07 April 2011 12:28:06AM * -1 points

From the research I have done in the last 5 minutes, it seems as though Popper believed that all good scientific theories should be subject to experiments that could prove them wrong.

Ex:

"the falsificationists or fallibilists say, roughly speaking, that what cannot (at present) in principle be overthrown by criticism is (at present) unworthy of being seriously considered; while what can in principle be so overthrown and yet resists all our critical efforts to do so may quite possibly be false, but is at any rate not unworthy of being seriously considered and perhaps even of being believed" -Popper

This seems to imply that theories can be proved false.

curi 07 April 2011 12:32:34AM -3 points

Replying to accusations of un scholarly criticism of Popper with an *unsourced* Popper quote is very silly.

That the quote doesn't say what you claim it does (as I read it), and you make no attempt to explain your reading of it, is also silly.

[deleted] 07 April 2011 12:59:57AM * 3 points

The quote came from *Conjectures and Refutations*, pg 309. I agree that it doesn't say what falenas108 claims. Plus a bit has been missed out at the end: " -- though only tentatively." Also, on the following page, Popper says:

For us [fallibilists] ... science has nothing to do with the quest for certainty or probability or reliability. We are not interested in establishing scientific theories as secure, or certain, or probable. Conscious of our own fallibility we are only interested in criticizing them and testing them, hoping to find out where we are mistaken; of learning from our mistakes; and , if we are lucky, of proceeding to better theories.

So Popper would not assert that theories can be established as definitely false.

Peterdjones 12 April 2011 08:28:49PM 0 points

Of course, in reality, fallibilism just means you don't look for certainty. You can and should look for more probable theories, or as P. calls them, "better theories".

curi 13 April 2011 01:53:00AM -2 points 

Hi Peter,

How did you find me here?

falenas108 07 April 2011 12:56:18AM 0 points 

Citation: Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*, New York: Harper and Row. Reprinted London: Routledge, 1974.

It says theories should resist being overthrown for them to be proper theories. That implies that it is possible for a theory to be overthrown.

curi 07 April 2011 01:10:57AM 1 point 

No page number isn't very nice. For anyone interested, it is on page 309, which is at the start of chapter 10 section 3.

If you read the context, you will find, for example, an explicit denouncement of the quest for certainty on the next page. Plus elaboration. Popper's position in these matters is not unclear.

curi 07 April 2011 12:59:11AM * 7 points 

A theory can be fallibly overthrown, but not definitely overthrown, in Popper's view. Quotes out of context are easy to misread when you are not familiar with the ideas, and when you make assumptions (e.g. that overthrowing must be definitive) that the author does not make.

Peterdjones 12 April 2011 08:25:35PM 1 point 

"A theory can be fallibly overthrown, but not definitely overthrown, in Popper's view. "

So maybe Jaynes was using "disprove" to mean "fallibly overthrow".

falenas108 07 April 2011 01:01:26AM 3 points 

Ok, thanks for correcting me.

[deleted] 07 April 2011 12:25:14AM * 4 points 

The naturalist philosopher Peter Godfrey Smith said this of Popper's position:

[F]or Popper, it is never possible to confirm or establish a theory by showing its agreement with observations. Confirmation is a myth. The only thing an observational test can do is to show that a theory is false...Popper, like Hume, was an inductive skeptic, and Popper was skeptical about all forms of confirmation and support other than deductive logic itself...This position, that we can never be completely certain about factual issues, is often known as fallibilism...According to Popper, we should always retain a tentative attitude towards our theories, no matter how successful they have been in the past...[a]ll we can do is try out one theory after another. A theory that we have failed to falsify up till now might, in fact, be true. But if so, we will never know this or even have reason to increase our confidence.

(From *Theory and Reality*, p. 59-61.) Is this not an accurate description? You seem to think Popper didn't believe in definitive falsification, but this doesn't seem to be a universally accepted interpretation. Note also that Godfrey-

Smith does refer to Popper's position as fallibilism, so he is not being "unscholarly." Though Popper may have held the position that falsification can't be perfectly certain, he definitely didn't take this idea too seriously because his description of science as a process (step one: come up with conjectures; step two: falsify them) makes use of falsification by experiment.

I think the answer to your overarching question can be found [here](#). If we know that certain events are more probable given that certain other events happened, i.e. [conditional probability](#), we can make inferences about the future.

curi 07 April 2011 12:44:42AM * 3 points



Is this not an accurate description?

No. To start with, it's extremely incomplete. It doesn't really discuss what Popper's position is. It just makes a few scattered statements which do not explain what Popper is about.

The word "show" is ambiguous in the phrase "show that a theory is false". To a Popperian, equivocation over the issue of what is meant there is an important issue. It's ambiguous between "show definitively" and "show fallibly".

The idea that we can show a theory is false by an experimental test (even fallibly) is also, strictly, false, as Popper explained in LScD. When you reach a contradiction, something in the whole system is false. It could be an idea you had about how to measure what you wanted to measure. There's many possibilities.

You seem to think Popper didn't believe in definitive falsification, but this doesn't seem to be a universally accepted interpretation.

It's right there in LScD on page 56. I think it's in most of his other books too. I am familiar with the field and know of no competent Popper scholars who say otherwise.

Anyone publishing to the contrary is simply incompetent, or believed low quality secondary sources without fact checking them.

Though Popper may have held the position that falsification can't be perfectly certain, he definitely didn't take this idea too seriously because his description of science as a process (step one: come up with conjectures; step two: falsify them) makes use of falsification by experiment.

You have misinterpreted when you took "falsify them" to mean "falsify them with certainty". Popper is a fallibilist.

If we know that certain events are more probable given that certain other events happened

This does not even attempt to address important problems in epistemology such as how explanatory or philosophical knowledge is created.

[deleted] 07 April 2011 01:07:57AM * 3 points



I'll agree that Godfrey-Smith's definition is incomplete, but I don't think it really matters for the purpose of this discussion: I've already said I agree that Popper did not believe in certain confirmation, and this seems to be your main problem with this quote and with the ones other people gave. You wrote:

You have misinterpreted when you took "falsify them" to mean "falsify them with certainty". Popper is a fallibilist.

No, that is not what I meant at all. What I meant was, Popper was content with the fact that experimental evidence *can* say that something is *probably* false. If he wasn't, he wouldn't have included this his view of science as a process. So even though Popper was a fallibilist, he thought that when an experimental result argued against a hypothesis, it was good enough for science.

Next:

The idea that we can show a theory is false by an experimental test (even fallibly) is also, strictly, false, as Popper explained in LScD. When you reach a contradiction, something in the whole system is false. It could be an idea you had about how to measure what you wanted to measure. There's many possibilities.

Yes, this is the old "underdetermination of theory by data" problem, which Solomonoff Induction solves--see the coinflipping example [here](#).

Moving on, you wrote:

This does not even attempt to address important problems in epistemology such as how explanatory or philosophical knowledge is created.

Would you mind elaborating on this? What specific problems are you referring to here?

curi 07 April 2011 01:37:39AM 3 points

Popper was content with the fact that experimental evidence can say that something is probably false

That is not Popper's position. That is not even close. In various passages he explicitly denies it like "not certain or probable". To Popper, the claims that the evidence tells us something is certainly true, or probably true, are cousins which share an underlying mistake. You're assuming Popper would agree with you about probability without reading any of his passages on probability in which he, well, doesn't.

Arguing what books say with people who haven't read them gets old fast. So how about you just imagine a hypothetical person who had the views I attribute to Popper and discuss that?

Would you mind elaborating on this? What specific problems are you referring to here?

For example, the answers to all questions that have a "why" in them. E.g. *why* is the Earth roughly spherical? Statements with "because" (sometimes implied) is a pretty accurate way to find explanations, e.g. "because gravity is a symmetrical force in all directions". Another example is all of moral philosophy. Another example is epistemology itself, which is a philosophy not an empirical field.

Yes, this is the old "underdetermination of theory by data" problem

Yes

Which Solomonoff Induction solves--see the coinflipping example here.

This does not solve the problem to my satisfaction. It orders theories which make identical predictions (about all our data, but not about the unknown) and then lets you differentiate by that order. But isn't that ordering arbitrary? It's just not true that short and simple theories are always best; sometimes the truth is complicated.

[deleted] 07 April 2011 01:58:09AM * -1 points

Actually, one of the reason I stood by this interpretation of Popper was because one of the quotes posted in one of the other threads here:

"the falsificationists or fallibilists say, roughly speaking, that what cannot (at present) in principle be overturned by criticism is (at present) unworthy of being seriously considered; while what can in principle be so overturned and yet resists all our critical efforts to do so may quite possibly be false, but is at any rate not unworthy of being seriously considered and perhaps even of being believed"

Which is apparently from *Conjectures and Refutations*, pg 309. Regardless, I don't care about this argument overmuch, since we seem to have moved on to some other points.

[Solomonoff Induction] does not solve the problem to my satisfaction. It orders theories which make identical predictions (about all our data, but not about the unknown) and then lets you differentiate by that order. But isn't that ordering arbitrary? It's just not true that short and simple theories are always best; sometimes the truth is complicated.

Remember that in Bayesian epistemology, probabilities represent our state of knowledge, so as you pointed out, the simplest hypothesis that fits the data so far may not be the true one because we *haven't* seen all of the data. But it is *necessarily* our best guess because of the [conjunction rule](#).

JoshuaZ 07 April 2011 02:42:59AM 1 point

Remember that in Bayesian epistemology, probabilities represent our state of knowledge, so as you pointed out, the simplest hypothesis that fits the data so far may not be the true one because we haven't seen all of the data. But it is necessarily our best guess because of the conjunction rule.

You are going to have to expand on this. I don't see how the conjunction rule implies that simpler hypotheses are in general more probable. This is true if we have two hypotheses where one is X and the other is "X and Y" but that's not how people generally apply this sort of thing. For example, I might have a sequence of numbers that for the first 10,000 terms has the nth term as the nth prime number. One hypothesis is that the nth term is always the nth prime number. But I could have as another hypothesis some high degree polynomial that matches the first 10,000 primes. That's clearly more complicated. But one can't use conjunction to argue that it is less likely.

[deleted] 07 April 2011 04:52:44AM * 1 point

Imagine that I have some set of propositions, A through Z, and I don't know the probabilities of any of these. Now let's say I'm using these propositions to explain some experimental result--since I would have uniform priors for A through Z, it follows that an explanation like "M did it" is more probable than "A and B did it," which in turn is more probable than "G and P and H did it."

JoshuaZ 07 April 2011 04:58:22AM 1 point

Yes, I agree with you there. But this is much weaker than any general form of Occam. See my example with primes. What we want to say in some form of Occam approach is much stronger than what you can get from simply using the conjunction argument.

curi 07 April 2011 02:22:42AM * 1 point

There are so many problems here that it's hard to choose a starting point.

- 1) the data set you are using is biased (it is selective. all observation is selective)
- 2) there is no such thing as "raw data" -- all your observations are interpreted. your interpretations may be mistaken.
- 3) what do you mean by "best guess"? one meaning is "most likely to be the final, perfect truth". but a different meaning is "most useful now".
- 4) You say "probabilities represent our state of knowledge". However there are infinitely many theories with the same probability. Or there would be, except for your solomonoff prior about simpler theories having higher probability. So the important part of "state of our knowledge" as represented by these probabilities consists mostly of the solomonoff prior and nothing else, because it, and it alone, is dealing with the hard problem of epistemology (dealing with theories which make identical predictions about everything we have data for).
- 5) you can have infinite data and still get all non-empirical issues wrong
- 6) regarding the conjunction rule, there is miscommunication. this does not address the point i was trying to make. i think you have a premise like "all more complicated theories are merely conjunctions of simpler theories". But that is to conceive of theories very differently than Popperians do, in what we see as a limited and narrow way. To begin to address these issues, let's consider what's better: a bald assertion, or an assertion *and* an explanation of why it is correct? If you want "most likely to happen to be the perfect, final truth" you are better off with only an unargued assertion (since any argument may be mistaken). But if you want to learn about the world, you are better off not relying on unargued assertions.

jimrandomh 07 April 2011 01:48:41AM 3 points

For example, the answers to all questions that have a "why" in them. E.g. why is the Earth roughly

spherical? Statements with "because" (sometimes implied) is a pretty accurate way to find explanations, e.g. "because gravity is a symmetrical force in all directions". Another example is all of moral philosophy. Another example is epistemology itself, which is a philosophy not an empirical field.

For a formal mathematical discussion of these sorts of problems, read Causality by Judea Pearl. He reduces cause to a combination of conditional independence and ordering, and from this he defines algorithms for discovering causal models from data, predicting the effect of interventions and computing counterfactuals.

curi 07 April 2011 01:51:03AM * 1 point

Could you give a short statement of the main ideas? How can morality be reduced to math? Or could you say something to persuade me that that book will address the issues in a way I won't think misses the point? (e.g. by showing you understand what I think the point is, otherwise I won't expect you to be able to judge if it misses the point in the way I would).

jimrandomh 07 April 2011 02:01:00AM 2 points

Sorry, I over-quoted there; Pearl only discusses causality, and a little bit of epistemology, but he doesn't talk about moral philosophy at all.

His book is all about causal models, which are directed graphs in which each vertex represents a variable and each edge represents a conditional dependence between variables. He shows that the properties of these graphs reproduce what we intuitively think of as "cause and effect", defines algorithms for building them from data and operating on them, and analyzes the circumstances under which causality can and can't be inferred from the data.

curi 07 April 2011 02:28:44AM 2 points

I don't understand the relevance.

jimrandomh 07 April 2011 02:39:41AM 2 points

Your quote seemed to be saying that that Bayesianism couldn't handle why/because questions, but Popperian philosophy could. I mentioned Pearl as a treatment of that class of question from a Bayes-compatible perspective.

curi 07 April 2011 02:54:20AM 1 point

Causality isn't explanation. X caused Y isn't the issue I was talking about.

For example, the statement "Murder is bad because it is illiberal" is an explanation of *why* it is bad. It is not a statement about causality.

You may say that "illiberal" is a short cut for various other ideas. And you may claim that eventually that reduce away to causal issues. But that would be reductionism. We do not accept that high level concepts are a mistake or that emergence isn't important.

[continue this thread »](#)

falenas108 07 April 2011 12:31:38AM -1 points

Sorry, didn't see you posted this before I replied too...

[deleted] 07 April 2011 12:34:48AM 0 points

Actually, I'm glad you replied as well--the more quotes about/by Popper that we unearth, the more accurate we will be.

paulfchristiano 07 April 2011 12:19:41AM * 5 points

I don't understand Popper's work beyond the Wikipedia summary of critical rationalism. That summary, as well as the debate here at LW, appear to be confused and essentially without value. If this is not the case, you should update this post to include not just a description of how supporters of Bayesianism don't understand Popper, but why they should care about this discussion--why Bayesianism is not, as it seems, obviously the correct answer to the question Popper is trying to answer.

If you want to make bets about the future, Bayesianism will beat whatever else you could use. To suggest that something else is an improved method of doing science is nothing more than to suggest that it is a more feasible approximation to Bayesianism. These things are mathematical facts, if you define Bayesianism and "winning" precisely.

It seems like the only possible room for debate is the choice of prior. Everyone is forced to either implicitly choose a prior or else bet in a way that is manifestly irrational. This is also a mathematical fact. The Solomonoff prior provably isn't too bad. You just have to get over the arbitrariness.

Edit: Lets make this more precise. I claim that if we play a betting game, I can reconstruct a prior from your strategy such that a Bayesian using that prior will beat you in expectation. Do you object to this mathematical statement, or do you object to the interpretation of this fact as "Bayesianism is correct"? I'm not sure which side of the fence you are on, but I suppose it must be one or the other, so if we get that sorted out maybe we can make progress.

curi 07 April 2011 12:30:32AM 1 point

I don't understand Popper's work beyond the Wikipedia summary of critical rationalism

FYI that won't work. Wikipedia doesn't understand Popper. Secondary sources promoting myths, like Jaynes did, is common. A pretty good overview is the Popper book by Bryan Magee (only like 100 pages).

without value

I posted criticisms of Jaynes' arguments (or more accurately, his assumptions). I posted an argument about support. Why don't you answer it?

You just have to get over the arbitrariness.

You are basically admitting that your epistemology is wrong. Given that Popper has an epistemology which does not have this feature, and the rejections of him by Bayesians are unscholarly mistakes, you should be interested in it!

Of course if I wrote up his whole epistemology and posted it here for you that would be nice. But that would take a long time, and it would repeat content from his books.

If you want somewhere to start online, you could read

<http://fallibleideas.com/>

If you want to make bets about the future

That is not primarily what we want. And what you're doing here is conflating Bayes' theorem (which is about probability, and which is a matter of logic, and which is correct) with Bayesian epistemology (the application of Bayes' theorem to epistemological problems, rather than to the math behind betting).

To suggest that something else is an improved method of doing science is nothing more than to suggest that it is a more feasible approximation to Bayesianism. These things are mathematical facts,

Are you open to the possibility that the general outline of your approach is itself mistaken, and there the theorems you have proven within your framework of assumptions are therefore not all true? Or:

It seems like the only possible room for debate is the choice of prior.

Are you so sure of yourself -- that you are right about many things -- that you will dismiss all rival ideas without even having to know what they say? Even when they offer things your approach doesn't have, such as *not having arbitrary foundations*.

What you're doing is accepting ideas which have been popular since Aristotle. When you think no other ways are possible, that's bias talking. Your ideas have become common sense (not the Bayes part, but the philosophical approach to epistemology you are taking which comes before you use Bayes's theorem at all).

Here let me ask you a question: has any Bayesian ever published any substantive criticism of an important idea in Popper's epistemology? Someone should have done it, right? And if no one ever has, then you should be interested in investigating, right? And also interested in investigating what is wrong with your movement that it never addressed rival ideas in scholarly debate. (I have looked for such a criticism. Never managed to find one.)

Peterdjones 15 April 2011 03:14:18PM 3 points

Why don't you fix the WP article?

timtyler 07 April 2011 12:48:52PM * 1 point

what you're doing here is conflating Bayes' theorem (which is about probability, and which is a matter of logic, and which is correct) with Bayesian epistemology (the application of Bayes' theorem to epistemological problems, rather than to the math behind betting).

That's because to a Bayesian, these things are the same thing. Epistemology is all about probability - and visa versa. Bayes's theorem includes induction and confirmation. You can't *accept* Bayes's theorem and *reject* induction without crazy inconsistency - and Bayes's theorem is just the math of probability theory.

[deleted] 07 April 2011 01:05:19PM 0 points

If I understand correctly, I think curi is saying that there's no *reason* for probability and epistemology to be the same thing. That said, I don't entirely understand his/her argument in this thread, as some of the criticisms he/she mentions are vague. For example, what are these "epistemological problems" that Popper solves but Bayes doesn't?

paulfchristiano 07 April 2011 01:12:55AM * 2 points

Having read the website you linked to in its entirety, I think we should defer this discussion (as a community) until the next time you explain why someone's particular belief is wrong, at which point you will be forced to make an actual claim which can be rejected.

In particular, if you ever try to make a claim of the form "You should not believe X, because Bayesianism is wrong, and undesirable Y will happen if you act on this belief" then I would be interested in the resulting discussion. We could do the same thing now, I guess, if you want to make such a claim of some historical decision.

Edit: changed wording to be less of an ass.

curi 07 April 2011 01:24:01AM 2 points

In its entirety? Assuming you spent 40 minutes reading, 0 minutes delay before you saw my post, 0 minutes reading my post here, and 2:23 writing your reply, then you read at a speed of around 833 words per minute. That is very impressive. Where did you learn to do that? How can I learn to do that too?

Given that I do make claims on my website, I wonder why you don't pick one and point out something you think is wrong with it.

paulfchristiano 07 April 2011 01:33:16AM * 2 points

Fair, fair. I should have thought more and been less heated. (My initial response was even worse!)

I did read the parts of your website that relate to the question at hand. I do skim at several hundred words per minute (in much more detail than was needed for this application), though I did not spend the entire time reading. Much of the content of the website (perfectly reasonably) is devoted to things not really germane to this discussion.

If you really want (because I am constitutively incapable of letting an argument on the internet go) you could point to a particular claim you make, of the form I asked for. My issue is not really that I have an objection to any of your arguments--its that you seem to offer no concrete points where your epistemology leads to a different conclusion than Bayesianism, or in which Bayesianism will get you into trouble. I don't think this is necessarily a flaw with your website--presumably it was not designed first and foremost as a response to Bayesianism--but given this observation I would rather defer discussion until such a claim does come up and I can argue in a more concrete way.

To be clear, what I am looking for is a statement of the form: "Based on Bayesian reasoning, you conclude that there is a 50% chance that a singularity will occur by 2060. This is a dangerous and wrong belief. By acting on it you will do damage. I would not believe such a thing because of my improved epistemology. Here is why my belief is more correct, and why your belief will do damage." Or whatever example it is you would like to use. Any example at all. Even an argument that Bayesian reasoning with the Solomonoff prior has been "wrong" where Popper would be clearly "right" at any historical point would be good enough to argue about.

curi 07 April 2011 01:47:06AM * 0 points

statement of the form: "Based on Bayesian reasoning, you conclude that there is a 50% chance that a singularity will occur by 2060. This is a dangerous and wrong belief. By acting on it you will do damage I would not believe such a thing because of my improved epistemology.

Do you assert that? It is wrong and has real world consequence. In *The Beginning of Infinity* Deutsch takes on a claim of a similar type (50% probability of humanity surviving the next century) using Popperian epistemology. You can find Deutsch explaining some of that material here: <http://groupspaces.com/oxfordtranshumanists/pages/past-talks>

While Fallible Ideas does not comment on Bayesian Epistemology directly, it takes a different approach. You do not find Bayesians advocating the same ways of thinking. They have a different (worse, IMO) emphasis.

I wonder if you think that all mathematically equivalent ways of thinking are equal. I believe they aren't because some are more convenient, some get to answers more directly, some make it harder to make mistakes, and so on. So even if my approach was compatible with the Bayesian approach, that wouldn't mean we agree or have nothing to discuss.

Fair, fair. I should have thought more and been less heated. (My initial response was even worse!)

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference? You learned Bayesian stuff but it apparently didn't solve your problem, whereas my epistemology did solve mine.

paulchristiano 07 April 2011 02:09:39AM 3 points

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference?

No. It provides an example of a way in which you are better than me. I am overwhelmingly confident that I can find ways in which I am better than you.

Do you assert that? It is wrong and has real world consequence. In *The Beginning of Infinity* Deutsch takes on a claim of a similar type (50% probability of humanity surviving the next century) using Popperian epistemology. You can find Deutsch explaining some of that material here: <http://groupspaces.com/oxfordtranshumanists/pages/past-talks>

Could you explain how a Popperian disputes such an assertion? Through only my own fault, I can't listen to an mp3 right now.

My understanding is that anyone would make that argument in the same way: by providing evidence in the Bayesian sense, which would convince a Bayesian. What I am really asking for is a description of why your beliefs aren't the same as mine but better. Why is it that a Popperian disagrees with a Bayesian in this case? What argument do they accept that a Bayesian wouldn't? What is the corresponding calculation a Popperian does when he has to decide how to gamble with the lives of six billion people on an uncertain assertion?

I wonder if you think that all mathematically equivalent ways of thinking are equal. I believe they aren't because some are more convenient, some get to answers more directly, some make it harder to make mistakes, and so on. So even if my approach was compatible with the Bayesian approach, that wouldn't mean we agree or have nothing to discuss.

I agree that different ways of thinking can be better or worse even when they come to the same conclusions. You seem to be arguing that Bayesianism is wrong, which is a very different thing. At best, you seem to be claiming that trying to come up with probabilities is a bad idea. I don't yet understand exactly what you mean. Would you never take a bet? Would never take an action that could possibly be bad and could possibly be good, which requires weighing two uncertain outcomes?

This brings me back to my initial query: give a specific case where Popperian reasoning diverges from Bayesian reasoning, explain why they diverge, and explain why Bayesianism is wrong. Explain why Bayesian's willingness to bet does harm. Explain why Bayesians are slower than Popperians at coming to the same conclusion. Whatever you want.

I do not plan to continue this discussion except in the pursuit of an example about which we could actually argue productively.

curi 07 April 2011 02:46:51AM 0 points



Could you explain how a Popperian disputes such an assertion? [(50% probability of humanity surviving the next century)]

e.g. by pointing out that whether we do or don't survive depends on human choices, which in turn depends on human knowledge. And the growth of knowledge is not predictable (exactly or probabilistically). If we knew its contents and effects now, we would already have that knowledge. So this is not prediction but prophecy. And prophecy has build in bias towards pessimism: because we can't make predictions about future knowledge, prophets in general make predictions that disregard future knowledge. These are explanatory, philosophical arguments which do not rely on evidence (that is appropriate because it is not a scientific or empirical mistake being criticized). No corresponding calculation is made at all.

You ask about how Popperians make decisions if not with such calculations. Well, say we want to decide if we should build a lot more nuclear power plants. This could be taken as gambling with a lot of lives, and maybe even all of them. Of course, *not* doing it could also be taken as a way of gambling with lives. There's no way to never face any potential dangers. So, how do Popperians decide? They *conjecture* an answer, e.g. "yes". Actually, they make many conjectures, e.g. also "no". Then they criticize the conjectures, and make more conjectures. So for example I would criticize "yes" for not providing enough explanatory detail about why it's a good idea. Thus "yes" would be rejected, but a variant of it like "yes, because nuclear power plants are safe, clean, and efficient, and all the criticisms of them are from silly luddites" would be better. If I didn't understand all the references to longer arguments being made there, I would criticize it and ask for the details. Meanwhile the "no" answer and its variants will get refuted by criticism. Sometimes entire infinite categories of conjectures will be refuted by a criticism, e.g. the anti-nuclear people might start arguing with conspiracy theories. By providing a general purpose argument against all conspiracy theories, I could deal with all their arguments of that type. Does this illustrate the general idea for you?

You seem to be arguing that Bayesianism is wrong, which is a very different thing.

I think it's wrong as an epistemology. For example because induction is wrong, and the notion of positive support is wrong. Of course Bayes' theorem is correct, and various math you guys have done is correct. I keep getting conflicting statements from people about whether Bayesianism conflicts with Popperism or not, and I don't want to speak for you guys, nor do I

want to discourage anyone from finding the shared ideas or discourage them from learning from both.

Would you never take a bet?

Bets are made on *events*, like which team wins a sports game. Probabilities are fine for *events*. Probabilities of *the truth of theories* is problematic (b/c e.g. there is no way to make them non-arbitrary). And it's not something a fallibilist can bet on because he accepts we never know the final truth for sure, so how are we to set up a decision procedure that decides who won the bet?

Would never take an action that could possibly be bad and could possibly be good, which requires weighing two uncertain outcomes?

We are not afraid of uncertainty. Popperian epistemology is fallibilist. It rejects certainty. Life is always uncertain. That does not imply probability is the right way to approach all types of uncertainty.

This brings me back to my initial query: give a specific case where Popperian reasoning diverges from Bayesian reasoning, explain why they diverge, and explain why Bayesianism is wrong. Explain why Bayesian's willingness to bet does harm. Explain why Bayesians are slower than Popperians at coming to the same conclusion. Whatever you want.

Bayesian reasoning diverges when it says that ideas can be positively supported. We diverge because Popper questioned the concept of positive support, as I posted in the original text on this page, and which no one has answered yet. The criticism of positive support begins by considering what it is (you tell me) and how it differs from consistency (you tell me).

Larks 08 April 2011 01:03:10AM 0 points

And the growth of knowledge is not predictable (exactly or probabilistically). If we knew its contents and effects now, we would already have that knowledge.

You're equivocating between "knowing exactly the contents of the new knowledge", which may be impossible for the reason you describe, and "know some things about the effect of the new knowledge", which we can do. As Eliezer said, I may not know which move Kasparov will make, but I know he will win.

jake987722 07 April 2011 03:24:11AM 6 points

So, how do Popperians decide? They conjecture an answer, e.g. "yes". Actually, they make many conjectures, e.g. also "no". Then they criticize the conjectures, and make more conjectures. So for example I would criticize "yes" for not providing enough explanatory detail about why it's a good idea. Thus "yes" would be rejected, but a variant of it like "yes, because nuclear power plants are safe, clean, and efficient, and all the criticisms of them are from silly luddites" would be better. If I didn't understand all the references to longer arguments being made there, I would criticize it and ask for the details. Meanwhile the "no" answer and its variants will get refuted by criticism. Sometimes entire infinite categories of conjectures will be refuted by a criticism, e.g. the anti-nuclear people might start arguing with conspiracy theories. By providing a general purpose argument against all conspiracy theories, I could deal with all their arguments of that type. Does this illustrate the general idea for you?

Almost, but you seem to have left out the rather important detail of how actually *make the decision*. Based on the process of criticizing conjectures you've described so far, it seems that there are two basic routes you can take to finish the decision process once the critical smoke has cleared.

First, you can declare that, since there is no such thing as confirmation, it turns out that no conjecture is better or worse than any other. In this way you don't actually make a decision and the problem remains unsolved.

Second, you can choose to go with the conjecture that best weathered the criticisms you were able to muster. That's fine, but then it's not clear that you've done anything different

from what a Bayesian would have done--you've simply avoided explicitly talking about things like probabilities and priors.

Which of these is a more accurate characterization of the Popperian decision process? Or is it something radically different from these two altogether?

curi 07 April 2011 03:59:34AM 2 points 

When you have exactly one non-refuted theory, you go with that.

The other cases are more complicated and difficult to understand.

Suppose I gave you the answer to the other cases, and we talked about it enough for you to understand it. What would you change your mind about? What would you concede?

If I convinced you of this one single issue (that there is a method for making the decision), would you follow up with a thousand other objections to Popperian epistemology, or would we have gotten somewhere?

If you have lots of other objections you are interested in, I would suggest you just accept for now that we have a method and focus on the other issues first.

[option 1] since there is no such thing as confirmation, it turns out that no conjecture is better or worse than any other.

But some are criticized and some aren't.

[option 2] conjecture that best weathered the criticisms you were able to muster

But how is that to be judged?

No, we always go with uncriticized ideas (which may be close variants of ideas that were criticized). Even the terminology is very tricky here -- the English language is not well adapted to expressing these ideas. (In particular, the concept "uncriticized" is a very substantive one with a lot of meaning, and the word for it may be misleading, but other words are even worse. And the straightforward meaning is OK for present purposes, but may be problematic in future discussion.).

[Or is it something radically different from these two altogether?

Yes, different. Both of these are justificationist ways of thinking. They consider how much justification each theory has. The first one rejects a standard source of justification, does not replace it, and ends up stuck. The second one replaces it, and ends up, as you say, reasonably similar to Bayesianism. It still uses the same basic method of tallying up how much of some good thing (which we call justification) each theory has, and then judging by what has the most.

Popperian epistemology does not justify. It uses criticism for a different purpose: a criticism is an explanation of a mistake. By finding mistakes, and explaining what the mistakes are, and conjecturing better ideas which we think won't have those mistakes, we learn and improve our knowledge.

[continue this thread »](#)

Desrtopa 07 April 2011 01:58:08AM 4 points 

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference? You learned Bayesian stuff but it apparently didn't solve your problem, whereas my epistemology did solve mine.

It doesn't take Popperian epistemology to learn social fluency. I've learned to limit conflict and improve the productivity of my discussions, and I am (to the best of my ability) Bayesian in my epistemology.

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

curi 07 April 2011 02:11:59AM -4 points

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

I have considered that. Popperian epistemology helps with these issues more. I don't want to argue about that now because it is an advanced topic and you don't know enough about my epistemology to understand it (correct me if I'm wrong), but I thought the example could help make a point to the person I was speaking to.

Desrtopa 07 April 2011 02:15:42AM * 2 points

If I don't understand your explanation and am interested in it, I'm prepared to do the research in order to understand it, but if you can only assert why your epistemology *should* result in better social learning and not demonstrate that it does so for people in general, I confess that I will probably not be interested enough to follow up.

I will note though, that stating the assumption that another does not understand, but leaving them free to correct you, strikes me as a markedly worse way to minimize conflict and aggression than *asking* if they have the familiarity necessary to understand the explanation.

curi 07 April 2011 02:25:58AM 1 point

You could begin by reading

<http://fallibleideas.com/emotions>

And the rest of the site. If you don't understand any connections between it and Popperian epistemology, feel free to ask.

I'm not asking you to be interested in this, but I do think you should have some interest in rival epistemologies.

[continue this thread »](#)

JoshuaZ 07 April 2011 02:07:05AM 2 points

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

That's a claim that only makes sense in certain epistemological systems...

curi 07 April 2011 02:09:13AM * 3 points

I don't have a problem with the main substance of that argument, which I agree with. Your implication that we would reject this idea is mistaken.

JoshuaZ 07 April 2011 02:36:12AM 0 points

I don't have a problem with the main substance of that argument, which I agree with. Your implication that we would reject this idea is mistaken.

Hmm? I'm not sure who you mean by we? If you mean that someone supporting a Popperian approach to epistemology would probably find this idea reasonable than I agree with you (at least empirically, people claiming to support some form of Popperian approach seem ok with this sort of thing. That's not to say I understand how they think it is implied/ok in a Popperian framework).

paulfchristiano 07 April 2011 01:02:58AM 9 points

Here let me ask you a question: has any Bayesian ever published any substantive criticism of an important idea in Popper's epistemology? Someone should have done it, right?

Most things in the space of possible documents can't be refuted, because they don't correspond to anything refutable. They are simply confused, and irredeemably. In the case of epistemology, virtually everything that has ever been said falls into this category. I am glad that I don't have to spend time thinking about it, because it is *solved*. I would not generally criticize a rival's ideas, because I no longer care. The problem is solved, and I can go work on things that still matter.

Are you so sure of yourself -- that you are right about many things -- that you will dismiss all rival ideas without even having to know what they say?

Once I know the definitive answer to a question, I will dismiss all other answers (rather than trying to poke holes in them). The only sort of argument which warrants response is an objection to my current definitive answer. So ignorance of Popper is essentially irrelevant (and I suspect I couldn't object to anything in his philosophy, because it has essentially no content concrete enough to be defeated by mere reasoning).

The real question, in fact the only question, is whether the arbitrariness of choosing a prior can be surmounted--whether my current answer is not actually definitive. If someone came to me and said they had a solution to this problem I would be interested, except that I am fairly confident the problem has no solution for what are essentially obvious reasons. Popper avoids this problem by not even describing his epistemology precisely enough to express the difficulty.

Really this entire discussion comes down to what we want out of epistemology.

That [guiding betting] is not primarily what we want.

What do you want? I don't understand at all. Whatever you specify, I would be shocked if critical rationality provided it. Here is what I want, and maybe you will agree:

I want to decide between action A and action B. To do this, I want to evaluate the consequences of action A and action B. To do this, I want to predict something about the world. In particular, by choosing B instead of A, I am making a bet about the consequences of A and B. I would like to make such bets in the best possible way.

Lo! This is precisely what Bayesianism allows me to do. Why is there more to say?

You can object that it involves knowing a prior. But from the problem statement it is *obvious* (as a mathematical fact) that there is a universe in which each possible prior is the best one. Is there a strategy that does better than Bayesianism with a reasonable prior in all possible universes? Maybe, but Popper's ideas aren't nearly precise enough to answer the question (by which I mean, not even at the point where this question, to me clearly the most important one, is meaningful). Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

If I have to bet, or make a decision that effects peoples lives which amounts to a bet, I am going to use Bayesianism, or a computational heuristic which I justify by Bayesianism. Doing something else seems irresponsible.

curi 07 April 2011 02:07:06AM 1 point

Most things in the space of possible documents can't be refuted, because they don't correspond to anything refutable. They are simply confused, and irredeemably.

You don't think confused things can be criticized? You can, for example, point out ambiguous passages. That would be a criticism. If they have no clarification to offer, then it would be (tentatively and fallibly) decisive (pending some reason to reconsider).

But you haven't provided any argument that Popper in particular was confused, irrefutable, or whatever. I don't know about you, but as someone who wants to improve my epistemological knowledge I think it's important to consider all the major ideas in the field at the very least enough to know one good criticism of each.

Refusing to address criticism because you think you already have the solution is very closed minded, is it not? You think you're done with thinking, you have the final truth, and that's that..?

The only sort of argument which warrants response is an objection to my current definitive answer.

Popper published several of those. Where's the response from Bayesians?

One thing to note is it's hard to understand his objections without understanding his philosophy a bit more broadly (or you will misread stuff, not knowing the broader context of what he is trying to say, what assumptions he does not share with you, etc...)

The real question, in fact the only question, is whether the arbitrariness of choosing a prior can be surmounted--whether my current answer is not actually definitive. If someone came to me and said they had a solution to this problem I would be interested

Popper solved that problem.

I am fairly confident the problem has no solution for what are essentially obvious reasons

The standard reasons seem obvious because of your cultural bias. Since Aristotle some philosophical assumptions have been taken for granted by almost everyone. Now most people regard them as obvious. Given those assumptions, I agree that your conclusion follows (no way to avoid arbitrariness). The assumptions are called "justificationism" by Popperians, and are criticized in detail. I think you ought to be interested in this.

One criticism of justificationism is that it *causes* the regress/arbitrariness/foundations problem. The problem doesn't exist automatically but is being created by your own assumptions.

Popper avoids this problem by not even describing his epistemology precisely enough to express the difficulty.

What are you talking about? You haven't read his books and claim he didn't give enough detail? He was something of a workaholic who didn't watch TV, didn't have a big social life, and worked and wrote all the time.

What do you want?

To create knowledge, including explanatory and non-instrumentalist knowledge. You come off like a borderline positivist to me, who has trouble with the notion that non-empirical stuff is even meaningful. (No offense intended, and I'm not assuming you actually are a positivist, but I'm not really seeing much difference yet.)

To do this, I want to evaluate the consequences of action A and action B. To do this, I want to predict something about the world.

To take one issue, besides predicting the physical results of your actions you also need a way to judge which results are good or bad. That is moral knowledge. I don't think Bayesianism addresses this well.

Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

Neither. You can and should do better!

David_Allen 07 April 2011 04:16:38PM 0 points



To take one issue, besides predicting the physical results of your actions you also need a way to judge which results are good or bad. That is moral knowledge. I don't think Bayesianism addresses this well.

Given well defined contexts and meanings for good and bad I don't see why Bayesianism could not be effectively applied to moral problems.

curi 07 April 2011 06:40:28PM 0 points



Yes, *given moral assertions* you can then analyze them. Well, sort of. You guys rely on empirical evidence. Most moral arguments don't.

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

David_Allen 07 April 2011 08:26:10PM 0 points



Yes, given moral assertions you can then analyze them. Well, sort of. You guys rely on empirical evidence. Most moral arguments don't.

First of all, you shouldn't lump me in with the Yudkowskyist Bayesians. Compared to them and to you I am in a distinct third party on epistemology.

Bayes' theorem is an abstraction. If you don't have a reasonable way to transform your problem to a form valid within that abstraction then of course you shouldn't use it. Also, if you have a problem that is solved more efficiently using another abstraction, then use that other abstraction.

This doesn't mean that Bayes' theorem is useless, it just means there are domains of reasonable usage. The same will be true for your [Popperian decision making](#).

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

These are just computable processes; if Bayesianism is in some sense [Turing complete](#) then it can be used to do all of this; it just might be very inefficient when compared to other approaches.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods. Other aspects should probably be accomplished using other methods.

curi 07 April 2011 08:41:46PM 0 points



First of all, you shouldn't lump me in with the Yudkowskyist Bayesians. Compared to them and to you I am in a distinct third party on epistemology.

Sorry. I have no idea who is who. Don't mind me.

This doesn't mean that Bayes' theorem is useless, it just means there are domains of reasonable usage. The same will be true for your Popperian decision making.

The Popperian method is universal.

if Bayesianism is in some sense Turing complete then it can be used to do all of this

Well, umm, yes but that's no help. my iMac is definitely Turing complete. It could run an AI. It could do whatever. But we don't know how to make it do that stuff. Epistemology should help us.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods.

Example or details?

David_Allen 07 April 2011 09:59:13PM 0 points



Sorry. I have no idea who is who. Don't mind me.

No problem, I'm just pointing out that there are other perspectives out here.

The Popperian method is universal.

Sure, in the sense it is Turing complete; but that doesn't make it the most efficient approach for all cases. For example I'm not going to use it to decide the answer to the statement " $2 + 3$ ", it is much more efficient for me to use the arithmetic abstraction.

But we don't know how to make it do that stuff. Epistemology should help us.

Agreed, it is one of the reasons that I am actively working on epistemology.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods.

Example or details?

[The naive Bayes classifier](#) can be an effective way to classify discrete input into independent classes. Certainly for some cases it could be used to classify something as "good" or "bad" based on example input.

[Bayesian networks](#) can capture the meaning within interdependent sets. For example the meaning of words forms a complex network; if the meaning of a single word shifts it will probably result in changes to the meanings of related words; and in a similar way ideas on morality form connected interdependent structures.

Within a culture a particular moral position may be dependent on other moral positions, or even other aspects of the culture. For example a combination of religious beliefs and inheritance traditions might result in a belief that a husband is justified in killing an unfaithful wife. A Bayesian network trained on information across cultures might be able to identify these kinds of relationships. With this you could start to answer questions like "Why is X moral in the UK but not in Saudi Arabia?"

curi 08 April 2011 12:37:39AM 0 points

Sure, in the sense it is Turing complete;

No, in the sense that it *directly* applies to all types of knowledge (which any epistemology applies to -- which i think is all of them, but that doesn't matter to universality).

Not in the sense that it's Turing complete so you could, by a roundabout way and using whatever methods, do anything.

I think the basic way we differ is you have despaired of philosophy getting anywhere, and you're trying to get rigor from math. But Popper saved philosophy. (And most people didn't notice.) Example:

With this you could start to answer questions like "Why is X moral in the UK but not in Saudi Arabia?"

You have very limited ambitious. You're trying to focus on small questions b/c you think bigger ones like: what *is* moral objectively? are too hard and, since you math won't answer them, it's hopeless.

[continue this thread »](#)

JoshuaZ 07 April 2011 06:58:15PM * 2 points

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

You've repeatedly claimed that the Popperian approach can somehow address moral issues. Despite requests you've shown no details of that claim other than to say that you do the same thing you would do but with moral claims. So let's work through a specific moral issue. Can you take an example of a real moral issue that has been controversial historically (like say slavery or free speech) and show how the Popperian would approach? An concrete worked out example would be very helpful.

curi 07 April 2011 07:00:42PM * -2 points

http://lesswrong.com/lw/552/reply_to_benelliott_about_popper_issues/3uv7

And it creates moral knowledge by conjecture and refutation, same as any other knowledge. If you understand how Popper approaches any kind of knowledge (which I have written about a bunch here), then you know how he approaches moral knowledge too.

JoshuaZ 07 April 2011 07:10:36PM 0 points

And it creates moral knowledge by conjecture and refutation, same as any other knowledge. If you understand how Popper approaches any kind of knowledge (which I have written about a bunch here), then you know how he approaches moral knowledge too.

Consider that you are replying to a statement I just said that all you've done is say that it would use the same methodologies. Given that, does this reply seem sufficient? Do I need to repeat my request for a worked example (which is not included in your link)?

[deleted] 07 April 2011 01:50:34AM -4 points

Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

Saying your epistemology has a "necessary flaw" is an admission of defeat, that it doesn't work. The "necessary flaw" is unavoidable if you are committed to the justificationist way of thinking. Popper saw that the whole idea of justification is wrong and he offered a different idea to replace it - an idea with no known flaws. You criticize Popper for being underspecified, yet he elaborated on his ideas in many books. And, furthermore, no amount of mathematical precision or formalism will paper over cracks in justificationist epistemologies.

curi 07 April 2011 01:54:23AM * -1 points

It's interesting how different Bayesians say different things. They don't seem to all agree with each other even about their basic claims. Sometimes Bayesianism is proved, other times it is acknowledged to have known flaws. Sometimes it may be completely compatible with Popper, other times it is dethroning Popper. It seems to me that perhaps Bayesianism is a bit underspecified. I wonder why they haven't sorted out these internal disputes.

JoshuaZ 07 April 2011 03:41:08AM * 3 points

Sometimes Bayesianism is proved, other times it is acknowledged to have known flaws. Sometimes it may be completely compatible with Popper, other times it is dethroning Popper. It seems to me that perhaps Bayesianism is a bit underspecified. I wonder why they haven't sorted out these internal disputes.

There are disputes among the Bayesians. But you are confusing different issues. First, the presence of internal disputes about the borders of an idea is not a priori a problem with an idea that is in progress. The fact that evolutionary biologists disagree about how much neutral drift matters isn't a reason to reject evolution. (It is possible that I'm reading an unintended implication here.)

Moreover, most of what you are talking about here are not contradictions but failure to understand. That Bayesianism has flaws is a distinct claim from when someone talks about something like Cox's theorem which is the sort of result that Bayesians are talking about that you refer to as "Sometimes Bayesianism is proved"(which incidentally is a terribly unhelpful and vague way of discussing the point). The point of results like Cox's theorem is that if one very broad attempts under certain very weak assumptions to formalize epistemology you must end up with some form of Bayesianism. At the

same time it is important to keep in mind that this isn't saying all that much. It doesn't for example say anything about what one's priors should be. Thus one has the classical disagreement between objective and subjective Bayesians based on what sort of priors to use (and within each of those there is further breakdown. LessWrong seems to mainly have objective Bayesians favoring some form Occam prior, although just what is not clear.) Similarly, when discussing whether or not Bayesianism is compatible with Popper depends a lot on what one means by "Bayesianism", "compatible" and "Popper". Bayesianism is certainly not compatible with a naive-Popperian approach, which is what many are talking about when they say that it is not compatible (and as you've already noted Popper himself wasn't a naive Popperian). But some people use Popper to mean the idea that given an interesting hypothesis one should search out for experiments which would be likely to falsify the hypothesis if it is false (an idea that actually predates Popper) but what one means by falsify can be a problem.

paulfchristiano 07 April 2011 01:53:27AM 2 points

Saying your epistemology has a "necessary flaw" is an admission of defeat,

In this case, its recognition of reality. I repeat that I would like to defer this conversation until we have something concrete to disagree about. Until then I don't care about that difference.

[deleted] 07 April 2011 02:24:53AM 0 points

The "necessary flaw" arises because all justificationist epistemologies lead to infinite regress or circular arguments or appeals to authority (or even sillier things). That you think there is no alternative to justificationism and I don't is something concrete we disagree about.

David_Allen 07 April 2011 04:02:37PM 1 point

Adding a reference for this comment: [Münchhausen Trilemma](#).

endoself 07 April 2011 12:05:45AM 2 points

The thing intended as the proof is most of chapter 2. I dislike Jaynes' assumptions there, since I find many of them superfluous compared to other proofs. You probably like them even less, since one is "Representation of degrees of plausibility by real numbers".

curi 07 April 2011 12:09:20AM * 2 points

It cannot be a proof of *Bayesian epistemology* itself if it makes assumptions like that.

It is merely a proof of some theorems in Bayesian epistemology given some premises that Bayesians like.

If you have a different proof which *does not make assumptions I disagree with*, then let's hear it. Otherwise you can give up on *proving* and start *arguing* why I should agree with your starting points. Or maybe even, say, engaging with Popper's arguments and pointing out mistakes in them (if you can find any).

Peterdjones 12 April 2011 08:31:43PM 1 point

You are complaining it is not a deduction of Bayes from no assumptions whatever. But all it needs to be is that those assumptions can be made to "work"--ie applied without contradiction, quodlibet or other disaster.

Peterdjones 15 April 2011 03:18:53PM 0 points

Remember, Popper himself said it all starts with common sense.

endoself 07 April 2011 02:42:12AM -1 points

I agree that it is by no means a complete proof of Bayesian epistemology. The book I pointed you to might have a more complete one, though I doubt it will be complete since it seems more like a book about using statistics than about rigorously understanding epistemology.

I am currently collecting the necessary knowledge to write the full proof myself, if it is possible (not because of this debate, because I kept being annoyed by unjustified assumptions that didn't even seem necessary).

curi 07 April 2011 02:55:59AM 0 points

Good luck. But, umm, do you have some argument against fallibilism? Because you're going to need one.

endoself 07 April 2011 03:35:59AM * 1 point

I think I massively overstated my intention. I meant the full proof of the stuff we know; the thing I think could be in *Mathematical Statistics, Volume 1: Basic and Selected Topics*.

Anyways, I think I accept fallibilism, at least from the Wikipedia page. Why do you think I don't? This is understandable, because I've been talking about idealized agents a lot more than about humans actually applying Bayesianism.

curi 07 April 2011 03:48:40AM 1 point

I think you are not a fallibilist because you want to prove philosophical ideas.

But we can't have certainty. So what do you even think it means to "prove" them? Why do you want to prove them instead of give good arguments on the matter?

endoself 07 April 2011 04:15:22AM * 0 points

I use the word prove because I'm doing it deductively in math. I already linked you to the $2+2=3$ thing, I believe. Also, the question of how I would, for example, change AI design if a well-known theorem is wrong (pretend it is the future and the best theorems proving Bayesianism are better-known and I am working on AI design) is both extremely hard to answer and unlikely to be necessary. Well unlikely is the wrong word; what is $P(X | \text{"There are no probabilities"})$? :)

calef 07 April 2011 05:10:07AM * 1 point

Probably the most damning criticism you'll find, curi, is that fallibilism isn't *useful* to the Bayesian.

The fundamental disagreement here is somewhere in the following statement:

"There exist true things, and we have a means of determining how likely it is for any given statement to be true. Furthermore, a statement that has a high likelihood of being true should be believed over a similar statement with a lower likelihood of being true."

I suspect your disagreement is in one of several places.

1) You disagree that there even exist epistemically "true" facts. 2) That we can determine how likely something is to be true. or 3) That likelihood of being true (as defined by us) is reason to believe the truth of something.

I can actually flesh out your objections to all of these things.

For 1, you could probably successfully argue that we aren't capable of determining if we've ever actually arrived at a true epistemic statement because real certainty doesn't exist, thus the existence or nonexistence of true epistemic statements is on the same epistemological footing as the existence of God--i.e. shaky to the point of not concerning oneself with them all together.

2 basically ties in with the above directly.

3 is a whole 'nother ball game, and I don't think it's really been broached yet by anyone, but it's certainly a valid point of contention. I'll leave it out unless you'd like to pursue it.

The Bayesian counter to all of these is simply, "That doesn't really do anything for me."

Declaring we have certainty, and quantifying it as best we can is *incredibly useful*. I can pick up an apple and let go. It will fall to the ground. I have an incredibly huge amount of certainty in my ability to repeat that experiment.

That I cannot foresee the philosophical paradigm that will uproot my hypothesis that dropped apples fall to the ground is not a very good reason to reject my relative certainty in the soundness of my hypothesis. Such a apples-aren't-falling-when-dropped paradigm would literally (and necessarily) uproot everything else we know about the world.

Basically, what I'm trying to say is that all you're ever going to get out of a Bayesian is, "No, I disagree. I think we can have certainty." And the only way you could disprove conclusions made by Bayesians are through means the Bayesian would have already seen, and thus the Bayesian would have already rejected said conclusion.

You've already outlined that the fallibilist will just keep tweaking explanations until an explanation with no criticism is reached. I think you might find Bayesianism more palatable if you just pretend that we aren't trying to find certainty, just say we're trying to minimize criticism.

This probably hasn't been a very satisfying answer. I certainly agree it's useful to have an understanding of the biases to our certainties. I also think Bayesianism happens to build that into itself quite well. Personally, I don't think there's anything I'm absolutely certain about, because to claim so would be silly.

endoseif 07 April 2011 05:32:57AM 1 point

Small nitpick: I don't like your use of the word 'certainty' here. Especially in philosophy, it has too much of a connotation of "literally impossible for me to be wrong" rather than "so ridiculously unlikely that I'm wrong that we can just ignore it", which may cause confusion.

calef 07 April 2011 05:40:16AM 0 points

Where don't you like it? I don't think anyone actually argues for your first definition, because, like I said, it's silly. I think curi's point is that fallibilism is predicated on your second definition not (ever?) being a valid claim.

My point is that the things we are "certain" about (as per your second definition) probably coincide almost exactly with "statements without criticism" as per curi's definition(s).

[continue this thread »](#)

curi 07 April 2011 04:35:29AM -3 points

I think we have very different goals, and that the Popperian ones are better.

There is more to epistemology, and to philosophy, than math.

I'd say you are practically trying to *eliminate all philosophy*. And that saying you have an epistemology at all is very misleading, because epistemology is a philosophical field.

JoshuaZ 07 April 2011 05:24:07AM * 3 points

I think we have very different goals, and that the Popperian ones are better.

So could you be more precise in how you think the goals differ and why the Popperian goals are better?

There is more to epistemology, and to philosophy, than math.

I'd say you are practically trying to eliminate all philosophy. And that saying you have an epistemology at all is very misleading, because epistemology is a philosophical field.

Huh? Do you mean that because the Bayesians have made precise mathematical claims it somehow ceases to be an epistemological system? What does that even mean? I don't incidentally know what it means to eliminate philosophy, but areas can certainly be carved off from philosophy into other branches. Indeed, this is generally what happens. Philosophy is the big grab bag of things that we don't have a very good precise feel for. As we get more precise understanding things break off. For example, biology broke off from philosophy (when it broke off isn't clear, but certainly by 1900 it was a separate field) with the philosophers now only focusing on the remaining tough issues like how to define "species". Similarly, economics broke off. Again, where it broke off is tough (that's why Bentham and Adam Smith are often both classified as philosophers). A recent break off has been psychology, which some might argue is still in the process. One thing that most people would still see as clearly in the philosophy realm is moral reasoning. Indeed, some would argue that the ultimate goal of philosophy is to eliminate itself.

If it helps at all, in claiming that the Bayesians lack an epistemology or are not trying to philosophy it might help to taboo both epistemology and philosophy and restate those statements. What do those claims mean in a precise way?

curi 07 April 2011 05:33:11AM 0 points



Different people are telling me different things. I have been told some very strong instrumentalist and anti-philosophy arguments in my discussions here. I don't know just how representative of all Bayesians that is.

For example, moral philosophy has been trashed by everyone who spoke to me about it so far. I get told its meaningless, or that Bayesian epistemology cannot create moral knowledge. No one has yet been like "oh my god, epistemology *should* be able to create moral and other philosophical (non-empirical, non-observational) knowledge! Bayesian stuff is wrong since it can't!" Rather, people don't seem to mind, and will argue at length that e.g. explanatory knowledge and non-empirical knowledge don't exist or are worthless and prediction is everything.

By "philosophy" I mean things which can't be experimentally/empirically tested (as opposed to "science" by which I mean things that can be). So for philosophy, no observations are directly relevant.

Make sense? Where do you stand on these issues?

And the way I think Popperian goals are better is that they value explanations which help us understand the world instead of being instrumentalist, positivist, anti-philosophical, or anything like that.

[continue this thread »](#)

endoself 07 April 2011 05:10:34AM 0 points



In order for any philosophy to be valid, the human brain must be able to evaluate deductive arguments; they are a huge component of philosophy, with many often being needed to argue a single idea. Wondering what to do in case these are wrong is not only unnecessary but impossible.

curi 07 April 2011 05:22:01AM 2 points



I don't have any criticism of deductive logic itself. But I do have criticisms of some of the premises i expect you to use. For example, they won't all be deductively argued for themselves. That raises the problem of: how will you sort out good ideas from bad ideas for use as premises? That gets into various proposed solutions to that problem, such as induction or Popperian epistemology. But if you get into that, right in the premises of your

supposed proof, then it won't be much of a proof because so much substantive content in the premises will be non-deductive.

[continue this thread »](#)



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 08 April 2011 09:18:03AM -7 points

http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf

Bayesian epistemology therefore complements traditional epistemology; it does not replace it or aim at replacing it.

Since Popper refuted traditional epistemology (source: his books, and the failure of anyone to come up with any good criticisms of his main ideas), and Bayesian Epistemology retains it, then Bayesian Epistemology is refuted too. And discussing this issue can be done without mentioning probability, Bayes' theorem, or Solomonoff induction. Bringing those up cannot be a relevant defense since traditional epistemology, which doesn't use them, is retained.

Bayesian epistemology is, in the first place, a philosophical project, and that it is its ambition to further progress in philosophy.

Why are most Less Wrong people anti-philosophy then? There's so much instrumentalism, empiricism, reductionism and borderline postivism. Not much interest in philosophy.

Section 2 introduces the probability calculus and explains why degrees of belief obey the probability calculus. Section 3 applies the formal machinery to an analysis of the notion of evidence, and high- lights potential application. Section 4 discusses Bayesian models of coherence and testimony, and section 5 ends this essay with a comparison of traditional epistemology and Bayesian epistemology.

Sections 2-4 are irrelevant. They are already assuming mistakes from traditional epistemology. Moving on to 5, which is only one page.

Bayesian epistemology, on the other hand, draws much of its power from the mathematical machinery of probability theory. It starts with a mathematical intuition.

Advocating intuitionism is very silly.

traditional epistemology inspires Bayesian accounts.

So Bayesians should care about criticisms of traditional epistemology, and be willing to engage with them directly without even mentioning any Bayesian stuff.

Both Bayesian epistemology and traditional epistemology do not much consider empirical data. Both are based on intuitions,

That's not even close to what most Less Wrong people told me. They mostly are very focussed on empirical data.

This might be a problem as privilege is given to the philosopher's intuitions.

Custom Search

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

- NEAREST MEETUPS**
- Denver Area LW November Meetup: 07 November 2017 08:00PM
 - Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

- RECENT COMMENTS**
- [Yeah, that's a bug. I am planning to](#) by Habryka on LW 2.0 Open Beta Live | 0 points
 - [Woo! Also if anyone else gets a](#) by korin43 on LW 2.0 Open Beta Live | 0 points
 - [Running through this to check that](#) by Multipartite on Sleeping Beauty gets counterfactually mugged | 0 points
 - [No, you don't need update you](#) by trickster on What Evidence Filtered Evidence? | 0 points
 - [I think that the core of religion—that](#) by adjuant on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

Might be? lol... What a hedge. They know it's a problem and equivocate.

non-philosophers may have different intuitions.

Also Popperian philosophers, and all other types that don't agree with you.

While it is debatable how serious these intuitions should be taken (maybe people are simply wrong!)

But not traditional philosophers, who have reliable intuitions? This is just plain silly.

It is therefore advisable that philosophers also keep on paying attention to other formal frameworks

But not informal frameworks, because your intuition says that formality is next to Godliness?

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all' by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- gwern (32)
- ingres (30)
- Lumifer (27)
- turchin (23)
- Habryka (22)
- Yosarian2 (21)
- IlyaShpitser (19)
- SaidAchmiz (19)
- gjm (19)
- fowlertm (17)
- SquirrelInHell (15)
- Dagon (15)
- Elo (14)
- Stuart_Armstrong (13)
- Gunnar_Zarncke (12)

RECENT KARMA AWARDS

[deleted] 09 April 2011 08:46:09AM -4 points

Good criticisms here, yet downvoted to -3. Do LWer's really want to be less wrong?

JoshuaZ 09 April 2011 04:20:48PM * 5 points

Good criticisms here, yet downvoted to -3. Do LWer's really want to be less wrong?

There's a general pattern here. Some of this comments are potentially good. But the general pattern either a) misses points b) doesn't actually grapple with what he is claiming it does and c) is uncivil. C is a major issue. Obnoxious remarks like "But not informal frameworks, because your intuition says that formality is next to Godliness?" are going to get downvoted.

People are inclined to downvote uncivil comments for a variety of reasons: 1) They reinforce emotionalism on all members of a discussion, making the actual exchange of ideas less likely. 2) They make the individual making the comments much less likely to acknowledge when they are wrong (this is due to standard cognitive biases). 3) They make communities less pleasant.

Uncivil comments that support common beliefs are also voted down. LWians are not perfect and you shouldn't be surprised if that is going to occur even more with comments that are uncivil and go against the consensus. In this particular case, it also doesn't help that most of the uncivility is at the end of the comment, so one moves directly from reading the unproductive, uncivil remarks to seeing the vote button.

[deleted] 08 April 2011 07:33:54PM * 0 points

I would recommend turning this into a discussion-level post--I doubt anyone will find this comment, as it's buried pretty deeply in this discussion.

curi 08 April 2011 07:35:36PM * 2 points

Do you think people other than you will like it? I think many will complain about the style and i didn't want to rewrite it more formally. also i dismissed most of the paper as irrelevant. i expect people to complain about that and don't particularly expect the discussion to go anywhere.

[deleted] 08 April 2011 09:03:27PM * 0 points

You'd probably have to change the style, yes. And no, I don't expect other people to like it, but I expect that they will respond. Also: you're probably going to have to either go into more depth or pick a specific example, or both.

curi 08 April 2011 09:48:47PM -3 points

What's in it for me? I think I got the gist of what less wrong has to offer already.

Desrtopa 08 April 2011 10:25:39PM 2 points

Can you explain how we believe that Bayesianism leads to better decisionmaking? I'm not even asking you to *do* it, I no longer have high expectations of productivity from this conversation and don't intend to prolong it, but know that if you can't, you don't understand what we offer at all.



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 08 April 2011 09:47:58AM -7 points

Does one mean for example to claim that pure mathematics works off of criticism?

yes

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

JoshuaZ 08 April 2011 01:45:36PM 4 points

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

You are confused about what that means. An appeal to authority is not intrinsically fallacious. An appeal to authority is problematic when the authority is irrelevant (e.g. a celebrity who plays a doctor on TV endorsing a product) or when one is claiming that one has a valid deduction in some logical system. Someone making an observation about what people in their profession actually do is not a bad appeal to authority in the same way. In any event, you ignored the next line of my comment:

Moreover, it isn't clear what it would even mean for us to try to do this as our primary method of inquiry. Are we supposed to spend all our time going through pre-existing proofs trying to find holes in them?

If you do think that mathematicians use Popperian reasoning then please explain how we do it.

curi 08 April 2011 04:49:03PM -2 points

An appeal to authority is not intrinsically fallacious.

It is in Popperian epistemology.

Could you point me to a Bayesian source that says they are OK? I'd love to have a quote of Yudkowsky advocating appeals to authority, for instance. Or could others comment? Do most people here think appeals to authority are good arguments?

Marius 08 April 2011 05:05:13PM * 2 points

An appeal to authority is not logically airtight, and if logic is about mathematical proofs, then it's going to be

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

a fallacy. But an appeal to an appropriate authority gives Bayesians strong evidence, provided that [X|Authority believes X] is sufficiently high. In many fields, authorities have sufficient track records that appeals to authority are good arguments. In other fields, not so much.

Of course, the Appeal to Insufficient Force fallacy is a different story from the Appeal to Inappropriate Authority

curi 08 April 2011 05:10:23PM * -5 points [-]

How do you judge:

[X|Authority believes X]

In general I judge it very low. Certainly in this case.

Can you provide a link to Yudkowsky or any well known Bayesian advocating appeals to authority?

Marius 08 April 2011 05:23:40PM * 4 points [-]

How do you judge: [X|Authority believes X]

Track record of statements/predictions, taking into account the prior likelihood of previous predictions and prior likelihood of current prediction.

Can you provide a link to Yudkowsky or any well known Bayesian advocating appeals to authority?

Are you asking us to justify appeals to authority by using an appeal to authority?

edit per wedrifid

wedrifid 08 April 2011 06:07:33PM 4 points [-]

Track record of statements/predictions, taking into account a priori likelihood of previous predictions and a priori likelihood of current prediction.

I would have said 'prior', not 'a priori'.

curi 08 April 2011 05:24:50PM -4 points [-]

Are you asking us to justify appeals to authority by using an appeal to authority?

No lol. I just wanted one to read. Some of my friends will be interested in it too.

Track record of statements/predictions

Since the guy who made the appeal to authority has little track record with me, and little of it good in my view, why would he expect me to concede to his appeal to authority?

benelliott 08 April 2011 05:22:42PM 1 point [-]

Robin Hanson does so [here](#).

curi 08 April 2011 05:30:31PM -5 points [-]

Too much ambiguity there. e.g. the word authority isn't used.

benelliott 08 April 2011 07:36:00PM 2 points [-]

This is silly. Whether or not he uses the word authority does not change the fact he is

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (32)

ingres (30)

Lumifer (27)

turchin (23)

Habryka (22)

Yosarian2 (21)

IlyaShpitser (19)

SaidAchmiz (19)

gjm (19)

fowlertm (17)

SquirrelInHell (15)

Dagon (15)

Elo (14)

Stuart_Armstrong (13)

Gunnar_Zarncke (12)

RECENT KARMA AWARDS

suggesting that we treat the opinions of experts as more accurate than our own opinions.

I had a lot of respect for you before you made this comment, but you have now lost most of it.

curi 08 April 2011 07:40:33PM * 1 point

The idea that appeals to authority are good arguments is not identical to the idea that the opinions of experts are more accurate. Suppose they are more accurate, on average. Does that make appealing to one a good argument? I don't think so and my friends won't. They won't know if Hanson thinks so.

For the purposes I wanted to use it for, this will not work well.

One thing I know about some of my friends is that they consider the word "authority" to be very nasty, but the word "expert" to be OK. They specifically differentiate between expertise (a legitimate concept) and authority (an illegitimate concept). Hanson's use of the expertise terminology, instead of the authority terminology, will matter to them. Explaining that he meant what they call authority will add complexity -- and scope for argument -- and be distracting. And people will find it boring and ignore it as a terminological debate.

And I'm not even quite sure what Hanson did mean. I don't think what he meant is identical to what the commenter I was speaking to meant.

Hanson speaks of, for example, "if you plan to mostly ignore the experts". That you shouldn't *ignore* them is a different claim than that appeals to their authority are good arguments.

JoshuaZ 09 April 2011 03:12:47PM 1 point

The idea that appeals to authority are good arguments is not identical to the idea that the opinions of experts are more accurate. Suppose they are more accurate, on average. Does that make appealing to one a good argument?

What do you mean by good argument? The Bayesians have an answer to this. They mean that $P(\text{claim}|\text{argument}) > P(\text{claim})$. Now, one might argue in that framework that if $P(\text{claim}|\text{argument})/P(\text{claim})$ is close to 1 then this isn't a good argument, or if $\log P(\text{claim}|\text{argument})/P(\text{claim})$ is small compared to the effort to present and evaluate the argument then it isn't a good argument.

However, that's obviously not what you mean. It isn't clear to me what you mean by "good argument" and how this connects to the notion of a fallacy. Please expand your definitions or taboo the terms.

benelliott 08 April 2011 09:32:16PM 2 points

He's stated before, I'm not sure where, that if you believe an expert has more knowledge about an issue than you then you should prefer their opinions to any argument you generate. This is because if they disagree with you it is almost certainly because they have considered and rejected your argument, not because they have not considered your argument.

One thing I know about some of my friends is that they consider the word "authority" to be very nasty, but the word "expert" to be OK. They specifically differentiate between expertise (a legitimate concept) and authority (an illegitimate concept). Hanson's use of the expertise terminology, instead of the authority terminology, will matter to them.

If your friends cannot differentiate between the content of an argument and its surface appearance then I would advise you find new friends [facetious].

[continue this thread »](#)



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 08 April 2011 09:47:58AM -7 points

Does one mean for example to claim that pure mathematics works off of criticism?

yes

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

JoshuaZ 08 April 2011 01:45:36PM 4 points

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

You are confused about what that means. An appeal to authority is not intrinsically fallacious. An appeal to authority is problematic when the authority is irrelevant (e.g. a celebrity who plays a doctor on TV endorsing a product) or when one is claiming that one has a valid deduction in some logical system. Someone making an observation about what people in their profession actually do is not a bad appeal to authority in the same way. In any event, you ignored the next line of my comment:

Moreover, it isn't clear what it would even mean for us to try to do this as our primary method of inquiry. Are we supposed to spend all our time going through pre-existing proofs trying to find holes in them?

If you do think that mathematicians use Popperian reasoning then please explain how we do it.

curi 08 April 2011 04:49:03PM -2 points

An appeal to authority is not intrinsically fallacious.

It is in Popperian epistemology.

Could you point me to a Bayesian source that says they are OK? I'd love to have a quote of Yudkowsky advocating appeals to authority, for instance. Or could others comment? Do most people here think appeals to authority are good arguments?

Marius 08 April 2011 05:05:13PM * 2 points

An appeal to authority is not logically airtight, and if logic is about mathematical proofs, then it's going to be

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

a fallacy. But an appeal to an appropriate authority gives Bayesians strong evidence, provided that [X|Authority believes X] is sufficiently high. In many fields, authorities have sufficient track records that appeals to authority are good arguments. In other fields, not so much.

Of course, the Appeal to Insufficient Force fallacy is a different story from the Appeal to Inappropriate Authority

curi 08 April 2011 05:10:23PM * -5 points [-]

How do you judge:

[X|Authority believes X]

In general I judge it very low. Certainly in this case.

Can you provide a link to Yudkowsky or any well known Bayesian advocating appeals to authority?

Marius 08 April 2011 05:23:40PM * 4 points [-]

How do you judge: [X|Authority believes X]

Track record of statements/predictions, taking into account the prior likelihood of previous predictions and prior likelihood of current prediction.

Can you provide a link to Yudkowsky or any well known Bayesian advocating appeals to authority?

Are you asking us to justify appeals to authority by using an appeal to authority?

edit per wedrifid

wedrifid 08 April 2011 06:07:33PM 4 points [-]

Track record of statements/predictions, taking into account a priori likelihood of previous predictions and a priori likelihood of current prediction.

I would have said 'prior', not 'a priori'.

curi 08 April 2011 05:24:50PM -4 points [-]

Are you asking us to justify appeals to authority by using an appeal to authority?

No lol. I just wanted one to read. Some of my friends will be interested in it too.

Track record of statements/predictions

Since the guy who made the appeal to authority has little track record with me, and little of it good in my view, why would he expect me to concede to his appeal to authority?

benelliott 08 April 2011 05:22:42PM 1 point [-]

Robin Hanson does so [here](#).

curi 08 April 2011 05:30:31PM -5 points [-]

Too much ambiguity there. e.g. the word authority isn't used.

benelliott 08 April 2011 07:36:00PM 2 points [-]

This is silly. Whether or not he uses the word authority does not change the fact he is

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (32)

ingres (30)

Lumifer (27)

turchin (23)

Habryka (22)

Yosarian2 (21)

IlyaShpitser (19)

SaidAchmiz (19)

gjm (19)

fowlertm (17)

SquirrelInHell (15)

Dagon (15)

Elo (14)

Stuart_Armstrong (13)

Gunnar_Zarncke (12)

RECENT KARMA AWARDS

suggesting that we treat the opinions of experts as more accurate than our own opinions.

I had a lot of respect for you before you made this comment, but you have now lost most of it.

curi 08 April 2011 07:40:33PM * 1 point

The idea that appeals to authority are good arguments is not identical to the idea that the opinions of experts are more accurate. Suppose they are more accurate, on average. Does that make appealing to one a good argument? I don't think so and my friends won't. They won't know if Hanson thinks so.

For the purposes I wanted to use it for, this will not work well.

One thing I know about some of my friends is that they consider the word "authority" to be very nasty, but the word "expert" to be OK. They specifically differentiate between expertise (a legitimate concept) and authority (an illegitimate concept). Hanson's use of the expertise terminology, instead of the authority terminology, will matter to them. Explaining that he meant what they call authority will add complexity -- and scope for argument -- and be distracting. And people will find it boring and ignore it as a terminological debate.

And I'm not even quite sure what Hanson did mean. I don't think what he meant is identical to what the commenter I was speaking to meant.

Hanson speaks of, for example, "if you plan to mostly ignore the experts". That you shouldn't *ignore* them is a different claim than that appeals to their authority are good arguments.

JoshuaZ 09 April 2011 03:12:47PM 1 point

The idea that appeals to authority are good arguments is not identical to the idea that the opinions of experts are more accurate. Suppose they are more accurate, on average. Does that make appealing to one a good argument?

What do you mean by good argument? The Bayesians have an answer to this. They mean that $P(\text{claim}|\text{argument}) > P(\text{claim})$. Now, one might argue in that framework that if $P(\text{claim}|\text{argument})/P(\text{claim})$ is close to 1 then this isn't a good argument, or if $\log P(\text{claim}|\text{argument})/P(\text{claim})$ is small compared to the effort to present and evaluate the argument then it isn't a good argument.

However, that's obviously not what you mean. It isn't clear to me what you mean by "good argument" and how this connects to the notion of a fallacy. Please expand your definitions or taboo the terms.

benelliott 08 April 2011 09:32:16PM 2 points

He's stated before, I'm not sure where, that if you believe an expert has more knowledge about an issue than you then you should prefer their opinions to any argument you generate. This is because if they disagree with you it is almost certainly because they have considered and rejected your argument, not because they have not considered your argument.

One thing I know about some of my friends is that they consider the word "authority" to be very nasty, but the word "expert" to be OK. They specifically differentiate between expertise (a legitimate concept) and authority (an illegitimate concept). Hanson's use of the expertise terminology, instead of the authority terminology, will matter to them.

If your friends cannot differentiate between the content of an argument and its surface appearance then I would advise you find new friends [facetious].

[continue this thread »](#)



MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

Bayesian Epistemology vs Popper

-1 curi 06 April 2011 11:50PM

I was directed to this book (<http://www-biba.inrialpes.fr/Jaynes/prob.html>) in conversation here:

http://lesswrong.com/lw/3ox/bayesianism_versus_critical_rationalism/3ug7?context=1#3ug7

I was told it had a proof of Bayesian epistemology in the first two chapters. One of the things we were discussing is Popper's epistemology.

Here are those chapters:

<http://www-biba.inrialpes.fr/Jaynes/cc01p.pdf>

<http://www-biba.inrialpes.fr/Jaynes/cc02m.pdf>

I have not found any proof here that Bayesian epistemology is correct. There is not even an attempt to prove it. Various things are assumed in the first chapter. In the second chapter, some things are proven given those assumptions.

Some first chapter assumptions are incorrect or unargued. It begins with an example with a policeman, and says his conclusion is not a logical deduction because the evidence is logically consistent with his conclusion being false. I agree so far. Next it says "we will grant that it had a certain degree of validity". But I will not grant that. Popper's epistemology explains that *this is a mistake* (and Jaynes makes no attempt at all to address Popper's arguments). In any case, simply assuming his readers will grant his substantive claims is no way to argue.

The next sentences blithely assert that we all reason in this way. Jaynes' is basically presenting the issues of this kind of reasoning as his topic. This simply ignores Popper and makes no attempt to prove Jaynes' approach is correct.

Jaynes goes on to give syllogisms, which he calls "weaker" than deduction, which he acknowledges are not deductively correct. And then he just says we use that kind of reasoning all the time. That sort of assertion only appeals to the already converted. Jaynes starts with arguments which appeal to the "intuition" of his readers, not on arguments which could persuade someone who disagreed with him (that is, good rational arguments). Later when he gets into more mathematical stuff which doesn't (directly) rest on appeals to intuition, it does rest on the ideas he (supposedly) established early on with his appeals to intuition.

The outline of the approach here is to quickly gloss over substantive philosophical assumptions, never provide serious arguments for them, take them as common sense, do not detail them, and then later provide arguments which are rigorous *given the assumptions glossed over earlier*. This is a mistake.

So we get, e.g., a section on Boolean Algebra which says it will state previous ideas more formally. This briefly acknowledges that the rigorous parts depend on the non-rigorous parts. Also the very important problem of carefully detailing how the mathematical objects discussed correspond to the real world things they are supposed to help us understand does not receive adequate attention.

Chapter 2 begins by saying we've now formulated our problem and the rest is just math. What I take from that is that the early assumptions won't be revisited but simply used as premises. So the rest is pointless if those early assumptions are mistaken, and Bayesian Epistemology cannot be proven in this way to anyone who doesn't grant the assumptions (such as a Popperian).

Moving on to Popper, Jaynes is ignorant of the topic and unscholarly. He writes:

<http://www-biba.inrialpes.fr/Jaynes/crefsy.pdf>

Custom Search

curi
 Karma Score
 12
 2

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
[Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

> Karl Popper is famous mostly through making a career out of the doctrine that theories may not be proved true, only false

This is pure fiction. Popper is a fallibilist and said (repeatedly) that theories cannot be proved false (or anything else).

It's important to criticize unscholarly books promoting myths about rival philosophers rather than addressing their actual arguments. That's a major flaw not just in a particular paragraph but in the author's way of thinking. It's especially relevant in this case since the author of the books tries to tell us about how to think.

Note that Yudkowsky made a similar unscholarly mistake, about the same rival philosopher, here:

<http://yudkowsky.net/rational/bayes>

> Previously, the most popular philosophy of science was probably Karl Popper's falsificationism - this is the old philosophy that the Bayesian revolution is currently dethroning. Karl Popper's idea that theories can be definitely falsified, but never definitely confirmed

Popper's philosophy is not falsificationism, it was never the most popular, and it is fallibilist: it says ideas cannot be definitely falsified. It's bad to make this kind of mistake about what a rival's basic claims are when claiming to be dethroning him. The correct method of dethroning a rival philosophy involves understanding what it does say and criticizing that.

If Bayesians wish to challenge Popper they should learn his ideas and address his arguments. For example he questioned the concept of positive support for ideas. Part of this argument involves asking the questions: "What is support?" (This is not asking for its essential nature or a perfect definition, just to explain clearly and precisely what the support idea actually says) and "What is the difference between "X supports Y" and "X is consistent with Y"?" If anyone has the answer, please tell me.

Article Navigation

Comments (226)

Tags: epistemology | popper | jaynes | Delete

Comments (226)

Sort By: New

Enter a comment here

Comment

Show help

Peterdjones 18 July 2011 12:03:38AM * 2 points [-]
If anyone can bear more of this, Poppers argument against induction using Bayes is being discussed [here](#)

Peterdjones 12 April 2011 08:23:30PM * 2 points [-]
""What is support?" (This is not asking for its essential nature or a perfect definition, just to explain clearly and precisely what the support idea actually says) and "What is the difference between "X supports Y" and "X is consistent with Y"?" If anyone has the answer, please tell me."
Bayesians appear to have answers to these questions. Moreovoer, far from wishing to refute Popper, they can actually incorporate a fomr of Popperianism.
"On the other hand, Popper's idea that there is only falsification and no such thing as confirmation turns out to be incorrect. Bayes' Theorem shows that falsification is very strong evidence compared to confirmation, but falsification is still probabilistic in nature; it is not governed by fundamentally different rules from confirmation, as Popper argued."
But of course Popper was a fallibilist as well as a falsificationist, so his falsifications aren't absolute and certain anyway. Bayes just brings out that where you don't have absolute falsification, you can't have absolute lack of positive support. Falsification of T has to support not-T. But the support gets spread thinly...

Peterdjones 12 April 2011 08:11:16PM 2 points [-]

LW 2.0 Open Beta Live
by Vaniver | 21v (33c)

In support of yak shaving part 2
by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag
by Raemon | 41v (106c)

European Community Weekend 2017
by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about "protecting" the epistemology
by AnnaSalamon | 73v (53c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"
by AnnaSalamon | 36v (43c)

Be secretly wrong
by Benquo | 32v (47c)

CFAR's new focus, and AI Safety
by AnnaSalamon | 30v (88c)

Fact Posts: How and Why
by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement
by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- gwern (32)
ingres (30)
Lumifer (27)
turchin (23)
Habryka (22)
Yosarian2 (21)
IlyaShpitser (19)
SaidAchmiz (19)
gjm (19)
fowlertm (17)
SquirrelInHell (15)
Dagon (15)
Elo (14)
Stuart_Armstrong (13)
Gunnar_Zarncke (12)

RECENT KARMA AWARDS

Curi,

"Some first chapter assumptions are incorrect or unargued. It begins with an example with a policeman, and says his conclusion is not a logical deduction because the evidence is logically consistent with his conclusion being false."

Popper's epistemology doesn't explain that the conclusion of the argument has no validity, in the sense of being certainly false. In fact, it requires that the conclusion is not certainly false. No conjecture is certainly false.

Perhaps you meant he shows that the argument is invalid in the sense of being a non sequitur. (A non sequitur can still have a plausible or true conclusion). Of course it is not valid in the sense of traditional, necessitarian deduction. The whole point is that it is something different. And the argument that this non-traditional, plausibility based deduction works is just the informal observation that we use it all the time and it seems to work. What else could it be? If were valid by traditional deduction it would BE traditional deduction.

" Later when he gets into more mathematical stuff which doesn't (directly) rest on appeals to intuition, it does rest on the ideas he (supposedly) established early on with his appeals to intuition."

The Popperian argument against probabilistic reasoning is that it can't be shown how it works. If Jaynes maths shows how it works, that objection is removed.

"This is pure fiction. Popper is a fallibilist and said (repeatedly) that theories cannot be proved false (or anything else)."

Of course he has to believe in some FAPP refutation. or he ends up saying nothing at all.

PeterdJones 14 April 2011 02:27:20PM 3 points



"Science, philosophy and rational thought must all start from common sense". KRP, Objective Knowledge, p33.

Starting with common sense is exactly what Jaynes is doing. (Popper says that what is important is not to take common sense as irrefutable).

prase 07 April 2011 01:30:14PM 19 points



I have skimmed through the comments here and smelled a weak odour of a flame war. Well, the discussion is still rather civil and far from a flame war as understood on most internet forums, but it somehow doesn't fit well within what I am used to see here on LW.

The main problem I have is that you (i.e. curi) have repeatedly asserted that the Bayesians, including most of LW users, don't understand Popperianism and that Bayesianism is in fact worse, without properly explaining your position. It is entirely possible, even probable, that most people here don't actually get all subtleties of Popper's worldview. But then, a better strategy may be to first write a post which explains these subtleties and tells why they are important. On the other hand, you don't need to tell us explicitly "you are unscholarly and misinterpret Popper". If you actually explain what you ought to (and if you are right about the issue), people here will likely understand that they were previously wrong, and they will do it without feeling that you seek confrontation rather than truth - which I mildly have.

curi 07 April 2011 07:46:41PM 0 points



By "properly explaining my position" I'm not sure what you want. Properly understanding it takes reading, say, 20 books (plus asking questions about them as you go, and having critical discussions about them, and so on). If I summarize, lots of precision is lost. I have tried to summarize.

I can't write "a (one) post" that explains the subtleties of Popper. It took Popper a career and many books.

Bayesianism has a regress/foundations problem. Yudkowsky acknowledges that. Popperism doesn't. So Popperism is better in a pretty straightforward way.

On the other hand, you don't need to tell us explicitly "you are unscholarly and misinterpret Popper".

But they were propagating myths about Popper. They were unscholarly. They didn't know wtf they were talking about, not even the basics. Basically *all* of Popper's books contradict those myths. It's really not cool to *attribute*

positions to someone he never advocated. This mistake is easy to avoid by the method: *don't publish about people you haven't read.* Bad scholarship is a big deal, IMO.

Desrtopa 08 April 2011 09:34:47PM * 4 points

Bayesianism has a regress/foundations problem. Yudkowsky acknowledges that. Popperism doesn't. So Popperism is better in a pretty straightforward way.

Any system with axioms can be infinitely regressed or rendered circular if you demand that it justify the axioms. Critical Rationalism has axioms, and can be infinitely regressed.

You were upvoted in the beginning for pointing out gaps in scholarship and raising ideas not in common circulation here. You yourself, however, have demonstrated a clear lack of understanding of Bayesianism, and have attracted frustration with your *own* lack of scholarship and confused arguments, along with failure to provide good reasons for us to be *interested* in the prospect of doing this large amount of reading you insist is necessary to properly understand Popper. If doing this reading were worthwhile, we would expect you to be able to give a better demonstration of why.

curi 08 April 2011 09:44:47PM -5 points

Any system with axioms can be infinitely regressed or rendered circular if you demand that it justify the axioms. Critical Rationalism has axioms, and can be infinitely regressed.

You haven't understood the basic point that this only works *if you accept that ideas should be justify.*

If you reject the demand for justification -- as CR does -- then then regress is gone. Hence no regress in CR.

Peterdjones 15 April 2011 03:04:54PM 1 point

How do you know which forms of criticism are valid? Do you justify them, or attempt to criticise them? Either way looks regressive to me.

curi 15 April 2011 05:44:19PM * -4 points

Tell me how you found me on Less Wrong or I'm definitely not answering any of your questions.

Desrtopa 08 April 2011 09:58:13PM * 0 points

I reject this position as vacuous.

The position might be self consistent if you accept its premises, and one of its premises may be that you can introduce any idea without justification, but it's not *reality consistent*.

curi 08 April 2011 10:00:50PM -5 points

b/c?

Desrtopa 08 April 2011 10:19:35PM * 3 points

It's no less founded on unproven axioms, and it's no less arbitrary than induction, it just contains the tenet "we don't reject propositions for being arbitrary" and pats itself on the back for being self consistent. This doesn't do a better job helping people generate true information, it does a worse one.

[deleted] 09 April 2011 07:42:37AM -1 points

Being arbitrary is a criticism, so a critical rationalist can and does reject propositions for being

arbitrary. Rejecting the idea of justification does not mean accepting any old arbitrary thing. If your idea doesn't stand up to criticism, including criticism that it is just arbitrary, then it is gone.

prase 07 April 2011 08:51:35PM 2 points

I have tried to summarize.

I acknowledge that, although I would have preferred if you did that before you have written this post.

I can't write "a (one) post" that explains the subtitles of Popper. It took Popper a career and many books.

Could be five posts.

Even if such a defense can be sometimes valid, it is too often used to defend confused positions (think about theology) to be much credible.

curi 07 April 2011 08:52:52PM -1 points

It would need to be 500 posts.

But anyway, they are written and published. By Popper not me. They already exist and they don't need to be published on this particular website.

prase 07 April 2011 08:57:01PM 3 points

Following your advice expressed elsewhere, isn't the fact that the basics of Popperianism cannot be explained in five posts a valid criticism of Popperianism, which should be therefore rejected?

curi 07 April 2011 08:59:26PM 1 point

Why is that a criticism? What's wrong with that?

Also maybe it could be. But I don't know how.

And the *basics* could be explained quickly, to someone who didn't have a bunch of anti-Popperian biases, but people do have those b/c they are built into our culture. And without the details and precision then people complain about 1) not understanding how to do it, what it says 2) it not having enough precision and rigor

prase 07 April 2011 09:06:08PM * 2 points

Why is that a criticism?

Actually I don't know what constitutes a criticism in your book (since you never specified), but you have also said that there are no rules for criticism, so I suppose that it is a criticism. If not, then please say why it is *not* a criticism.

I am not going to engage in a discussion about my and your biases, since such debates rarely lead to an agreement.

curi 07 April 2011 09:11:10PM * 0 points

You can conjecture standards of criticism, or use the ones from your culture. If you find a problem with them, you can change them or conjecture different ones.

For many purposes I'm pretty happy with common sense notions of standards of criticism, which I think you understand, but which are hard to explain in words. If you have a *relevant* problem with the, you can say it.

[deleted] 07 April 2011 08:55:41PM 4 points 

One thing you could do is write a post highlighting a specific example where Bayes is wrong and Popper is right. A lot of people have asked for specific examples in this thread; if you could give a detailed discussion of one, that would move the discussion to more fertile ground.

curi 07 April 2011 08:57:01PM * 1 point 

Can you give me a link to a canonical essay on Bayesian *epistemology/philosophy*, and I'll pick from there?

Induction and justificationism are examples but I've been talking about them. I think you want something else. Not entirely sure what.

[deleted] 07 April 2011 09:04:46PM 1 point 

It's not at all canonical, but a paper that neatly summarizes Bayesian epistemology is "Bayesian Epistemology" by Stephan Hartmann and Jan Sprenger.

curi 07 April 2011 09:09:44PM 1 point 

Found it.

http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf

Will take a look in a bit.

[deleted] 07 April 2011 09:14:16PM 1 point 

Excellent, thanks.

curi 08 April 2011 09:18:03AM -7 points 

http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf

Bayesian epistemology therefore complements traditional epistemology; it does not replace it or aim at replacing it.

Since Popper refuted traditional epistemology (source: his books, and the failure of anyone to come up with any good criticisms of his main ideas), and Bayesian Epistemology retains it, then Bayesian Epistemology is refuted too. And discussing this issue can be done without mentioning probability, Bayes' theorem, or Solomonoff induction. Bringing those up cannot be a relevant defense since traditional epistemology, which doesn't use them, is retained.

Bayesian epistemology is, in the first place, a philosophical project, and that it is its ambition to further progress in philosophy.

Why are most Less Wrong people anti-philosophy then? There's so much instrumentalism, empiricism, reductionism and borderline postivism. Not much interest in philosophy.

Section 2 introduces the probability calculus and explains why degrees of belief obey the probability calculus. Section 3 applies the formal machinery to an analysis of the notion of evidence, and highlights potential application. Section 4 discusses Bayesian models of coherence and testimony, and section 5 ends this essay with a comparison of traditional epistemology and Bayesian epistemology.

Sections 2-4 are irrelevant. They are already assuming mistakes from traditional

epistemology. Moving on to 5, which is only one page.

Bayesian epistemology, on the other hand, draws much of its power from the mathematical machinery of probability theory. It starts with a mathematical intuition.

Advocating intuitionism is very silly.

traditional epistemology inspires Bayesian accounts.

So Bayesians should care about criticisms of traditional epistemology, and be willing to engage with them directly without even mentioning any Bayesian stuff.

Both Bayesian epistemology and traditional epistemology do not much consider empirical data. Both are based on intuitions,

That's not even close to what most Less Wrong people told me. They mostly are very focussed on empirical data.

This might be a problem as privilege is given to the philosopher's intuitions.

Might be? lol... What a hedge. They know it's a problem and equivocate.

non-philosophers may have different intuitions.

Also Popperian philosophers, and all other types that don't agree with you.

While it is debatable how serious these intuitions should be taken (maybe people are simply wrong!)

But not traditional philosophers, who have reliable intuitions? This is just plain silly.

It is therefore advisable that philosophers also keep on paying attention to other formal frameworks

But not informal frameworks, because your intuition says that formality is next to Godliness?

[continue this thread »](#)

Desrtopa 07 April 2011 02:49:52PM * 3 points 

Upvoted and agreed. I feel at this point like further addressing the discussion on present terms would be simply irresponsible, more likely to become adversarial than productive. If curi wrote up such a post, it would hopefully give a meaningful place to continue from.

Edit: It seems that curi has [created such a post](#). I'm not entirely convinced that continuing the discussion is a good idea, but perhaps it's worth humoring the effort.

TheOtherDave 07 April 2011 02:24:22PM 2 points 

For what it's worth, I have that feeling more than mildly and consequently stopped paying attention to the curi-exchange a while ago. Too much heat, not enough light.

I've been considering downvoting the whole thread on the grounds that I want less of it, but haven't yet, roughly on the grounds that I consider it irresponsible to do so without paying more careful attention to it and don't currently consider it worth paying more attention to.

David_Gerard 07 April 2011 12:15:23PM * 2 points 

It has occurred to me before that the lack of a proper explanation on LessWrong of Bayesian *epistemology* (and not just saying "Here's Bayes' theorem and how it works, with a neat Java applet") is a serious lack. I've been

reduced to linking the Stanford Encyclopedia of Philosophy article, which is really not well written at all.

It is also clear from the comments on this post that people are talking about it without citable sources, and are downvoting as a mark of disagreement rather than anything else. This is bad as it directly discourages thought or engagement on the topic from those trying to disagree in good faith, as curi is here.

Is there a decent explanation of Bayesian epistemology per se (not the theorem, the epistemology) that doesn't start by talking about Popper or something else, that the Bayesian epistemology advocates here could link to? This would lead to a much more productive discussion, as everyone might at least start on approximately the same page.

benelliott 07 April 2011 12:40:43PM * 0 points

I don't know if these are what you're looking for but:

[Probability Theory: The Logic of Science](#) by Jaynes, spends its first chapter explaining why we need a 'calculus of plausibility' and what such a calculus should hope to achieve. The rest of the book is mostly about setting it up and showing what it can do. (The link does not contain the whole book, only the first few chapters, you may need to buy or borrow it to get the rest).

[Yudkowsky's Technical explanation](#), which assumes the reader is already familiar with the theorem, explains some of its implications for scientific thinking in general.

David_Gerard 07 April 2011 01:29:13PM 1 point

See [here](#) for what I see the absence of. There's a hole that needs filling here.

benelliott 07 April 2011 06:29:42AM * 5 points

I gave a description of how a Bayesian sees the difference between "X supports Y" and "X is consistent with Y" in our previous discussion. I don't know if you saw it, you haven't responded to it and you aren't acting like you accepted it so I'll give it again here:

"X is consistent with Y" is not really a Bayesian way of putting things, I can see two ways of interpreting it. One is as $P(X \& Y) > 0$, meaning it is at least theoretically possible that both X and Y are true. The other is that $P(X|Y)$ is reasonably large, i.e. that X is plausible if we assume Y.

"X supports Y" means $P(Y|X) > P(Y)$, X supports Y if and only if Y becomes more plausible when we learn of X. Bayes tells us that this is equivalent to $P(X|Y) > P(X)$, i.e. if Y would suggest that X is more likely that we might think otherwise then X is support of Y.

Suppose we make X the statement "the first swan I see today is white" and Y the statement "all swans are white". $P(X|Y)$ is very close to 1, $P(X|\sim Y)$ is less than 1 so $P(X|Y) > P(X)$, so seeing a white swan offers support for the view that all swans are white. Very, very weak support, but support nonetheless.

For a Popperian definition, you guys are allowed to criticise something right? In that case could we say that support for a proposition is logically equivalent to a criticism of its negation?

The whole 'there is no positive support' thing seems like an overreaction to the whole Cartesian 'I can prove ideas with certainty thing'. I agree that certain support is a flawed concept, but you seem to be throwing the baby out with the bathwater by saying uncertain support is guilty by association and should be rejected as well.

Also, I'm a little incredulous here, do you really reject the policeman's syllogism? Would you say he is wrong to chase the man down the road? If you encountered such a person, would you genuinely treat them as you would treat anyone else?

curi 07 April 2011 07:07:21AM * 0 points

I missed your comment. I found it now. I will reply there.

http://lesswrong.com/lw/3ox/bayesianism_versus_critical_rationalism/3uld?context=1#3uld

could we say that support for a proposition is logically equivalent to a criticism of its negation?

No. The negation of a universal theory is not universal, and the negation of an explanatory theory is not explanatory. So, the interesting theories would still be criticism only, and the uninteresting ones (e.g. "there is a cat") support only. And the meaning of "support" is rather circumscribed there.

If you want to say theories of the type "the following explanation isn't true:" get "supported" it doesn't contribute anything useful to epistemology. the support idea, as it is normally conceived, is still wrong, and this rescues none of the substance.

The other issue is that criticism isn't the same kind of thing as support. It's not in the same category of concept.

Yes I really reject the policeman's syllogism. In the sense of: I don't think the argument in the book is any good. There are other arguments which are OK for reaching the conclusion (but which rely on things the book left unstated, e.g. background knowledge and context. Without adding anything at all, no cultural biases or assumptions or hidden claims, and even doing our best to not use the biases and assumptions built into the English language, then no there isn't any way to guess what's more likely).

Peterdjones 15 April 2011 03:08:22PM 1 point

If the Policeman's argument is only valid in the light of background assumptions, why would they need to be stated? Surely we would only need to make the same tacit assumptions to agree with the conclusions. Everyday reasoning differs from formal logic in various ways, and mainly because it takes short cuts. I don't think that invalidates it.

JGWeissman 07 April 2011 06:19:50AM 4 points

A huge strength of Bayesian epistemology is that it tells me how to [program computers to form accurate beliefs](#). Has Popperian epistemology guided the development of any computer program as awesome as Gmail's spam filter?

curi 07 April 2011 06:59:03AM -2 points

Bayesian epistemology didn't do that. Bayes' theorem did. See the difference?

JGWeissman 07 April 2011 04:40:10PM 3 points

Bayesian epistemology didn't do that. Bayes' theorem did.

Bayes' theorem is part of probability theory. Bayesian epistemology essentially says to take probability theory seriously as a normative description of degrees of belief.

If you don't buy that and really want to split the hair, then I am willing to modify my question to: Has the math behind Popperian epistemology guided the development of any computer program as awesome as Gmail's spam filter? (Is there math behind Popperian epistemology?)

curi 07 April 2011 05:59:18PM -1 points

gmail's spam filter does not have degrees of belief or belief.

It has things which you could call by those words if you really wanted to. But it wouldn't make them into the same things those words mean when referring to people.

JGWeissman 07 April 2011 06:14:51PM 4 points

But it wouldn't make them into the same things those words mean when referring to people.

I want the program to find the correct belief, and then take good actions based on that correct belief. I don't care if lacks the conscious experience of believing.

You are [disputing definitions](#) and ignoring my actual question. Your next reply should answer the question, or admit that you do not know of an answer.

Alicorn 07 April 2011 06:12:49PM 4 points 

gmail's spam filter does not have degrees of belief or belief.

It has things which you could call by those words if you really wanted to. But it wouldn't make them into the same things those words mean when referring to people.

Augh, this reminded me of a quote that I can't seem to find based on my tentative memory of its component words... it was something to the effect that we anthropomorphize computers and talk about them "knowing" things or "communicating" with each other, and some people think that's wrong and they don't really do those things, and the quote-ee was of the opinion that computers were clarifying what we meant by those concepts all along. Anybody know what I'm talking about?

curi 07 April 2011 06:37:10PM 1 point 

To be clear, I think computers *can* do those things and AIs *will*, and *that* will help clarify the concepts a lot.

But I don't think that microsoft word does it. Nor any game "AI" today. Nor gmail's spam filter which just does mindlessly math.

jimrandomh 07 April 2011 01:02:59AM 12 points 

The assumptions behind Cox's theorem are:

1. Representation of degrees of plausibility by real numbers
2. Qualitative correspondence with common sense
3. Consistency

Would you please clearly state which of these you disagree with, and why? And if you disagree with (1), is it because you don't think degrees of plausibility should be represented, or because you think they should be represented by something other than real numbers, and if so, then what? (Please do not give an answer that cannot be defined precisely by mapping it to a mathematical set. And please do not suggest a representation that is obviously inadequate, such as booleans.)

curi 07 April 2011 03:00:06AM 1 point 

Could you explain what you're talking about a bit more? For example you state "consistency" as an assumption. What are you assuming is (should be?) consistent with what?

Larks 08 April 2011 01:10:11AM 3 points 

If only [Jaynes](#) had clearly listed them on page 114!

JoshuaZ 07 April 2011 03:25:19AM 11 points 

You may have valid points to make but it might help in getting people to listen to you if you don't exhibit apparent double standards. In particular, your main criticism seems to be that people aren't reading Popper's texts and related texts enough. Yet, at the same time, you are apparently unaware of the basic philosophical arguments for Bayesianism. This doesn't reduce the validity of anything you have to say but as an issue of trying to get people to listen, it isn't going to work well with fallible humans.

curi 07 April 2011 03:33:31AM -6 points 

It's fine if most people haven't read Popper. But they should be able to point to some Bayesian somewhere who did, or they should know at least one good argument against a major Popperian idea. or they should be interested and ask more about him instead of posting incorrect arguments about why his basic claims are false.

I do know, offhand, several arguments against Bayesian epistemology (e.g. it's inability to create moral knowledge, and i know many arguments against induction, each decisive). And anyway I came here to learn more about it. One particular thing I would be interested in is a Bayesian criticism of Popper. Are there any? By contrast (maybe), Popper did criticize Bayesian epistemology in LScD and elsewhere. And I am familiar with those criticisms.

Learning enough Bayesian stuff to sound like a Bayesian so people want to listen to me more sounds to me like more trouble than it's worth, no offense. I'm perfectly willing to read more things when I make a mistake and there is a specific thing which explains the issue. I have been reading various things people refer me to. If you wanted me to study Bayesian stuff for a month before speaking, well, I'd get bored because I would see flaws and then see them repeated, and then read arguments which depend on them. I did read the whole HP fic if that helps.

One thing that interests me, which I posted about in the initial post, is how unscholarly some Bayesian scholars are. Can anyone correct that? Are there any with higher scholarly standards? I would like there to be. I don't want to just read stuff until I happen to find something good, I want to be pointed to something considerably better than the unscholarly stuff I criticized. I don't know where to find that.

JoshuaZ 07 April 2011 04:55:06AM * 17 points

[...]

It's fine if most people haven't read Popper. But they should be able to point to some Bayesian somewhere who did, or they should know at least one good argument against a major Popperian idea. or they should be interested and ask more about him instead of posting incorrect arguments about why his basic claims are false.

Really? Much of that seems questionable. There are many different ideas out there and practically speaking, there are too many ideas out there for people to have to deal with every single one. Sure, making incorrect arguments is bad. And making arguments against strawmen is very bad. But people don't have time actually research every single idea out there or even know which one's to look at. Now, I think that Popper is important enough and has relevant enough points that he should be on the short list of philosophers that people can grapple with at least to some limited extent. But frankly, speaking as someone who is convinced of that point, you are making a very weak case for it.

I do know, offhand, several arguments against Bayesian epistemology (e.g. it's inability to create moral knowledge, and i know many arguments against induction, each decisive).

This paragraph seems to reflect a general problem you are having here in making assertions without providing any information other than vague claims of existence. I am for example aware of a large variety of arguments against induction (the consistency of anti-induction frameworks seems to be a major argument) but calling them "decisive" is a very strong claim, and isn't terribly relevant in so far as Bayesianism is not an inductive system in many senses of the term.

You've also referred to before to this claim that Popperian system can lead to moral knowledge and that's a claim I'd be very curious to hear expanded with a short summary of how that works. Generally when I see a claim that an epistemological system can create moral knowledge my initial guess is that someone has managed to bury the naturalistic fallacy somewhere or has managed to smuggle in additional moral premises that aren't really part of the epistemology. I'd be pleasantly surprised to see something that didn't function that way.

One particular thing I would be interested in is a Bayesian criticism of Popper. Are there any? By contrast (maybe), Popper did criticize Bayesian epistemology in LScD and elsewhere.

I haven't read it myself but I've been told that Earman's "Bayes or Bust" deals with a lot of the philosophical criticisms of Bayesianism as well as giving a lot of useful references. It should do a decent job in regards to the scholarly concerns.

As to Popper's criticism of Bayesianism, the discussion of it in LScD is quite small, which is understandable in that Bayesianism was not nearly as developed in that time as it is now. (You may incidentally be engaging in a classical philosophical fallacy here in focusing on a specific philosopher's

personal work rather than the general framework of ideas that followed from it. There's a lot of criticism of Bayesianism that is not in Popper that is potentially strong. Not everything is about Popper.)

Learning enough Bayesian stuff to sound like a Bayesian so people want to listen to me more sounds to me like more trouble than it's worth, no offense.

As a non-Bayesian, offense taken. You can't expect to go to a room full of people with a specific set of viewpoints offer a contrary view, act like the onus is on them to translate into your notation and terminology, and then be shocked when they don't listen to you. Moreover, knowing the basics of Cox's theorem is not asking you to "sound like a Bayesian" anyhow.

If you wanted me to study Bayesian stuff for a month before speaking, well, I'd get bored because I would see flaws and then see them repeated, and then read arguments which depend on them. I did read the whole HP fic if that helps.

What? I don't know how to respond to that. I'm not sure an exclamation exists in standard English to express my response to that last sentence. I'm thinking of saying "By every deity in the full Tegmark ensemble" but maybe I should wait for a better time to use it. You are repeatedly complaining about people not knowing much about Popper while your baseline for Bayesianism is that you've *read an incomplete Harry Potter fanfic*? This fanfic hasn't even addressed Bayesianism other than in passing. This seems akin to someone thinking they understand rocketry because they've watched "Apollo 13".

endoself 29 May 2011 09:32:12AM 2 points

By every deity in the full Tegmark ensemble

Can I steal this?

JoshuaZ 29 May 2011 06:57:37PM 0 points

Yes, by all means feel free.

curi 07 April 2011 09:21:30AM 1 point

Really? Much of that seems questionable. There are many different ideas out there and practically speaking, there are too many ideas out there for people to have to deal with every single one.

The number of major ideas in epistemology is not very large. After Aristotle, there wasn't very much innovation for a long time. It's a small enough field you can actually trace ideas all the way back to the start of written history. Any professional can look at everything important. Some Bayesian should have. Maybe some did, but I haven't seen anything of decent quality.

You've also referred to before to this claim that Popperian system can lead to moral knowledge and that's a claim I'd be very curious to hear expanded with a short summary of how that works.

It works exactly identically to how Popperian epistemology creates any other kind of knowledge. There's nothing special for morality.

Knowledge is created by an evolutionary process involving *conjecture* and *refutation*. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

You may incidentally be engaging in a classical philosophical fallacy here in focusing on a specific philosopher's personal work rather than the general framework of ideas that followed from it.

I have a lot of familiarity with the other Popperians. But Popper and Deutsch are by far the best. There isn't really anything non-Popperian that draws on Popper much. Everyone who has understood Popper is a Popperian, IMO. If you disagree, do tell.

As to Popper's criticism of Bayesianism, the discussion of it in LScD is quite small

Small is not a criticism; substance matters not length. Do you have a criticism of his arguments in LScD or not? Also he dealt with it elsewhere, as I stated.

JoshuaZ 07 April 2011 03:40:44PM * 2 points

The number of major ideas in epistemology is not very large. After Aristotle, there wasn't very much innovation for a long time. It's a small enough field you can actually trace ideas all the way back to the start of written history. Any professional can look at everything important. Some Bayesian should have. Maybe some did, but I haven't seen anything of decent quality.

As to a professional, I already referred you to Earman. Incidentally, you seem to be narrowing the claim somewhat. Note that I didn't say that the set of major ideas in epistemology isn't small, I referred to the much larger class of philosophical ideas (although I can see how that might not be clear from my wording). And the set is indeed very large. However, I think that your claim about "after Aristotle" is both wrong and misleading. There's a lot of what thought about epistemological issues in both the Islamic and Christian worlds during the Middle Ages. Now, you might argue that that's not helpful or relevant since it gets tangled up in theology and involves bad assumptions. But that's not to say that material doesn't exist. And that's before we get to non-Western stuff (which admittedly I don't know much about at all).

(I agree when you restrict to professionals, and have already recommended Earman to you.)

It works exactly identically to how Popperian epistemology creates any other kind of knowledge. There's nothing special for morality.

Knowledge is created by an evolutionary process involving conjecture and refutation. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

This is a deeply puzzling set of claims. First of all, a major point of his epistemological system is falsifiability based on data (at least as I understand it from LScD). How that would at all interact with moral issues is unclear to me. Indeed, the semi-canonical example of a non-falsifiable claim in the Popperian sense is Marxism, a set of ideas that has a large set of attached moral claims.

I also don't see how this works given that moral claims can always be criticized by the essential sociopathic argument "I don't care. Why should you?" Obviously, that line of thinking can be/should be expanded. To use your earlier example, how would you discuss "murder is wrong" in a Popperian framework? I would suggest that this isn't going to be any different than simply discussing moral ideas based on shared intuitions with particular attention to the edge cases. You're welcome to expand on these claims, but right now, nothing you've said in this regard is remotely convincing or even helpful since it amounts to just saying "well, do the same thing."

I have a lot of familiarity with the other Popperians. But Popper and Deutsch are by far the best. There isn't really anything non-Popperian that draws on Popper much. Everyone who has understood Popper is a Popperian, IMO. If you disagree, do tell.

I'm going to be obnoxious and quote a friend of mine "Everyone who understands Christianity is a Christian." I don't have any deep examples of other individuals although I would tentatively say that I understood Popper's views in Logic of Scientific Discovery just fine.

Do you have a criticism of his arguments in LScD or not?

Sure. The most obvious one is when he is discussing the law of large numbers and frequentist v. Bayesian interpretations (incidentally to understand those passages it is helpful to note that he uses the term "subjective" to describe Bayesians rather than Bayesian which is consistent with the language of the time, but in modern terminology has a very different meaning (used to distinguish between subject and objective Bayesians)). In that section he argues that (I don't have the page number unfortunately since I'm using my Kindle edition. I have a hard copy somewhere but I don't know where) that "it must be inadmissible to give *after* the deduction of Bernoulli's theorem a meaning to p different from the one which was given to it before the deduction." This is, simply put, wrong. Mathematicians all the time prove something in one framework and then interpret it in another framework. You just need to show that all the properties of the relevant frameworks overlap in sufficiently non-pathological cases. If someone wrote this as a complaint about say using the complex exponential to understand the symmetries of the Euclidean plane, we'd

immediately see this as a bad claim. There's an associated issue in this section which also turns up but it is more subtle; Popper doesn't appreciate what you can do with measure theory and L_p spaces and related ideas to move back and forth between different notions of probability and different metrics on spaces. That's ok, it was a very new idea when he wrote LScD (although the connections were to some extent definitely there). But it does render a lot of what he says simply irrelevant or outright wrong.

curi 07 April 2011 06:58:34PM * 0 points

[...]

As to a professional, I already referred you to Earman.

Which you stated you had not read. I have rather low standards for recommendations of things to read, but "I never read it myself" isn't good enough.

I don't agree with "restrict to professionals". How is it to be determined who is a professional? I don't want to set up arbitrary, authoritative criteria for dismissing ideas based on their source.

First of all, a major point of his epistemological system is falsifiability based on data (at least as I understand it from LScD).

That is a major point *for scientific research* where the problem "how do we use evidence?" is important. And the answer is "criticisms can refer to evidence". Note by "science" here I mean any empirical field. What do you do in non-scientific fields? You simply make criticisms that don't refer to evidence. Same method, just missing one type of criticism which is rather useful in science but not fundamental to the methodology.

Indeed, the semi-canonical example of a non-falsifiable claim in the Popperian sense is Marxism, a set of ideas that has a large set of attached moral claims.

It is not *empirically falsifiable*. It is criticizable. For example Popper criticized Marx in *The Open Society and its Enemies*..

I also don't see how this works given that moral claims can always be criticized by the essential sociopathic argument "I don't care. Why should you?"

Any argument which works against everything fails at the task of differentiating better and worse ideas. So it is a bad argument. So we can reject it and all other things in that category, by this criticism.

To use your earlier example, how would you discuss "murder is wrong" in a Popperian framework?

The short answer is: since we don't care to have justified foundations, you can discuss it any way you like. You can say it's bad because it hurts people. You can say it's good because it prevents overpopulation. You can say it's bad because it's mean. These kinds of normal arguments, made by normal people, are not deemed automatically invalid and ignored. Many of them are indeed mistakes. But some make good points.

For more on morality, please join this discussion:

http://lesswrong.com/lw/552/reply_to_benelliott_about_popper_issues/3uv7

I would tentatively say that I understood Popper's views in Logic of Scientific Discovery just fine.

He has like 20 books. There's way more to it. When one reads a lot of them, a whole worldview comes across that is very hard to understand from just a couple books. And I wasn't trying to *argue* with that statement, I was just commenting. I mentioned it because of a comment to do with whether I had studied results of non-Popperians using Popperian ideas.

"it must be inadmissible to give after the deduction of Bernoulli's theorem a meaning to p different from the one which was given to it before the deduction." This is, simply put, wrong.

Are you really telling me that you can prove something, then take the conclusion, redefine a term, and work with that, and consider it still proven? You could only do that *if* you created a

second proof that the change doesn't break anything, you can't just do it. I'm not sure you took what Popper was saying literally enough; I don't think your examples later actually do what he criticized. Changing the meaning of a term in a conclusion statement, and considering a conclusion from a different perspective, are different.

Popper doesn't appreciate what you can do with measure theory and L_p spaces

Would you understand if I said this has no relevance at all to 99.99% of Popper's philosophy? Note that his later books generally have considerably less mention of math or logic.

JoshuaZ 08 April 2011 02:43:45AM * 3 points

Which you stated you had not read. I have rather low standards for recommendations of things to read, but "I never read it myself" isn't good enough.

Earman is a philosopher and the book has gotten positive reviews from other philosophers. I don't know what else to say in that regard.

I don't agree with "restrict to professionals". How is it to be determined who is a professional? I don't want to set up arbitrary, authoritative criteria for dismissing ideas based on their source.

Hrrm? You mentioned professionals first. I'm not sure why you are now objecting to the use of professionals as a relevant category.

That is a major point for scientific research where the problem "how do we use evidence?" is important. And the answer is "criticisms can refer to evidence". Note by "science" here I mean any empirical field. What do you do in non-scientific fields? You simply make criticisms that don't refer to evidence. Same method, just missing one type of criticism which is rather useful in science but not fundamental to the methodology

I'm not at all convinced that this is what Popper intended (but again I've only read LScD) but if this is accurate then Popper isn't just wrong in an interesting way but is just wrong. Does one mean for example to claim that pure mathematics works off of criticism? I'm a mathematician. We don't do this. Moreover, it isn't clear what it would even mean for us to try to do this as our primary method of inquiry. Are we supposed to spend all our time going through pre-existing proofs trying to find holes in them?

He has like 20 books. There's way more to it. When one reads a lot of them, a whole worldview comes across that is very hard to understand from just a couple books.

Yes, and I'm quite sure that I get much more of a worldview if I read all of Hegel rather than just some of it. That doesn't mean I need to read all of it. Similar remarks would apply to Aquinas or more starkly the New Testament. Do you need to read all of the New Testament to decide that Christianity is bunk? Do you need to read the entire Talmud to decide that Judaism is incorrect? But you get a whole worldview that you don't obtain from just reading the major texts.

The short answer is: since we don't care to have justified foundations, you can discuss it any way you like. You can say it's bad because it hurts people. You can say it's good because it prevents overpopulation. You can say it's bad because it's mean. These kinds of normal arguments, made by normal people, are not deemed automatically invalid and ignored. Many of them are indeed mistakes. But some make good points

Right, and then we just the criticism "why bother" or "and how does that maximize the number of paperclip in the universe?" Or one can say "mean" "good" bad" are all hideously ill-defined. In any event, does it not bother you that you are essentially claiming that your moral discussion with your great epistemological system looks just like a discussion about morality by a bunch of random individuals? There's nothing in the above that uses your epistemology in any substantial way.

Are you really telling me that you can prove something, then take the conclusion, redefine a term, and work with that, and consider it still proven? You could only do that if you created a second proof that the change doesn't break anything, you can't just do it.

Right! And conveniently in the case Popper cares about you can prove that.

Popper doesn't appreciate what you can do with measure theory and

Would you understand if I said this has no relevance at all to 99.99% of Popper's philosophy? Note that his later books generally have considerably less mention of math or logic.

Do you mean understand or do you mean care? I don't understand why you are making this statement given that my remark was addressing the question you asked of whether I had specific problems with Popper's handling of Bayesianism in LScD. This is a specific problem there.

AlephNeil 08 April 2011 06:12:28PM * 1 point

Does one mean for example to claim that pure mathematics works off of criticism? I'm a mathematician. We don't do this.

I don't know what Popper himself would say, but one of his more insightful followers, namely Lakatos, argues for [exactly that position](#).

I read Proofs and Refutations too many years ago to say anything precise about it. I remember finding it interesting but also frustrating. Lakatos seems determined to ignore/deny/downplay the fact of mathematical practice that we only call something a 'theorem' when we've got a proof, and we only call something a 'proof' when it's logically watertight in such a way that no 'refutations' are possible. Still, it's well-researched (in its use of a historical case-study) and he comes up with some decent ideas along the way (e.g. about "monster barring" and "proof-oriented definitions".)

[continue this thread »](#)

curi 08 April 2011 09:47:58AM -7 points

Does one mean for example to claim that pure mathematics works off of criticism?

yes

I'm a mathematician. We don't do this.

Instead, you make appeals to authority?

[continue this thread »](#)

Randaly 07 April 2011 01:39:46PM 2 points

Err, Bayesian probability doesn't have anything special for morality either. People on LW tend to be moral non-realists, ie people who deny that there is objective moral knowledge, if that's what you're talking about (not sure- sorry!), but that's completely orthogonal to this discussion: there's nothing in Bayesianism that leads inevitably to non-realism. (Also, I'm not convinced that moral realism is right, so saying "Bayesianism leads to moral non-realism" isn't a very effective argument.)

curi 07 April 2011 07:08:51PM 0 points

Bayesian epistemology *doesn't create moral knowledge* because it only functions when fed in *observation data* (or assumptions). I get a lot of conflicting statements here, but some people tell me they *only care about prediction*, they are *instrumentalists*, and that is what Bayes stuff is for, and they don't regard it as a bad thing that it doesn't address morality at all.

Now what you have in mind, I think, is that if you make a ton of assumptions you could *then* talk about morality using Bayes. Popperism doesn't require a bunch of arbitrary starting

assumptions to create moral knowledge, it just can deal with it directly.

If I'm wrong, explain how you can deal with figuring out, e.g., what are good moral values to have (without assuming a utility function or something).

Randy 09 April 2011 04:22:53AM 2 points

As I tried to say (and probably explained really poorly- sorry!), the LW consensus is that morality is not objective. Therefore, the idea of figuring out what good moral values would be is, according to moral non-realism, impossible: any decision about what a good moral value is must rely on your pre-existing values, if an objective morality is not out there to be discovered. Using this as a criticism of Bayesianism is sorta like criticizing thermodynamics because it claims it's impossible to exactly specify the position and velocity of each particle: not only is the criticism unrelated to the subject matter, but satisfying it would require the theory to do something that is to the best of our knowledge incorrect.

[deleted] 07 April 2011 11:19:12AM * 2 points

Knowledge is created by an evolutionary process involving conjecture and refutation. By criticizing flaws in ideas, we seek to improve them (by making better conjectures we hope will eliminate the flaws).

I'm inclined to take this formula seriously, but I'd like to start by applying it to innate knowledge, knowledge we are born with, because here we are definitely talking about an evolutionary process involving mutation and natural selection. Some mutations add what amounts to a new innate conjecture (hypothesis, belief) into the cognitive architecture of the creature.

However, what occurs at this point is not that a creature with a false innate conjecture is eliminated. The creature isn't being tested purely against reality in isolation. It's being tested against other members of its species. The creature with the least-false, or least-perilously-false conjecture will tend to do better than the competitors. The competition for survival amounts to a competition between rival conjectures. The truest, or most-usefully-true, or least-wrong, or least-dangerously-wrong innate belief will tend to outdo its competitors and ultimately spread through the species. (With the odd usefully-wrong belief surviving.)

The occasional appearance of new innate conjectures resembles the conjecture part of Popperian conjecture and refutation. However, the contest between rival innate conjectures that occurs as the members of the species struggle against each other for survival seems less Popperian than Bayesian.

The relative success of the members of the species who carry the more successful hypothesis vaguely resembles Bayesian updating, because the winners increase their relative numbers and the losers decrease their relative numbers, which resembles the shift in the probabilities assigned to rival hypotheses that occurs in Bayesian updating. Consider the following substitutions applied to Bayes' formula:

$$P(H|E) = P(E|H)P(H) / P(E)$$

- $P(H|E)$ is the new proportion (i.e. in the next generation) of the species carrying the hypothesis H , given that event E occurred (E is "everything that happened to the generation")
- $P(E|H)$ is the degree to which H predicts and thus prepares the individual to handle E (measured in expected number of offspring given E)
- $P(H)$ is the old proportion (i.e. in the previous generation) of the species carrying H
- $P(E)$ is the degree to which the average member of the species predicts and thus is prepared to handle E (measured in expected number of offspring given E)

With these assignments, what the equation means is:

The new proportion of the species with H is equal to the old proportion of the species with H , times the expected number of offspring of members with H , divided by the expected number of offspring of the average member of the species.

One difference between this process and Bayesian updating is that this process allows the occasional introduction of new hypotheses over time, with what amounts to a modest but not vanishing initial prior.

curi 07 April 2011 07:19:31PM 0 points



I'm not sure if we're interested in the same stuff. But taking up one topic:

I think you regard innate/genetic ideas as important. I do not. Because people are universal knowledge creators, and can change any idea they start with, it doesn't matter very much.

The reason people are so biased is not in their genes but their memes.

There are two major replication strategies that memes use.

- 1) a meme can be useful and rational. it spreads because of its value
- 2) a meme can sabotage its holders creativity to prevent him from criticizing it, and to take away his choice not to spread it

The second type dominated all cultures on Earth for a long time. The transition to the first type is incomplete.

More details one memes and universality can be found in *The Beginning of Infinity* by David Deutsch

[deleted] 07 April 2011 10:10:10PM 2 points



I think you regard innate/genetic ideas as important. I do not. Because people are universal knowledge creators, and can change any idea they start with, it doesn't matter very much.

You misunderstand. I bring it up as a model of learning, and my choice was based on your own remarks. You said that knowledge is created by an evolutionary process. That way of putting it suggests an analogy with Darwin's theory of evolution as proceeding by random variation and natural selection. And indeed there is an analogy between Popper's conjectures and refutations and variation and natural selection, and it is this: a conjecture is something like variation (mutation), and refutation is something like natural selection.

However, what I found was that the closer I looked at knowledge which is actually acquired through natural selection - what we might call innate knowledge or instinctive knowledge - the more the process of acquisition resembled Bayesian updating rather than Popperian conjecture and refutation. I explained why.

In Bayesian updating, there are competing hypotheses, and the one for which actual events are less of a surprise (i.e., the hypothesis H_i for which $P(E|H_i)$ is higher) is strengthened relative to the one for which events are more of a surprise. I find a parallel to this in competition among alleles under natural selection, which I described.

Essential to Bayesian updating is the coexistence of competing hypotheses, and essential to natural selection is the coexistence of competing variants in a species. In contrast, Popper talks about conjecture and refutation, which is a more lonely process that need not involve more than one conjecture and a set of observations which have the potential to falsify it. Popper talks about improving the conjecture in response to refutation, but this process more resembles Lamarckian evolution than Darwinian evolution, because in Lamarckian evolution the individuals improve themselves in response to environmental challenges, much as Popper would have us improve our conjectures in response to observational challenges. Also, in Lamarckian evolution, as in the Popperian process of conjecture and refutation, competing variants (compare: competing hypotheses) do not play an *essential* role (though I'm sure they could be introduced). Rather, the picture is of a *single* animal (compare: a single hypothesis) facing existential environmental challenges (compare: facing the potential for falsification) improving itself in response (which improvement is passed to offspring).

The Popperian process of conjecture, refutation, and improvement of the conjecture, can as

it happens be understood from a Bayesian standpoint. It does implement Bayesian updating in a certain way. Specifically, when a particular conjecture is refuted and the scientist modifies the conjecture - at that point, there are two competing hypotheses. So at that point, the process of choosing between these two competing hypotheses can be characterized as Bayesian updating. The less successful hypothesis is weakened, and the more successful hypothesis is strengthened.

In short, if you want to take seriously the analogy that does exist between evolution through natural selection and knowledge acquisition of whatever type, then you may want to take a closer look at Bayesian updating as conforming more closely to the Darwinian model.

curi 07 April 2011 10:12:27PM 1 point

In short, if you want to take seriously the analogy

I wasn't talking about an analogy.

Evolution is a theory which applies to any type of replicator. Not by analogy by literally applies.

Make sense so far?

[continue this thread »](#)

jimrandomh 07 April 2011 03:18:22AM * 4 points

Cox's theorem is a proof of Bayes rule, from the conditions above. "Consistency" in this context means (Jaynes 19): If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result; we always take into account all of the evidence we have relevant to a question; and we always represent equivalent states of knowledge by equivalent plausibility assignments. By "reason in more than one way", we specifically mean adding the same pieces of evidence in different orders.

(Edit: It's page 114 in the PDF you linked. That seems to be the same text as my printed copy, but with the numbering starting in a different place for some reason.)

[deleted] 07 April 2011 03:43:28AM -5 points

Assigning degrees of plausibility to theories is an attempt to justify them. Cox's theorem just assumes you can do this. Popper argued that justification, including probabilistic justification, is impossible. How does just assuming something that Popper refuted show anything?

timtyler 07 April 2011 12:36:50PM * 0 points

Assigning degrees of plausibility to theories is an attempt to justify them. Cox's theorem just assumes you can do this.

Do you think that [grue and bleen](#) are as plausible as blue and green? Would you like to bet?

benelliott 07 April 2011 12:44:50PM 1 point

Nitpicking here, grue and bleen aren't statements and thus can't be assigned probabilities. "This object is grue" and "this object is bleen" are statements.

timtyler 07 April 2011 01:21:50PM * 0 points

Yes, I left making up more specific examples as an exercise for the reader.

[deleted] 07 April 2011 11:55:24PM 0 points

Assuming that the object in question is an emerald, then grue is in conflict with our best explanations about emeralds whereas there are no known problems with the idea that the emerald is green. So I go with green, but not because I have assigned degrees of plausibility but because I see no problem with green.

benelliott 07 April 2011 06:16:17AM 5 points

One argument for plausibility would be this.

At some point you may be called on to base a decision on whether something is true or false. The simplest of these decisions can be reduced to betting for or against something, and you cannot always choose not to bet. There must be some odds at which you switch from betting on falsity to betting on truth, and those can be taken to demonstrate your plausibility assignment.

[deleted] 07 April 2011 07:41:01AM 0 points

How does betting on the truth of a universal theory work? I can't see a bookie ever paying out on that, although it would be good business to get punters to take such bets.

timtyler 07 April 2011 12:31:50PM 2 points

So: just bet on things the theory predicts instead.

[deleted] 07 April 2011 11:25:40PM 0 points

Having the prediction turn out doesn't make the theory true or more likely, it is just consistent evidence. There are an infinitude of other theories that the same evidence is consistent with.

timtyler 08 April 2011 12:16:34PM 4 points

To give a simple example, consider flipping a coin. You observe HHH. Is this a fair coin? or a double-headed one? or a biased coin? Different theories describe these situations, and you could be asked to bet on them. Imagine you then further observe HHHH - making a total of HHHHHHH. This makes your estimate of the chances of the "double-headed coin" hypothesis go up. Other hypotheses may increase in probability too - but we are not troubled by there being an infinity of them, since we give extra weight to the simpler ones, using Occam's razor.

benelliott 07 April 2011 11:16:26AM * 4 points

The usual way on Less Wrong is to bring in Omega, the all powerful all knowing entity who spends his free time playing games with us mortals, and for some reason most of his games illustrate some point of probability or decision theory. With Omega acting as the bookie you can be forced to assign a probability to any meaningful statement. Some people just respond to such scenarios by asserting that Omega is impossible, I don't know if you're one of those people but I'll try a different approach anyway.

Imagine that in 2050 the physicists have narrowed down all their candidates for a Theory of Everything to just two possibilities, creatively named X-theory and Y-theory.

An engineer who is a passionate supporter of X-theory has designed and built a new power plant. If X-theory is correct, his power plant will produce a limitless supply of free energy and bring humanity into a post-scarcity era.

However, a number of physicists have had a look at his designs, and have shown that if Y-theory is correct his power plant will create a black hole and wipe out humanity as soon as it is turned on. Somehow, it has ended up being your decision whether or not it goes on.

This is one such 'bet', it may not a very likely scenario but you should still be able to handle it. If we combine it with many slightly altered dilemma we can figure out your probability estimate of theory X being correct, whether you admit to having one or not.

[deleted] 07 April 2011 11:18:05PM * 0 points [-]

You've presented this as a scenario in which you have to make a choice between two conflicting theories. But the problem you face isn't should I choose X or should I choose Y; the problem you face is that given you have this conflict, what should I do now. This problem is objective, it is different to the problem of whether X is right or Y is right, and it is solvable. Given that this is the year 2050 and humanity won't in fact be wanting, the best solution to the problem may be to wait, pending further research to resolve the conflict. This isn't an implicit bet against X and for Y, it is a solution to a different problem to the problems X and Y address.

benelliott 08 April 2011 06:56:28AM * 2 points [-]

For sake of argument we say that the plant requires a rare and unstable isotope to get started. Earth's entire supply is contained in the plant and will decay in 24 hours.

I could also ask you a similar dilemma, but this time there is only one theory which acknowledges that whether the plant works or creates a black hole depends on a single quantum event, which has a 50% chance of going either way. What do you do? If you wouldn't launch I can ask the same question but now there's only a 25% chance of a black hole, and so on until I learn the ratio of the utility values that you assign to "post scarcity future" and "extinction of humanity". This might for example tell me that the chance of a black hole has to be less than 30% for you to press the button.

Then I ask you the original dilemma, and learn whether the probability you assign to theory X is above or below 70%. If I have far too much time on my hands I can keep modifying the dilemma with slightly altered pay-offs until I pinpoint your estimate.

wedrifid 08 April 2011 07:07:57AM 1 point [-]

| creates a black whole

I suppose you get that when the container containing the black dye explodes....

[continue this thread »](#)

JoshuaZ 08 April 2011 03:32:23AM 1 point [-]

This avoids the question. If it helps, try to construct a version of this in [the least convenient possible world](#). For example, one obvious thing to do would be that something about theory X means the plant can only be turned on at a certain celestial conjunction (yes, this is silly but it gets the point across. That's why it is a least convenient world) and otherwise would need to wait a thousand years.

One can vary the situation. For example, it might be that under theory X, medicine A will save a terminally ill cancer patient, and under theory Y, medicine B will save them. And A and B together will kill the patient according to both theories.

JoshuaZ 07 April 2011 01:00:33AM 5 points

There's an associated problem here that may be getting ignored: Popper isn't a terribly good writer. "The Logic of Scientific Discovery" was one of the first phil-sci books I ever read and it almost turned me off of phil-sci. This is in contrast for example with Lakatos or Kuhn who are very readable. Some of the difficulty with reading Popper and understanding his viewpoints is that he's just tough to read.

That said, I think that chapter 3 of that books makes clear that Popper's notion of falsification is more subtle than what I would call "naive Popperism". But Popper never fully gave an explanation of how to distinguish between strict falsification theory and his notions.

There's an associated important issue: many people claim to support naive Popperism as an epistemological position, either as a demarcation between science and non-science or as a general epistemological approach. In so far as both are somewhat popular viewpoints (especially among scientists) responding to and explaining what is wrong with that approach is important even as one should acknowledge that Popper's own views were arguably more nuanced.

curi 07 April 2011 03:03:25AM 0 points

I do not find Popper hard to read.

Popper never fully gave an explanation of how to distinguish between strict falsification theory and his notions.

Did you read his later books? He does explain his position. One distinguishing difference is that Popper is not a justificationist and they are. Tell me if you don't know what that means.

falenas108 07 April 2011 12:28:06AM * -1 points

From the research I have done in the last 5 minutes, it seems as though Popper believed that all good scientific theories should be subject to experiments that could prove them wrong.

Ex:

"the falsificationists or fallibilists say, roughly speaking, that what cannot (at present) in principle be overthrown by criticism is (at present) unworthy of being seriously considered; while what can in principle be so overthrown and yet resists all our critical efforts to do so may quite possibly be false, but is at any rate not unworthy of being seriously considered and perhaps even of being believed" -Popper

This seems to imply that theories can be proved false.

curi 07 April 2011 12:32:34AM -3 points

Replying to accusations of un scholarly criticism of Popper with an *unsourced* Popper quote is very silly.

That the quote doesn't say what you claim it does (as I read it), and you make no attempt to explain your reading of it, is also silly.

[deleted] 07 April 2011 12:59:57AM * 3 points

The quote came from *Conjectures and Refutations*, pg 309. I agree that it doesn't say what falenas108 claims. Plus a bit has been missed out at the end: " -- though only tentatively." Also, on the following page, Popper says:

For us [fallibilists] ... science has nothing to do with the quest for certainty or probability or reliability. We are not interested in establishing scientific theories as secure, or certain, or probable. Conscious of our own fallibility we are only interested in criticizing them and testing them, hoping to find out where we are mistaken; of learning from our mistakes; and , if we are lucky, of proceeding to better theories.

So Popper would not assert that theories can be established as definitely false.

Peterdjones 12 April 2011 08:28:49PM 0 points

Of course, in reality, fallibilism just means you don't look for certainty. You can and should look for more probable theories, or as P. calls them, "better theories".

curi 13 April 2011 01:53:00AM -2 points 

Hi Peter,

How did you find me here?

falenas108 07 April 2011 12:56:18AM 0 points 

Citation: Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*, New York: Harper and Row. Reprinted London: Routledge, 1974.

It says theories should resist being overthrown for them to be proper theories. That implies that it is possible for a theory to be overthrown.

curi 07 April 2011 01:10:57AM 1 point 

No page number isn't very nice. For anyone interested, it is on page 309, which is at the start of chapter 10 section 3.

If you read the context, you will find, for example, an explicit denouncement of the quest for certainty on the next page. Plus elaboration. Popper's position in these matters is not unclear.

curi 07 April 2011 12:59:11AM * 7 points 

A theory can be fallibly overthrown, but not definitely overthrown, in Popper's view. Quotes out of context are easy to misread when you are not familiar with the ideas, and when you make assumptions (e.g. that overthrowing must be definitive) that the author does not make.

Peterdjones 12 April 2011 08:25:35PM 1 point 

"A theory can be fallibly overthrown, but not definitely overthrown, in Popper's view. "

So maybe Jaynes was using "disprove" to mean "fallibly overthrow".

falenas108 07 April 2011 01:01:26AM 3 points 

Ok, thanks for correcting me.

[deleted] 07 April 2011 12:25:14AM * 4 points 

The naturalist philosopher Peter Godfrey Smith said this of Popper's position:

[F]or Popper, it is never possible to confirm or establish a theory by showing its agreement with observations. Confirmation is a myth. The only thing an observational test can do is to show that a theory is false...Popper, like Hume, was an inductive skeptic, and Popper was skeptical about all forms of confirmation and support other than deductive logic itself...This position, that we can never be completely certain about factual issues, is often known as fallibilism...According to Popper, we should always retain a tentative attitude towards our theories, no matter how successful they have been in the past...[a]ll we can do is try out one theory after another. A theory that we have failed to falsify up till now might, in fact, be true. But if so, we will never know this or even have reason to increase our confidence.

(From *Theory and Reality*, p. 59-61.) Is this not an accurate description? You seem to think Popper didn't believe in definitive falsification, but this doesn't seem to be a universally accepted interpretation. Note also that Godfrey-

Smith does refer to Popper's position as fallibilism, so he is not being "unscholarly." Though Popper may have held the position that falsification can't be perfectly certain, he definitely didn't take this idea too seriously because his description of science as a process (step one: come up with conjectures; step two: falsify them) makes use of falsification by experiment.

I think the answer to your overarching question can be found [here](#). If we know that certain events are more probable given that certain other events happened, i.e. [conditional probability](#), we can make inferences about the future.

curi 07 April 2011 12:44:42AM * 3 points



Is this not an accurate description?

No. To start with, it's extremely incomplete. It doesn't really discuss what Popper's position is. It just makes a few scattered statements which do not explain what Popper is about.

The word "show" is ambiguous in the phrase "show that a theory is false". To a Popperian, equivocation over the issue of what is meant there is an important issue. It's ambiguous between "show definitively" and "show fallibly".

The idea that we can show a theory is false by an experimental test (even fallibly) is also, strictly, false, as Popper explained in LScD. When you reach a contradiction, something in the whole system is false. It could be an idea you had about how to measure what you wanted to measure. There's many possibilities.

You seem to think Popper didn't believe in definitive falsification, but this doesn't seem to be a universally accepted interpretation.

It's right there in LScD on page 56. I think it's in most of his other books too. I am familiar with the field and know of no competent Popper scholars who say otherwise.

Anyone publishing to the contrary is simply incompetent, or believed low quality secondary sources without fact checking them.

Though Popper may have held the position that falsification can't be perfectly certain, he definitely didn't take this idea too seriously because his description of science as a process (step one: come up with conjectures; step two: falsify them) makes use of falsification by experiment.

You have misinterpreted when you took "falsify them" to mean "falsify them with certainty". Popper is a fallibilist.

If we know that certain events are more probable given that certain other events happened

This does not even attempt to address important problems in epistemology such as how explanatory or philosophical knowledge is created.

[deleted] 07 April 2011 01:07:57AM * 3 points



I'll agree that Godfrey-Smith's definition is incomplete, but I don't think it really matters for the purpose of this discussion: I've already said I agree that Popper did not believe in certain confirmation, and this seems to be your main problem with this quote and with the ones other people gave. You wrote:

You have misinterpreted when you took "falsify them" to mean "falsify them with certainty". Popper is a fallibilist.

No, that is not what I meant at all. What I meant was, Popper was content with the fact that experimental evidence *can* say that something is *probably* false. If he wasn't, he wouldn't have included this his view of science as a process. So even though Popper was a fallibilist, he thought that when an experimental result argued against a hypothesis, it was good enough for science.

Next:

The idea that we can show a theory is false by an experimental test (even fallibly) is also, strictly, false, as Popper explained in LScD. When you reach a contradiction, something in the whole system is false. It could be an idea you had about how to measure what you wanted to measure. There's many possibilities.

Yes, this is the old "underdetermination of theory by data" problem, which Solomonoff Induction solves--see the coinflipping example [here](#).

Moving on, you wrote:

This does not even attempt to address important problems in epistemology such as how explanatory or philosophical knowledge is created.

Would you mind elaborating on this? What specific problems are you referring to here?

curi 07 April 2011 01:37:39AM 3 points

Popper was content with the fact that experimental evidence can say that something is probably false

That is not Popper's position. That is not even close. In various passages he explicitly denies it like "not certain or probable". To Popper, the claims that the evidence tells us something is certainly true, or probably true, are cousins which share an underlying mistake. You're assuming Popper would agree with you about probability without reading any of his passages on probability in which he, well, doesn't.

Arguing what books say with people who haven't read them gets old fast. So how about you just imagine a hypothetical person who had the views I attribute to Popper and discuss that?

Would you mind elaborating on this? What specific problems are you referring to here?

For example, the answers to all questions that have a "why" in them. E.g. *why* is the Earth roughly spherical? Statements with "because" (sometimes implied) is a pretty accurate way to find explanations, e.g. "because gravity is a symmetrical force in all directions". Another example is all of moral philosophy. Another example is epistemology itself, which is a philosophy not an empirical field.

Yes, this is the old "underdetermination of theory by data" problem

Yes

Which Solomonoff Induction solves--see the coinflipping example here.

This does not solve the problem to my satisfaction. It orders theories which make identical predictions (about all our data, but not about the unknown) and then lets you differentiate by that order. But isn't that ordering arbitrary? It's just not true that short and simple theories are always best; sometimes the truth is complicated.

[deleted] 07 April 2011 01:58:09AM * -1 points

Actually, one of the reason I stood by this interpretation of Popper was because one of the quotes posted in one of the other threads here:

"the falsificationists or fallibilists say, roughly speaking, that what cannot (at present) in principle be overturned by criticism is (at present) unworthy of being seriously considered; while what can in principle be so overturned and yet resists all our critical efforts to do so may quite possibly be false, but is at any rate not unworthy of being seriously considered and perhaps even of being believed"

Which is apparently from *Conjectures and Refutations*, pg 309. Regardless, I don't care about this argument overmuch, since we seem to have moved on to some other points.

[Solomonoff Induction] does not solve the problem to my satisfaction. It orders theories which make identical predictions (about all our data, but not about the unknown) and then lets you differentiate by that order. But isn't that ordering arbitrary? It's just not true that short and simple theories are always best; sometimes the truth is complicated.

Remember that in Bayesian epistemology, probabilities represent our state of knowledge, so as you pointed out, the simplest hypothesis that fits the data so far may not be the true one because we *haven't* seen all of the data. But it is *necessarily* our best guess because of the [conjunction rule](#).

JoshuaZ 07 April 2011 02:42:59AM 1 point

Remember that in Bayesian epistemology, probabilities represent our state of knowledge, so as you pointed out, the simplest hypothesis that fits the data so far may not be the true one because we haven't seen all of the data. But it is necessarily our best guess because of the conjunction rule.

You are going to have to expand on this. I don't see how the conjunction rule implies that simpler hypotheses are in general more probable. This is true if we have two hypotheses where one is X and the other is "X and Y" but that's not how people generally apply this sort of thing. For example, I might have a sequence of numbers that for the first 10,000 terms has the nth term as the nth prime number. One hypothesis is that the nth term is always the nth prime number. But I could have as another hypothesis some high degree polynomial that matches the first 10,000 primes. That's clearly more complicated. But one can't use conjunction to argue that it is less likely.

[deleted] 07 April 2011 04:52:44AM * 1 point

Imagine that I have some set of propositions, A through Z, and I don't know the probabilities of any of these. Now let's say I'm using these propositions to explain some experimental result--since I would have uniform priors for A through Z, it follows that an explanation like "M did it" is more probable than "A and B did it," which in turn is more probable than "G and P and H did it."

JoshuaZ 07 April 2011 04:58:22AM 1 point

Yes, I agree with you there. But this is much weaker than any general form of Occam. See my example with primes. What we want to say in some form of Occam approach is much stronger than what you can get from simply using the conjunction argument.

curi 07 April 2011 02:22:42AM * 1 point

There are so many problems here that it's hard to choose a starting point.

- 1) the data set you are using is biased (it is selective. all observation is selective)
- 2) there is no such thing as "raw data" -- all your observations are interpreted. your interpretations may be mistaken.
- 3) what do you mean by "best guess"? one meaning is "most likely to be the final, perfect truth". but a different meaning is "most useful now".
- 4) You say "probabilities represent our state of knowledge". However there are infinitely many theories with the same probability. Or there would be, except for your solomonoff prior about simpler theories having higher probability. So the important part of "state of our knowledge" as represented by these probabilities consists mostly of the solomonoff prior and nothing else, because it, and it alone, is dealing with the hard problem of epistemology (dealing with theories which make identical predictions about everything we have data for).
- 5) you can have infinite data and still get all non-empirical issues wrong
- 6) regarding the conjunction rule, there is miscommunication. this does not address the point i was trying to make. i think you have a premise like "all more complicated theories are merely conjunctions of simpler theories". But that is to conceive of theories very differently than Popperians do, in what we see as a limited and narrow way. To begin to address these issues, let's consider what's better: a bald assertion, or an assertion *and* an explanation of why it is correct? If you want "most likely to happen to be the perfect, final truth" you are better off with only an unargued assertion (since any argument may be mistaken). But if you want to learn about the world, you are better off not relying on unargued assertions.

jimrandomh 07 April 2011 01:48:41AM 3 points

For example, the answers to all questions that have a "why" in them. E.g. why is the Earth roughly

spherical? Statements with "because" (sometimes implied) is a pretty accurate way to find explanations, e.g. "because gravity is a symmetrical force in all directions". Another example is all of moral philosophy. Another example is epistemology itself, which is a philosophy not an empirical field.

For a formal mathematical discussion of these sorts of problems, read Causality by Judea Pearl. He reduces cause to a combination of conditional independence and ordering, and from this he defines algorithms for discovering causal models from data, predicting the effect of interventions and computing counterfactuals.

curi 07 April 2011 01:51:03AM * 1 point

Could you give a short statement of the main ideas? How can morality be reduced to math? Or could you say something to persuade me that that book will address the issues in a way I won't think misses the point? (e.g. by showing you understand what I think the point is, otherwise I won't expect you to be able to judge if it misses the point in the way I would).

jimrandomh 07 April 2011 02:01:00AM 2 points

Sorry, I over-quoted there; Pearl only discusses causality, and a little bit of epistemology, but he doesn't talk about moral philosophy at all.

His book is all about causal models, which are directed graphs in which each vertex represents a variable and each edge represents a conditional dependence between variables. He shows that the properties of these graphs reproduce what we intuitively think of as "cause and effect", defines algorithms for building them from data and operating on them, and analyzes the circumstances under which causality can and can't be inferred from the data.

curi 07 April 2011 02:28:44AM 2 points

I don't understand the relevance.

jimrandomh 07 April 2011 02:39:41AM 2 points

Your quote seemed to be saying that that Bayesianism couldn't handle why/because questions, but Popperian philosophy could. I mentioned Pearl as a treatment of that class of question from a Bayes-compatible perspective.

curi 07 April 2011 02:54:20AM 1 point

Causality isn't explanation. X caused Y isn't the issue I was talking about.

For example, the statement "Murder is bad because it is illiberal" is an explanation of *why* it is bad. It is not a statement about causality.

You may say that "illiberal" is a short cut for various other ideas. And you may claim that eventually that reduce away to causal issues. But that would be reductionism. We do not accept that high level concepts are a mistake or that emergence isn't important.

[continue this thread »](#)

falenas108 07 April 2011 12:31:38AM -1 points

Sorry, didn't see you posted this before I replied too...

[deleted] 07 April 2011 12:34:48AM 0 points

Actually, I'm glad you replied as well--the more quotes about/by Popper that we unearth, the more accurate we will be.

paulfchristiano 07 April 2011 12:19:41AM * 5 points

I don't understand Popper's work beyond the Wikipedia summary of critical rationalism. That summary, as well as the debate here at LW, appear to be confused and essentially without value. If this is not the case, you should update this post to include not just a description of how supporters of Bayesianism don't understand Popper, but why they should care about this discussion--why Bayesianism is not, as it seems, obviously the correct answer to the question Popper is trying to answer.

If you want to make bets about the future, Bayesianism will beat whatever else you could use. To suggest that something else is an improved method of doing science is nothing more than to suggest that it is a more feasible approximation to Bayesianism. These things are mathematical facts, if you define Bayesianism and "winning" precisely.

It seems like the only possible room for debate is the choice of prior. Everyone is forced to either implicitly choose a prior or else bet in a way that is manifestly irrational. This is also a mathematical fact. The Solomonoff prior provably isn't too bad. You just have to get over the arbitrariness.

Edit: Lets make this more precise. I claim that if we play a betting game, I can reconstruct a prior from your strategy such that a Bayesian using that prior will beat you in expectation. Do you object to this mathematical statement, or do you object to the interpretation of this fact as "Bayesianism is correct"? I'm not sure which side of the fence you are on, but I suppose it must be one or the other, so if we get that sorted out maybe we can make progress.

curi 07 April 2011 12:30:32AM 1 point

I don't understand Popper's work beyond the Wikipedia summary of critical rationalism

FYI that won't work. Wikipedia doesn't understand Popper. Secondary sources promoting myths, like Jaynes did, is common. A pretty good overview is the Popper book by Bryan Magee (only like 100 pages).

without value

I posted criticisms of Jaynes' arguments (or more accurately, his assumptions). I posted an argument about support. Why don't you answer it?

You just have to get over the arbitrariness.

You are basically admitting that your epistemology is wrong. Given that Popper has an epistemology which does not have this feature, and the rejections of him by Bayesians are unscholarly mistakes, you should be interested in it!

Of course if I wrote up his whole epistemology and posted it here for you that would be nice. But that would take a long time, and it would repeat content from his books.

If you want somewhere to start online, you could read

<http://fallibleideas.com/>

If you want to make bets about the future

That is not primarily what we want. And what you're doing here is conflating Bayes' theorem (which is about probability, and which is a matter of logic, and which is correct) with Bayesian epistemology (the application of Bayes' theorem to epistemological problems, rather than to the math behind betting).

To suggest that something else is an improved method of doing science is nothing more than to suggest that it is a more feasible approximation to Bayesianism. These things are mathematical facts,

Are you open to the possibility that the general outline of your approach is itself mistaken, and there the theorems you have proven within your framework of assumptions are therefore not all true? Or:

It seems like the only possible room for debate is the choice of prior.

Are you so sure of yourself -- that you are right about many things -- that you will dismiss all rival ideas without even having to know what they say? Even when they offer things your approach doesn't have, such as *not having arbitrary foundations*.

What you're doing is accepting ideas which have been popular since Aristotle. When you think no other ways are possible, that's bias talking. Your ideas have become common sense (not the Bayes part, but the philosophical approach to epistemology you are taking which comes before you use Bayes's theorem at all).

Here let me ask you a question: has any Bayesian ever published any substantive criticism of an important idea in Popper's epistemology? Someone should have done it, right? And if no one ever has, then you should be interested in investigating, right? And also interested in investigating what is wrong with your movement that it never addressed rival ideas in scholarly debate. (I have looked for such a criticism. Never managed to find one.)

Peterdjones 15 April 2011 03:14:18PM 3 points

Why don't you fix the WP article?

timtyler 07 April 2011 12:48:52PM * 1 point

what you're doing here is conflating Bayes' theorem (which is about probability, and which is a matter of logic, and which is correct) with Bayesian epistemology (the application of Bayes' theorem to epistemological problems, rather than to the math behind betting).

That's because to a Bayesian, these things are the same thing. Epistemology is all about probability - and visa versa. Bayes's theorem includes induction and confirmation. You can't *accept* Bayes's theorem and *reject* induction without crazy inconsistency - and Bayes's theorem is just the math of probability theory.

[deleted] 07 April 2011 01:05:19PM 0 points

If I understand correctly, I think curi is saying that there's no *reason* for probability and epistemology to be the same thing. That said, I don't entirely understand his/her argument in this thread, as some of the criticisms he/she mentions are vague. For example, what are these "epistemological problems" that Popper solves but Bayes doesn't?

paulfchristiano 07 April 2011 01:12:55AM * 2 points

Having read the website you linked to in its entirety, I think we should defer this discussion (as a community) until the next time you explain why someone's particular belief is wrong, at which point you will be forced to make an actual claim which can be rejected.

In particular, if you ever try to make a claim of the form "You should not believe X, because Bayesianism is wrong, and undesirable Y will happen if you act on this belief" then I would be interested in the resulting discussion. We could do the same thing now, I guess, if you want to make such a claim of some historical decision.

Edit: changed wording to be less of an ass.

curi 07 April 2011 01:24:01AM 2 points

In its entirety? Assuming you spent 40 minutes reading, 0 minutes delay before you saw my post, 0 minutes reading my post here, and 2:23 writing your reply, then you read at a speed of around 833 words per minute. That is very impressive. Where did you learn to do that? How can I learn to do that too?

Given that I do make claims on my website, I wonder why you don't pick one and point out something you think is wrong with it.

paulfchristiano 07 April 2011 01:33:16AM * 2 points

Fair, fair. I should have thought more and been less heated. (My initial response was even worse!)

I did read the parts of your website that relate to the question at hand. I do skim at several hundred words per minute (in much more detail than was needed for this application), though I did not spend the entire time reading. Much of the content of the website (perfectly reasonably) is devoted to things not really germane to this discussion.

If you really want (because I am constitutively incapable of letting an argument on the internet go) you could point to a particular claim you make, of the form I asked for. My issue is not really that I have an objection to any of your arguments--its that you seem to offer no concrete points where your epistemology leads to a different conclusion than Bayesianism, or in which Bayesianism will get you into trouble. I don't think this is necessarily a flaw with your website--presumably it was not designed first and foremost as a response to Bayesianism--but given this observation I would rather defer discussion until such a claim does come up and I can argue in a more concrete way.

To be clear, what I am looking for is a statement of the form: "Based on Bayesian reasoning, you conclude that there is a 50% chance that a singularity will occur by 2060. This is a dangerous and wrong belief. By acting on it you will do damage. I would not believe such a thing because of my improved epistemology. Here is why my belief is more correct, and why your belief will do damage." Or whatever example it is you would like to use. Any example at all. Even an argument that Bayesian reasoning with the Solomonoff prior has been "wrong" where Popper would be clearly "right" at any historical point would be good enough to argue about.

curi 07 April 2011 01:47:06AM * 0 points

statement of the form: "Based on Bayesian reasoning, you conclude that there is a 50% chance that a singularity will occur by 2060. This is a dangerous and wrong belief. By acting on it you will do damage I would not believe such a thing because of my improved epistemology.

Do you assert that? It is wrong and has real world consequence. In *The Beginning of Infinity* Deutsch takes on a claim of a similar type (50% probability of humanity surviving the next century) using Popperian epistemology. You can find Deutsch explaining some of that material here: <http://groupspaces.com/oxfordtranshumanists/pages/past-talks>

While Fallible Ideas does not comment on Bayesian Epistemology directly, it takes a different approach. You do not find Bayesians advocating the same ways of thinking. They have a different (worse, IMO) emphasis.

I wonder if you think that all mathematically equivalent ways of thinking are equal. I believe they aren't because some are more convenient, some get to answers more directly, some make it harder to make mistakes, and so on. So even if my approach was compatible with the Bayesian approach, that wouldn't mean we agree or have nothing to discuss.

Fair, fair. I should have thought more and been less heated. (My initial response was even worse!)

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference? You learned Bayesian stuff but it apparently didn't solve your problem, whereas my epistemology did solve mine.

paulchristiano 07 April 2011 02:09:39AM 3 points

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference?

No. It provides an example of a way in which you are better than me. I am overwhelmingly confident that I can find ways in which I am better than you.

Do you assert that? It is wrong and has real world consequence. In *The Beginning of Infinity* Deutsch takes on a claim of a similar type (50% probability of humanity surviving the next century) using Popperian epistemology. You can find Deutsch explaining some of that material here: <http://groupspaces.com/oxfordtranshumanists/pages/past-talks>

Could you explain how a Popperian disputes such an assertion? Through only my own fault, I can't listen to an mp3 right now.

My understanding is that anyone would make that argument in the same way: by providing evidence in the Bayesian sense, which would convince a Bayesian. What I am really asking for is a description of why your beliefs aren't the same as mine but better. Why is it that a Popperian disagrees with a Bayesian in this case? What argument do they accept that a Bayesian wouldn't? What is the corresponding calculation a Popperian does when he has to decide how to gamble with the lives of six billion people on an uncertain assertion?

I wonder if you think that all mathematically equivalent ways of thinking are equal. I believe they aren't because some are more convenient, some get to answers more directly, some make it harder to make mistakes, and so on. So even if my approach was compatible with the Bayesian approach, that wouldn't mean we agree or have nothing to discuss.

I agree that different ways of thinking can be better or worse even when they come to the same conclusions. You seem to be arguing that Bayesianism is wrong, which is a very different thing. At best, you seem to be claiming that trying to come up with probabilities is a bad idea. I don't yet understand exactly what you mean. Would you never take a bet? Would never take an action that could possibly be bad and could possibly be good, which requires weighing two uncertain outcomes?

This brings me back to my initial query: give a specific case where Popperian reasoning diverges from Bayesian reasoning, explain why they diverge, and explain why Bayesianism is wrong. Explain why Bayesian's willingness to bet does harm. Explain why Bayesians are slower than Popperians at coming to the same conclusion. Whatever you want.

I do not plan to continue this discussion except in the pursuit of an example about which we could actually argue productively.

curi 07 April 2011 02:46:51AM 0 points



Could you explain how a Popperian disputes such an assertion? [(50% probability of humanity surviving the next century)]

e.g. by pointing out that whether we do or don't survive depends on human choices, which in turn depends on human knowledge. And the growth of knowledge is not predictable (exactly or probabilistically). If we knew its contents and effects now, we would already have that knowledge. So this is not prediction but prophecy. And prophecy has built in bias towards pessimism: because we can't make predictions about future knowledge, prophets in general make predictions that disregard future knowledge. These are explanatory, philosophical arguments which do not rely on evidence (that is appropriate because it is not a scientific or empirical mistake being criticized). No corresponding calculation is made at all.

You ask about how Popperians make decisions if not with such calculations. Well, say we want to decide if we should build a lot more nuclear power plants. This could be taken as gambling with a lot of lives, and maybe even all of them. Of course, *not* doing it could also be taken as a way of gambling with lives. There's no way to never face any potential dangers. So, how do Popperians decide? They *conjecture* an answer, e.g. "yes". Actually, they make many conjectures, e.g. also "no". Then they criticize the conjectures, and make more conjectures. So for example I would criticize "yes" for not providing enough explanatory detail about why it's a good idea. Thus "yes" would be rejected, but a variant of it like "yes, because nuclear power plants are safe, clean, and efficient, and all the criticisms of them are from silly luddites" would be better. If I didn't understand all the references to longer arguments being made there, I would criticize it and ask for the details. Meanwhile the "no" answer and its variants will get refuted by criticism. Sometimes entire infinite categories of conjectures will be refuted by a criticism, e.g. the anti-nuclear people might start arguing with conspiracy theories. By providing a general purpose argument against all conspiracy theories, I could deal with all their arguments of that type. Does this illustrate the general idea for you?

You seem to be arguing that Bayesianism is wrong, which is a very different thing.

I think it's wrong as an epistemology. For example because induction is wrong, and the notion of positive support is wrong. Of course Bayes' theorem is correct, and various math you guys have done is correct. I keep getting conflicting statements from people about whether Bayesianism conflicts with Popperism or not, and I don't want to speak for you guys, nor do I

want to discourage anyone from finding the shared ideas or discourage them from learning from both.

Would you never take a bet?

Bets are made on *events*, like which team wins a sports game. Probabilities are fine for *events*. Probabilities of *the truth of theories* is problematic (b/c e.g. there is no way to make them non-arbitrary). And it's not something a fallibilist can bet on because he accepts we never know the final truth for sure, so how are we to set up a decision procedure that decides who won the bet?

Would never take an action that could possibly be bad and could possibly be good, which requires weighing two uncertain outcomes?

We are not afraid of uncertainty. Popperian epistemology is fallibilist. It rejects certainty. Life is always uncertain. That does not imply probability is the right way to approach all types of uncertainty.

This brings me back to my initial query: give a specific case where Popperian reasoning diverges from Bayesian reasoning, explain why they diverge, and explain why Bayesianism is wrong. Explain why Bayesian's willingness to bet does harm. Explain why Bayesians are slower than Popperians at coming to the same conclusion. Whatever you want.

Bayesian reasoning diverges when it says that ideas can be positively supported. We diverge because Popper questioned the concept of positive support, as I posted in the original text on this page, and which no one has answered yet. The criticism of positive support begins by considering what it is (you tell me) and how it differs from consistency (you tell me).

Larks 08 April 2011 01:03:10AM 0 points

And the growth of knowledge is not predictable (exactly or probabilistically). If we knew its contents and effects now, we would already have that knowledge.

You're equivocating between "knowing exactly the contents of the new knowledge", which may be impossible for the reason you describe, and "know some things about the effect of the new knowledge", which we can do. As Eliezer said, I may not know which move Kasparov will make, but I know he will win.

jake987722 07 April 2011 03:24:11AM 6 points

So, how do Popperians decide? They conjecture an answer, e.g. "yes". Actually, they make many conjectures, e.g. also "no". Then they criticize the conjectures, and make more conjectures. So for example I would criticize "yes" for not providing enough explanatory detail about why it's a good idea. Thus "yes" would be rejected, but a variant of it like "yes, because nuclear power plants are safe, clean, and efficient, and all the criticisms of them are from silly luddites" would be better. If I didn't understand all the references to longer arguments being made there, I would criticize it and ask for the details. Meanwhile the "no" answer and its variants will get refuted by criticism. Sometimes entire infinite categories of conjectures will be refuted by a criticism, e.g. the anti-nuclear people might start arguing with conspiracy theories. By providing a general purpose argument against all conspiracy theories, I could deal with all their arguments of that type. Does this illustrate the general idea for you?

Almost, but you seem to have left out the rather important detail of how actually *make the decision*. Based on the process of criticizing conjectures you've described so far, it seems that there are two basic routes you can take to finish the decision process once the critical smoke has cleared.

First, you can declare that, since there is no such thing as confirmation, it turns out that no conjecture is better or worse than any other. In this way you don't actually make a decision and the problem remains unsolved.

Second, you can choose to go with the conjecture that best weathered the criticisms you were able to muster. That's fine, but then it's not clear that you've done anything different

from what a Bayesian would have done--you've simply avoided explicitly talking about things like probabilities and priors.

Which of these is a more accurate characterization of the Popperian decision process? Or is it something radically different from these two altogether?

curi 07 April 2011 03:59:34AM 2 points 

When you have exactly one non-refuted theory, you go with that.

The other cases are more complicated and difficult to understand.

Suppose I gave you the answer to the other cases, and we talked about it enough for you to understand it. What would you change your mind about? What would you concede?

If I convinced you of this one single issue (that there is a method for making the decision), would you follow up with a thousand other objections to Popperian epistemology, or would we have gotten somewhere?

If you have lots of other objections you are interested in, I would suggest you just accept for now that we have a method and focus on the other issues first.

[option 1] since there is no such thing as confirmation, it turns out that no conjecture is better or worse than any other.

But some are criticized and some aren't.

[option 2] conjecture that best weathered the criticisms you were able to muster

But how is that to be judged?

No, we always go with uncriticized ideas (which may be close variants of ideas that were criticized). Even the terminology is very tricky here -- the English language is not well adapted to expressing these ideas. (In particular, the concept "uncriticized" is a very substantive one with a lot of meaning, and the word for it may be misleading, but other words are even worse. And the straightforward meaning is OK for present purposes, but may be problematic in future discussion.).

[Or is it something radically different from these two altogether?

Yes, different. Both of these are justificationist ways of thinking. They consider how much justification each theory has. The first one rejects a standard source of justification, does not replace it, and ends up stuck. The second one replaces it, and ends up, as you say, reasonably similar to Bayesianism. It still uses the same basic method of tallying up how much of some good thing (which we call justification) each theory has, and then judging by what has the most.

Popperian epistemology does not justify. It uses criticism for a different purpose: a criticism is an explanation of a mistake. By finding mistakes, and explaining what the mistakes are, and conjecturing better ideas which we think won't have those mistakes, we learn and improve our knowledge.

[continue this thread »](#)

Desrtopa 07 April 2011 01:58:08AM 4 points 

Using my epistemology I have learned not to do that kind of thing. Would that serve as an example of a practical benefit of it, and a substantive difference? You learned Bayesian stuff but it apparently didn't solve your problem, whereas my epistemology did solve mine.

It doesn't take Popperian epistemology to learn social fluency. I've learned to limit conflict and improve the productivity of my discussions, and I am (to the best of my ability) Bayesian in my epistemology.

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

curi 07 April 2011 02:11:59AM -4 points

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

I have considered that. Popperian epistemology helps with these issues more. I don't want to argue about that now because it is an advanced topic and you don't know enough about my epistemology to understand it (correct me if I'm wrong), but I thought the example could help make a point to the person I was speaking to.

Desrtopa 07 April 2011 02:15:42AM * 2 points

If I don't understand your explanation and am interested in it, I'm prepared to do the research in order to understand it, but if you can only assert why your epistemology *should* result in better social learning and not demonstrate that it does so for people in general, I confess that I will probably not be interested enough to follow up.

I will note though, that stating the assumption that another does not understand, but leaving them free to correct you, strikes me as a markedly worse way to minimize conflict and aggression than *asking* if they have the familiarity necessary to understand the explanation.

curi 07 April 2011 02:25:58AM 1 point

You could begin by reading

<http://fallibleideas.com/emotions>

And the rest of the site. If you don't understand any connections between it and Popperian epistemology, feel free to ask.

I'm not asking you to be interested in this, but I do think you should have some interest in rival epistemologies.

[continue this thread »](#)

JoshuaZ 07 April 2011 02:07:05AM 2 points

If you want to credit a particular skill to your epistemology, you should first see whether it's more likely to arise among those who share your epistemology than those who don't.

That's a claim that only makes sense in certain epistemological systems...

curi 07 April 2011 02:09:13AM * 3 points

I don't have a problem with the main substance of that argument, which I agree with. Your implication that we would reject this idea is mistaken.

JoshuaZ 07 April 2011 02:36:12AM 0 points

I don't have a problem with the main substance of that argument, which I agree with. Your implication that we would reject this idea is mistaken.

Hmm? I'm not sure who you mean by we? If you mean that someone supporting a Popperian approach to epistemology would probably find this idea reasonable than I agree with you (at least empirically, people claiming to support some form of Popperian approach seem ok with this sort of thing. That's not to say I understand how they think it is implied/ok in a Popperian framework).

paulfchristiano 07 April 2011 01:02:58AM 9 points

Here let me ask you a question: has any Bayesian ever published any substantive criticism of an important idea in Popper's epistemology? Someone should have done it, right?

Most things in the space of possible documents can't be refuted, because they don't correspond to anything refutable. They are simply confused, and irredeemably. In the case of epistemology, virtually everything that has ever been said falls into this category. I am glad that I don't have to spend time thinking about it, because it is *solved*. I would not generally criticize a rival's ideas, because I no longer care. The problem is solved, and I can go work on things that still matter.

Are you so sure of yourself -- that you are right about many things -- that you will dismiss all rival ideas without even having to know what they say?

Once I know the definitive answer to a question, I will dismiss all other answers (rather than trying to poke holes in them). The only sort of argument which warrants response is an objection to my current definitive answer. So ignorance of Popper is essentially irrelevant (and I suspect I couldn't object to anything in his philosophy, because it has essentially no content concrete enough to be defeated by mere reasoning).

The real question, in fact the only question, is whether the arbitrariness of choosing a prior can be surmounted--whether my current answer is not actually definitive. If someone came to me and said they had a solution to this problem I would be interested, except that I am fairly confident the problem has no solution for what are essentially obvious reasons. Popper avoids this problem by not even describing his epistemology precisely enough to express the difficulty.

Really this entire discussion comes down to what we want out of epistemology.

That [guiding betting] is not primarily what we want.

What do you want? I don't understand at all. Whatever you specify, I would be shocked if critical rationality provided it. Here is what I want, and maybe you will agree:

I want to decide between action A and action B. To do this, I want to evaluate the consequences of action A and action B. To do this, I want to predict something about the world. In particular, by choosing B instead of A, I am making a bet about the consequences of A and B. I would like to make such bets in the best possible way.

Lo! This is precisely what Bayesianism allows me to do. Why is there more to say?

You can object that it involves knowing a prior. But from the problem statement it is *obvious* (as a mathematical fact) that there is a universe in which each possible prior is the best one. Is there a strategy that does better than Bayesianism with a reasonable prior in all possible universes? Maybe, but Popper's ideas aren't nearly precise enough to answer the question (by which I mean, not even at the point where this question, to me clearly the most important one, is meaningful). Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

If I have to bet, or make a decision that effects peoples lives which amounts to a bet, I am going to use Bayesianism, or a computational heuristic which I justify by Bayesianism. Doing something else seems irresponsible.

curi 07 April 2011 02:07:06AM 1 point

Most things in the space of possible documents can't be refuted, because they don't correspond to anything refutable. They are simply confused, and irredeemably.

You don't think confused things can be criticized? You can, for example, point out ambiguous passages. That would be a criticism. If they have no clarification to offer, then it would be (tentatively and fallibly) decisive (pending some reason to reconsider).

But you haven't provided any argument that Popper in particular was confused, irrefutable, or whatever. I don't know about you, but as someone who wants to improve my epistemological knowledge I think it's important to consider all the major ideas in the field at the very least enough to know one good criticism of each.

Refusing to address criticism because you think you already have the solution is very closed minded, is it not? You think you're done with thinking, you have the final truth, and that's that..?

The only sort of argument which warrants response is an objection to my current definitive answer.

Popper published several of those. Where's the response from Bayesians?

One thing to note is it's hard to understand his objections without understanding his philosophy a bit more broadly (or you will misread stuff, not knowing the broader context of what he is trying to say, what assumptions he does not share with you, etc...)

The real question, in fact the only question, is whether the arbitrariness of choosing a prior can be surmounted--whether my current answer is not actually definitive. If someone came to me and said they had a solution to this problem I would be interested

Popper solved that problem.

I am fairly confident the problem has no solution for what are essentially obvious reasons

The standard reasons seem obvious because of your cultural bias. Since Aristotle some philosophical assumptions have been taken for granted by almost everyone. Now most people regard them as obvious. Given those assumptions, I agree that your conclusion follows (no way to avoid arbitrariness). The assumptions are called "justificationism" by Popperians, and are criticized in detail. I think you ought to be interested in this.

One criticism of justificationism is that it *causes* the regress/arbitrariness/foundations problem. The problem doesn't exist automatically but is being created by your own assumptions.

Popper avoids this problem by not even describing his epistemology precisely enough to express the difficulty.

What are you talking about? You haven't read his books and claim he didn't give enough detail? He was something of a workaholic who didn't watch TV, didn't have a big social life, and worked and wrote all the time.

What do you want?

To create knowledge, including explanatory and non-instrumentalist knowledge. You come off like a borderline positivist to me, who has trouble with the notion that non-empirical stuff is even meaningful. (No offense intended, and I'm not assuming you actually are a positivist, but I'm not really seeing much difference yet.)

To do this, I want to evaluate the consequences of action A and action B. To do this, I want to predict something about the world.

To take one issue, besides predicting the physical results of your actions you also need a way to judge which results are good or bad. That is moral knowledge. I don't think Bayesianism addresses this well.

Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

Neither. You can and should do better!

David_Allen 07 April 2011 04:16:38PM 0 points



To take one issue, besides predicting the physical results of your actions you also need a way to judge which results are good or bad. That is moral knowledge. I don't think Bayesianism addresses this well.

Given well defined contexts and meanings for good and bad I don't see why Bayesianism could not be effectively applied to moral problems.

curi 07 April 2011 06:40:28PM 0 points



Yes, *given moral assertions* you can then analyze them. Well, sort of. You guys rely on empirical evidence. Most moral arguments don't.

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

David_Allen 07 April 2011 08:26:10PM 0 points



Yes, given moral assertions you can then analyze them. Well, sort of. You guys rely on empirical evidence. Most moral arguments don't.

First of all, you shouldn't lump me in with the Yudkowskyist Bayesians. Compared to them and to you I am in a distinct third party on epistemology.

Bayes' theorem is an abstraction. If you don't have a reasonable way to transform your problem to a form valid within that abstraction then of course you shouldn't use it. Also, if you have a problem that is solved more efficiently using another abstraction, then use that other abstraction.

This doesn't mean that Bayes' theorem is useless, it just means there are domains of reasonable usage. The same will be true for your [Popperian decision making](#).

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

These are just computable processes; if Bayesianism is in some sense [Turing complete](#) then it can be used to do all of this; it just might be very inefficient when compared to other approaches.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods. Other aspects should probably be accomplished using other methods.

curi 07 April 2011 08:41:46PM 0 points



First of all, you shouldn't lump me in with the Yudkowskyist Bayesians. Compared to them and to you I am in a distinct third party on epistemology.

Sorry. I have no idea who is who. Don't mind me.

This doesn't mean that Bayes' theorem is useless, it just means there are domains of reasonable usage. The same will be true for your Popperian decision making.

The Popperian method is universal.

if Bayesianism is in some sense Turing complete then it can be used to do all of this

Well, umm, yes but that's no help. my iMac is definitely Turing complete. It could run an AI. It could do whatever. But we don't know how to make it do that stuff. Epistemology should help us.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods.

Example or details?

David_Allen 07 April 2011 09:59:13PM 0 points



Sorry. I have no idea who is who. Don't mind me.

No problem, I'm just pointing out that there are other perspectives out here.

The Popperian method is universal.

Sure, in the sense it is Turing complete; but that doesn't make it the most efficient approach for all cases. For example I'm not going to use it to decide the answer to the statement " $2 + 3$ ", it is much more efficient for me to use the arithmetic abstraction.

But we don't know how to make it do that stuff. Epistemology should help us.

Agreed, it is one of the reasons that I am actively working on epistemology.

Aspects of coming up with moral ideas and judging which ones are good would probably be accomplished well with Bayesian methods.

Example or details?

[The naive Bayes classifier](#) can be an effective way to classify discrete input into independent classes. Certainly for some cases it could be used to classify something as "good" or "bad" based on example input.

[Bayesian networks](#) can capture the meaning within interdependent sets. For example the meaning of words forms a complex network; if the meaning of a single word shifts it will probably result in changes to the meanings of related words; and in a similar way ideas on morality form connected interdependent structures.

Within a culture a particular moral position may be dependent on other moral positions, or even other aspects of the culture. For example a combination of religious beliefs and inheritance traditions might result in a belief that a husband is justified in killing an unfaithful wife. A Bayesian network trained on information across cultures might be able to identify these kinds of relationships. With this you could start to answer questions like "Why is X moral in the UK but not in Saudi Arabia?"

curi 08 April 2011 12:37:39AM 0 points

Sure, in the sense it is Turing complete;

No, in the sense that it *directly* applies to all types of knowledge (which any epistemology applies to -- which i think is all of them, but that doesn't matter to universality).

Not in the sense that it's Turing complete so you could, by a roundabout way and using whatever methods, do anything.

I think the basic way we differ is you have despaired of philosophy getting anywhere, and you're trying to get rigor from math. But Popper saved philosophy. (And most people didn't notice.) Example:

With this you could start to answer questions like "Why is X moral in the UK but not in Saudi Arabia?"

You have very limited ambitious. You're trying to focus on small questions b/c you think bigger ones like: what *is* moral objectively? are too hard and, since you math won't answer them, it's hopeless.

[continue this thread »](#)

JoshuaZ 07 April 2011 06:58:15PM * 2 points

You can't create moral ideas in the first place, or judge which are good (without, again, assuming a moral standard that you can't evaluate).

You've repeatedly claimed that the Popperian approach can somehow address moral issues. Despite requests you've shown no details of that claim other than to say that you do the same thing you would do but with moral claims. So let's work through a specific moral issue. Can you take an example of a real moral issue that has been controversial historically (like say slavery or free speech) and show how the Popperian would approach? An concrete worked out example would be very helpful.

curi 07 April 2011 07:00:42PM * -2 points

http://lesswrong.com/lw/552/reply_to_benelliott_about_popper_issues/3uv7

And it creates moral knowledge by conjecture and refutation, same as any other knowledge. If you understand how Popper approaches any kind of knowledge (which I have written about a bunch here), then you know how he approaches moral knowledge too.

JoshuaZ 07 April 2011 07:10:36PM 0 points

And it creates moral knowledge by conjecture and refutation, same as any other knowledge. If you understand how Popper approaches any kind of knowledge (which I have written about a bunch here), then you know how he approaches moral knowledge too.

Consider that you are replying to a statement I just said that all you've done is say that it would use the same methodologies. Given that, does this reply seem sufficient? Do I need to repeat my request for a worked example (which is not included in your link)?

[deleted] 07 April 2011 01:50:34AM -4 points

Should I use a theory which I understand and which has an apparently necessary flaw, or a theory which is underspecified and therefore "avoids" this difficulty?

Saying your epistemology has a "necessary flaw" is an admission of defeat, that it doesn't work. The "necessary flaw" is unavoidable if you are committed to the justificationist way of thinking. Popper saw that the whole idea of justification is wrong and he offered a different idea to replace it - an idea with no known flaws. You criticize Popper for being underspecified, yet he elaborated on his ideas in many books. And, furthermore, no amount of mathematical precision or formalism will paper over cracks in justificationist epistemologies.

curi 07 April 2011 01:54:23AM * -1 points

It's interesting how different Bayesians say different things. They don't seem to all agree with each other even about their basic claims. Sometimes Bayesianism is proved, other times it is acknowledged to have known flaws. Sometimes it may be completely compatible with Popper, other times it is dethroning Popper. It seems to me that perhaps Bayesianism is a bit underspecified. I wonder why they haven't sorted out these internal disputes.

JoshuaZ 07 April 2011 03:41:08AM * 3 points

Sometimes Bayesianism is proved, other times it is acknowledged to have known flaws. Sometimes it may be completely compatible with Popper, other times it is dethroning Popper. It seems to me that perhaps Bayesianism is a bit underspecified. I wonder why they haven't sorted out these internal disputes.

There are disputes among the Bayesians. But you are confusing different issues. First, the presence of internal disputes about the borders of an idea is not a priori a problem with an idea that is in progress. The fact that evolutionary biologists disagree about how much neutral drift matters isn't a reason to reject evolution. (It is possible that I'm reading an unintended implication here.)

Moreover, most of what you are talking about here are not contradictions but failure to understand. That Bayesianism has flaws is a distinct claim from when someone talks about something like Cox's theorem which is the sort of result that Bayesians are talking about that you refer to as "Sometimes Bayesianism is proved"(which incidentally is a terribly unhelpful and vague way of discussing the point). The point of results like Cox's theorem is that if one very broad attempts under certain very weak assumptions to formalize epistemology you must end up with some form of Bayesianism. At the

same time it is important to keep in mind that this isn't saying all that much. It doesn't for example say anything about what one's priors should be. Thus one has the classical disagreement between objective and subjective Bayesians based on what sort of priors to use (and within each of those there is further breakdown. LessWrong seems to mainly have objective Bayesians favoring some form Occam prior, although just what is not clear.) Similarly, when discussing whether or not Bayesianism is compatible with Popper depends a lot on what one means by "Bayesianism", "compatible" and "Popper". Bayesianism is certainly not compatible with a naive-Popperian approach, which is what many are talking about when they say that it is not compatible (and as you've already noted Popper himself wasn't a naive Popperian). But some people use Popper to mean the idea that given an interesting hypothesis one should search out for experiments which would be likely to falsify the hypothesis if it is false (an idea that actually predates Popper) but what one means by falsify can be a problem.

paulfchristiano 07 April 2011 01:53:27AM 2 points

Saying your epistemology has a "necessary flaw" is an admission of defeat,

In this case, its recognition of reality. I repeat that I would like to defer this conversation until we have something concrete to disagree about. Until then I don't care about that difference.

[deleted] 07 April 2011 02:24:53AM 0 points

The "necessary flaw" arises because all justificationist epistemologies lead to infinite regress or circular arguments or appeals to authority (or even sillier things). That you think there is no alternative to justificationism and I don't is something concrete we disagree about.

David_Allen 07 April 2011 04:02:37PM 1 point

Adding a reference for this comment: [Münchhausen Trilemma](#).

endoself 07 April 2011 12:05:45AM 2 points

The thing intended as the proof is most of chapter 2. I dislike Jaynes' assumptions there, since I find many of them superfluous compared to other proofs. You probably like them even less, since one is "Representation of degrees of plausibility by real numbers".

curi 07 April 2011 12:09:20AM * 2 points

It cannot be a proof of *Bayesian epistemology* itself if it makes assumptions like that.

It is merely a proof of some theorems in Bayesian epistemology given some premises that Bayesians like.

If you have a different proof which *does not make assumptions I disagree with*, then let's hear it. Otherwise you can give up on *proving* and start *arguing* why I should agree with your starting points. Or maybe even, say, engaging with Popper's arguments and pointing out mistakes in them (if you can find any).

Peterdjones 12 April 2011 08:31:43PM 1 point

You are complaining it is not a deduction of Bayes from no assumptions whatever. But all it needs to be is that those assumptions can be made to "work"--ie applied without contradiction, quodlibet or other disaster.

Peterdjones 15 April 2011 03:18:53PM 0 points

Remember, Popper himself said it all starts with common sense.

endoself 07 April 2011 02:42:12AM -1 points

I agree that it is by no means a complete proof of Bayesian epistemology. The book I pointed you to might have a more complete one, though I doubt it will be complete since it seems more like a book about using statistics than about rigorously understanding epistemology.

I am currently collecting the necessary knowledge to write the full proof myself, if it is possible (not because of this debate, because I kept being annoyed by unjustified assumptions that didn't even seem necessary).

curi 07 April 2011 02:55:59AM 0 points

Good luck. But, umm, do you have some argument against fallibilism? Because you're going to need one.

endoself 07 April 2011 03:35:59AM * 1 point

I think I massively overstated my intention. I meant the full proof of the stuff we know; the thing I think could be in *Mathematical Statistics, Volume 1: Basic and Selected Topics*.

Anyways, I think I accept fallibilism, at least from the Wikipedia page. Why do you think I don't? This is understandable, because I've been talking about idealized agents a lot more than about humans actually applying Bayesianism.

curi 07 April 2011 03:48:40AM 1 point

I think you are not a fallibilist because you want to prove philosophical ideas.

But we can't have certainty. So what do you even think it means to "prove" them? Why do you want to prove them instead of give good arguments on the matter?

endoself 07 April 2011 04:15:22AM * 0 points

I use the word prove because I'm doing it deductively in math. I already linked you to the $2+2=3$ thing, I believe. Also, the question of how I would, for example, change AI design if a well-known theorem is wrong (pretend it is the future and the best theorems proving Bayesianism are better-known and I am working on AI design) is both extremely hard to answer and unlikely to be necessary. Well unlikely is the wrong word; what is $P(X | \text{"There are no probabilities"})$? :)

calef 07 April 2011 05:10:07AM * 1 point

Probably the most damning criticism you'll find, curi, is that fallibilism isn't *useful* to the Bayesian.

The fundamental disagreement here is somewhere in the following statement:

"There exist true things, and we have a means of determining how likely it is for any given statement to be true. Furthermore, a statement that has a high likelihood of being true should be believed over a similar statement with a lower likelihood of being true."

I suspect your disagreement is in one of several places.

1) You disagree that there even exist epistemically "true" facts. 2) That we can determine how likely something is to be true. or 3) That likelihood of being true (as defined by us) is reason to believe the truth of something.

I can actually flesh out your objections to all of these things.

For 1, you could probably successfully argue that we aren't capable of determining if we've ever actually arrived at a true epistemic statement because real certainty doesn't exist, thus the existence or nonexistence of true epistemic statements is on the same epistemological footing as the existence of God--i.e. shaky to the point of not concerning oneself with them all together.

2 basically ties in with the above directly.

3 is a whole 'nother ball game, and I don't think it's really been broached yet by anyone, but it's certainly a valid point of contention. I'll leave it out unless you'd like to pursue it.

The Bayesian counter to all of these is simply, "That doesn't really do anything for me."

Declaring we have certainty, and quantifying it as best we can is *incredibly useful*. I can pick up an apple and let go. It will fall to the ground. I have an incredibly huge amount of certainty in my ability to repeat that experiment.

That I cannot foresee the philosophical paradigm that will uproot my hypothesis that dropped apples fall to the ground is not a very good reason to reject my relative certainty in the soundness of my hypothesis. Such a apples-aren't-falling-when-dropped paradigm would literally (and necessarily) uproot everything else we know about the world.

Basically, what I'm trying to say is that all you're ever going to get out of a Bayesian is, "No, I disagree. I think we can have certainty." And the only way you could disprove conclusions made by Bayesians are through means the Bayesian would have already seen, and thus the Bayesian would have already rejected said conclusion.

You've already outlined that the fallibilist will just keep tweaking explanations until an explanation with no criticism is reached. I think you might find Bayesianism more palatable if you just pretend that we aren't trying to find certainty, just say we're trying to minimize criticism.

This probably hasn't been a very satisfying answer. I certainly agree it's useful to have an understanding of the biases to our certainties. I also think Bayesianism happens to build that into itself quite well. Personally, I don't think there's anything I'm absolutely certain about, because to claim so would be silly.

endoseif 07 April 2011 05:32:57AM 1 point

Small nitpick: I don't like your use of the word 'certainty' here. Especially in philosophy, it has too much of a connotation of "literally impossible for me to be wrong" rather than "so ridiculously unlikely that I'm wrong that we can just ignore it", which may cause confusion.

calef 07 April 2011 05:40:16AM 0 points

Where don't you like it? I don't think anyone actually argues for your first definition, because, like I said, it's silly. I think curi's point is that fallibilism is predicated on your second definition not (ever?) being a valid claim.

My point is that the things we are "certain" about (as per your second definition) probably coincide almost exactly with "statements without criticism" as per curi's definition(s).

[continue this thread »](#)

curi 07 April 2011 04:35:29AM -3 points

I think we have very different goals, and that the Popperian ones are better.

There is more to epistemology, and to philosophy, than math.

I'd say you are practically trying to *eliminate all philosophy*. And that saying you have an epistemology at all is very misleading, because epistemology is a philosophical field.

JoshuaZ 07 April 2011 05:24:07AM * 3 points

I think we have very different goals, and that the Popperian ones are better.

So could you be more precise in how you think the goals differ and why the Popperian goals are better?

There is more to epistemology, and to philosophy, than math.

I'd say you are practically trying to eliminate all philosophy. And that saying you have an epistemology at all is very misleading, because epistemology is a philosophical field.

Huh? Do you mean that because the Bayesians have made precise mathematical claims it somehow ceases to be an epistemological system? What does that even mean? I don't incidentally know what it means to eliminate philosophy, but areas can certainly be carved off from philosophy into other branches. Indeed, this is generally what happens. Philosophy is the big grab bag of things that we don't have a very good precise feel for. As we get more precise understanding things break off. For example, biology broke off from philosophy (when it broke off isn't clear, but certainly by 1900 it was a separate field) with the philosophers now only focusing on the remaining tough issues like how to define "species". Similarly, economics broke off. Again, where it broke off is tough (that's why Bentham and Adam Smith are often both classified as philosophers). A recent break off has been psychology, which some might argue is still in the process. One thing that most people would still see as clearly in the philosophy realm is moral reasoning. Indeed, some would argue that the ultimate goal of philosophy is to eliminate itself.

If it helps at all, in claiming that the Bayesians lack an epistemology or are not trying to philosophy it might help to taboo both epistemology and philosophy and restate those statements. What do those claims mean in a precise way?

curi 07 April 2011 05:33:11AM 0 points



Different people are telling me different things. I have been told some very strong instrumentalist and anti-philosophy arguments in my discussions here. I don't know just how representative of all Bayesians that is.

For example, moral philosophy has been trashed by everyone who spoke to me about it so far. I get told its meaningless, or that Bayesian epistemology cannot create moral knowledge. No one has yet been like "oh my god, epistemology *should* be able to create moral and other philosophical (non-empirical, non-observational) knowledge! Bayesian stuff is wrong since it can't!" Rather, people don't seem to mind, and will argue at length that e.g. explanatory knowledge and non-empirical knowledge don't exist or are worthless and prediction is everything.

By "philosophy" I mean things which can't be experimentally/empirically tested (as opposed to "science" by which I mean things that can be). So for philosophy, no observations are directly relevant.

Make sense? Where do you stand on these issues?

And the way I think Popperian goals are better is that they value explanations which help us understand the world instead of being instrumentalist, positivist, anti-philosophical, or anything like that.

[continue this thread »](#)

endoself 07 April 2011 05:10:34AM 0 points



In order for any philosophy to be valid, the human brain must be able to evaluate deductive arguments; they are a huge component of philosophy, with many often being needed to argue a single idea. Wondering what to do in case these are wrong is not only unnecessary but impossible.

curi 07 April 2011 05:22:01AM 2 points



I don't have any criticism of deductive logic itself. But I do have criticisms of some of the premises i expect you to use. For example, they won't all be deductively argued for themselves. That raises the problem of: how will you sort out good ideas from bad ideas for use as premises? That gets into various proposed solutions to that problem, such as induction or Popperian epistemology. But if you get into that, right in the premises of your

supposed proof, then it won't be much of a proof because so much substantive content in the premises will be non-deductive.

[continue this thread »](#)



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [Delete](#) | [jaynes](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 10:12:27PM 1 point

In short, if you want to take seriously the analogy

I wasn't talking about an analogy.

Evolution is a theory which applies to any type of replicator. Not by analogy by literally applies.

Make sense so far?

[deleted] 07 April 2011 10:22:10PM 1 point

That only strengthens my argument.

curi 07 April 2011 10:24:48PM * 0 points

You said we were discussing an analogy. That was a mistake. How can having made a mistake strength your argument? When you make a mistake, and find out, you should be like "uh oh. maybe i made 2. or 3. i better rethink things a bit more carefully. maybe the mistake is caused by a misunderstanding that could cause multiple mistakes." I don't think glossing over mistakes is rational or wise.

Make sense so far?

Randaly 08 April 2011 02:40:12PM 4 points

Because if there is only an analogy between evolution and knowledge acquisition, there are some aspects of each that do are not the same, and it is possible that these differences mean that the specific factor under consideration is not the same; but if the two processes are literally the same, that is not possible.

"How can having a mistake strengthen your argument?"

Example: During WWII, many American leaders didn't believe that Germany was actually committing massacres, as they were disillusioned from similar but inaccurate WWI propaganda; however, they still believed that Nazi aggression was morally wrong. Later, the death camps were discovered. Clearly, given that they were mistaken in disbelieving in the Holocaust, they were mistaken in believing that the Nazis were morally wrong- because how can making a mistake strength your argument?

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 02:54:20AM 1 point

Causality isn't explanation. X caused Y isn't the issue I was talking about.

For example, the statement "Murder is bad because it is illiberal" is an explanation of *why* it is bad. It is not a statement about causality.

You may say that "illiberal" is a short cut for various other ideas. And you may claim that eventually that reduce away to causal issues. But that would be reductionism. We do not accept that high level concepts are a mistake or that emergence isn't important.

JoshuaZ 07 April 2011 03:02:20AM 0 points

Huh? It may be that I haven't read Logic of Scientific Discovery in a long time, but as far as I remember/can tell, Popper doesn't care about moral whys like "why is murder bad" at all. That seems to be an issue generally independent of both Bayesian and Popperian epistemology. One could be a Bayesian and be a utilitarian, or a virtue ethicist, or some form of deontologist. What am I missing?

curi 07 April 2011 03:09:00AM * 3 points

Huh? It may be that I haven't read Logic of Scientific Discovery in a long time, but as far as I remember/can tell, Popper doesn't care about moral whys like "why is murder bad" at all.

He doesn't discuss them in LScD (as far as I remember). He does elsewhere, e.g. in The World of Parmenides. Whether he published moral arguments or not, his epistemology applies to them and works with them -- it is general purpose.

Epistemology is about how we get knowledge. Any epistemology which can't deal with entire categories of knowledge has a big problem. It would mean a second epistemology would be needed for that other category of knowledge. And that would raise questions like: if this second one works where the first failed, why not use it for everything?

Popper's method does not rely on only empirical criticism but also allows for all types of philosophical criticism. So it's not restricted to only empirical issues.

ShardPhoenix 07 April 2011 04:38:13AM * 1 point

You seem to be assuming that "morality" is a fact about the universe. Most people here think it's a fact about human minds.

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

(ie we aren't moral realists, at least not in the sense that a religious person is).

curi 07 April 2011 04:40:28AM -2 points

Yes, morality is objective.

I don't want to argue terminology.

There are objective facts about how to live, call them what you will. Or, maybe you'll say there aren't. If there are, then it's not objectively wrong to be a mass murderer. Do you really want to go there into full blown relativism and subjectivism?

ShardPhoenix 07 April 2011 04:42:47AM 5 points

Well, that's just like, your opinion, man.

Seriously: Morality is in the brain. Murder is "wrong" because I, and people sufficiently similar to me, don't like it. There's nothing more objective about it than any of my other opinions and desires. If you can't even agree on this, then coming here and arguing is hopeless - you might as well be a Christian and try to tell us to believe in God.

zaph 07 April 2011 02:32:45PM * 0 points

Well stated. And I would further add that there are issues with significant minority interests that staunchly disagree with majority opinion. Take the debates on homosexual marriage or abortion. The various sides have such different viewpoints that there isn't a common ground where any agreeably objective position can be reached. The "we all agree mass murder is wrong" is a cop out, because it implies all moral questions are that black and white. And even then, if it's such a universal moral, why does it happen in the first place? In the brain based morality model, I can say Dennis Rader's just a substantially different brain. With universal morality, you're stuck with the problem of people knowing something is wrong, but doing it anyway.

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- gwern (32)
- ingres (30)
- Lumifer (27)
- turchin (23)
- Habryka (22)
- Yosarian2 (21)
- IlyaShpitser (19)
- SaidAchmiz (19)
- gjm (19)
- fowlertm (17)
- SquirrelInHell (15)
- Dagon (15)
- Elo (14)
- Stuart_Armstrong (13)
- Gunnar_Zarncke (12)

RECENT KARMA AWARDS



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 03:59:34AM 2 points

When you have exactly one non-refuted theory, you go with that.

The other cases are more complicated and difficult to understand.

Suppose I gave you the answer to the other cases, and we talked about it enough for you to understand it. What would you change your mind about? What would you concede?

If i convinced you of this one single issue (that there is a method for making the decision), would you follow up with a thousand other objections to Popperian epistemology, or would we have gotten somewhere?

If you have lots of other objections you are interested in, I would suggest you just accept for now that we have a method and focus on the other issues first.

[option 1] since there is no such thing as confirmation, it turns out that no conjecture is better or worse than any other.

But some are criticized and some aren't.

[option 2] conjecture that best weathered the criticisms you were able to muster

But how is that to be judged?

No, we always go with uncriticized ideas (which may be close variants of ideas that were criticized). Even the terminology is very tricky here -- the English language is not well adapted to expressing these ideas. (In particular, the concept "uncriticized" is a very substantive one with a lot of meaning, and the word for it may be misleading, but other words are even worse. And the straightforward meaning is OK for present purposes, but may be problematic in future discussion.)

Or is it something radically different from these two altogether?

Yes, different. Both of these are justificationist ways of thinking. They consider how much justification each theory has. The first one rejects a standard source of justification, does not replace it, and ends up stuck. The second one replaces it, and ends up, as you say, reasonably similar to Bayesianism. It still uses the same basic method of tallying up how much of some good thing (which we call justification) each theory has, and then judging by what has the most.

Popperian epistemology does not justify. It uses criticism for a different purpose: a criticism is an explanation of a mistake. By finding mistakes, and explaining what the mistakes are, and conjecturing better ideas which we think won't have those mistakes, we learn and improve our knowledge.

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

jake987722 07 April 2011 04:37:27AM 0 points



If I convinced you of this one single issue (that there is a method for making the decision), would you follow up with a thousand other objections to Popperian epistemology, or would we have gotten somewhere?

Yes, we will have gotten somewhere. This issue is my primary criticism of Popperian epistemology. That is, given what I understand about the set of ideas, it is not clear to me how we would go about making practical scientific decisions. With that said, I can't reasonably guarantee that I will not have later objections as well before we've even had the discussion!

So let me see if I'm understanding this correctly. What we are looking for is the one conjecture which appears to be completely impervious to any criticism that we can muster against it, given our current knowledge. Once we have found such a conjecture, we -- I don't want to say "assume that it's true," because that's probably not correct -- we *behave as if it were true* until it finally is criticized and, hopefully, replaced by a new conjecture. Is that basically right?

I'm not really seeing how this is fundamentally anti-justificationist. It seems to me that the Popperian epistemology still depends on a form of justification, but that it relies on a sort of boolean all-or-nothing justification rather than allowing graded degrees of justification. For example, when we say something like, "in order to make a decision, we need to have a guiding theory which is currently impervious to criticism" (my current understanding of Popper's idea, roughly illustrated), isn't this just another way of saying: "the fact that this theory is currently impervious to criticism is what **justifies** our reliance on it in making this decision?"

In short, isn't imperviousness to criticism a type of justification in itself?

curi 07 April 2011 05:02:40AM 2 points



Yes, we will have gotten somewhere. This issue is my primary criticism of Popperian epistemology.

OK then :-). Should we go somewhere else to discuss, rather than heavily nested comments? Would a new discussion topic page be the right place?

Is that basically right?

That is the general idea (but incomplete).

The reason we behave as if it's true is that it's the best option available. All the other theories are criticized (= we have an explanation of what we think is a mistake/ flaw in them). We wouldn't want to act on an idea that we (thought we) saw a mistake in, over one we don't think we see any mistake with -- we should use what (fallible) knowledge we have.

A justification is a reason a conjecture is *good*. Popperian epistemology basically has no such thing. There are no positive arguments, only negative. What we have instead of positive arguments is explanations. These are to help people understand an idea (what it says, what problem it is intended to solve, how it solves it, why they might like it, etc...), but they do not justify the theory, they play an advisory role (also note: they pretty much *are* the theory, they are the content that we care about in general).

One reason that not being criticized isn't a justification is that saying it is gets you a regress problem. So let's not say that! The other reason is: what would that be adding as compared with not saying it? It's not helpful (and if you give specific details/claims of how it is helpful, which are in line with the justificationist tradition, then I can give you specific criticisms of those).

Terminology isn't terribly important. David Deutsch used the word justification in his explanation of this in the dialog chapter of *The Fabric of Reality* (highly recommended). I don't like to use it. But the important thing is not to mean anything that causes a regress problem, or to expect justification to come from authority, or various other mistakes. If you want to take the Popperian conception of a good theory and label it "justified" it doesn't matter so much.

jake987722 07 April 2011 05:43:49AM 0 points



Should we go somewhere else to discuss, rather than heavily nested comments? Would a new discussion topic page be the right place?

I agree that the nested comment format is a little cumbersome (in fact, this is a bit of a complaint of mine

LW 2.0 Open Beta Live

by Vaniver | 21v (33c)

In support of yak shaving part 2

by Elo | 4v (7c)

Project Hufflepuff: Planting the Flag

by Raemon | 41v (106c)

European Community Weekend 2017

by DreamFlasher | 16v (16c)

"Flinching away from truth" is often about "protecting" the epistemology

by AnnaSalamon | 73v (53c)

Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"

by AnnaSalamon | 36v (43c)

Be secretly wrong

by Benquo | 32v (47c)

CFAR's new focus, and AI Safety

by AnnaSalamon | 30v (88c)

Fact Posts: How and Why

by sarahconstantin | 76v (32c)

Double Crux — A Strategy for Resolving Disagreement

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all

by vaultDweller on Rationality Quotes

April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

gwern (32)

ingres (30)

Lumifer (27)

turchin (23)

Habryka (22)

Yosarian2 (21)

IlyaShpitser (19)

SaidAchmiz (19)

gjm (19)

fowlertm (17)

SquirrelInHell (15)

Dagon (15)

Elo (14)

Stuart_Armstrong (13)

Gunnar_Zarncke (12)

RECENT KARMA AWARDS

about the LW format in general), but it's not clear that this discussion warrants an entirely new topic.

Terminology isn't terribly important . . . If you want to take the Popperian conception of a good theory and label it "justified" it doesn't matter so much.

Okay. So what is really at issue here is whether or not the Popperian conception of a good theory, whatever we call that, leads to regress problems similar to those experienced by "justificationist" systems.

It seems to me that it does! You claim that the particular feature of justificationist systems that leads to a regress is their reliance on positive arguments. Popper's system is said to avoid this issue because it denies positive arguments and instead only recognizes negative arguments, which circumvents the regress issue so long as we accept modus tollens. But I claim that Popper's system *does* in fact rely on positive arguments at least implicitly, and that this opens the system to regress problems. Let me illustrate.

According to Popper, we ought to act on whatever theory we have that has not been falsified. But that itself represents a positive argument in favor of any non-falsified theory! We might ask: okay, but *why* ought we to act only on theories which have not been falsified? We could probably come up with a pretty reasonable answer to this question--but as you can see, the regress has begun.

curi 07 April 2011 06:42:57AM 3 points



I think it's a big topic. Began answering your question here:

http://lesswrong.com/r/discussion/lw/551/popperian_decision_making/

curi 07 April 2011 06:33:37AM 2 points



We might ask: okay, but why ought we to act only on theories which have not been falsified? We could probably come up with a pretty reasonable answer to this question--but as you can see, the regress has begun.

No regress has begun. I already answered why:

The reason we behave as if it's true is that it's the best option available. All the other theories are criticized (= we have an explanation of what we think is a mistake/ flaw in them). We wouldn't want to act on an idea that we (thought we) saw a mistake in, over one we don't think we see any mistake with -- we should use what (fallible) knowledge we have.

Try to regress me.

It is possible, if you want, to create a regress of some kind which isn't the same one and isn't important. The crucial issue is: are the questions that continue the regress any good? Do they have some kind of valid point to them? If not, then I won't regard it as a real regress problem of the same type. You'll probably wonder how that's evaluated, but, well, it's not such a big deal. We'll quickly get to the point where your attempts to create regress look silly to you. That's different than the regresses inductivists face where it's the person trying to defend induction who runs out of stuff to say.



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

wedrifid comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 02:25:58AM 1 point

You could begin by reading

<http://fallibleideas.com/emotions>

And the rest of the site. If you don't understand any connections between it and Popperian epistemology, feel free to ask.

I'm not asking you to be interested in this, but I do think you should have some interest in rival epistemologies.

Desrtopa 07 April 2011 03:00:48AM 2 points

I studied philosophy as part of a double major (which I eventually dropped because of the amount of confusion and sophistry I was being expected to humor,) and my acquaintance with Popper, although not as deep as yours, I'm sure, precedes my acquaintance with Bayes. Although it may be that others who I have not read better presented and refined his ideas, Popper's philosophy did not particularly impress me, whereas the ideas presented by Bayesianism immediately struck me as deserving of further investigation. It's possible that I haven't given Popper his fair shakes, but it's not for lack of interest in other epistemologies that I've come to identify as Bayesian.

I wouldn't describe the link as unhelpful, exactly, but I also wouldn't say that it's among the best advice for controlling one's emotions that I've received (this was a process I put quite a bit of effort into learning, and I've received a fair amount,) so I don't see how it functions as a demonstration of the superiority of Popperian epistemology.

curi 07 April 2011 03:05:54AM 3 points

You say Popper didn't impress you. Why not? Did you have any criticism of his ideas? Any substantive argument against them?

Do you have any criticism of the linked ideas? You just said it doesn't seem that good to you, but you didn't give any kind of substantive argument.

Desrtopa 07 April 2011 03:38:56AM 1 point

With regards to the link, it's simply that it's less in depth than other advice I've received. There are techniques that it doesn't cover in meaningful detail, like manipulation of cognitive dissonance (habitually behaving in certain ways to convince yourself to feel certain ways,) or recognition of various cognitive

Custom Search

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

biases which will alter our feelings. It's not that bad as an introduction, but it could do a better job opening up connections to specific techniques to practice or biases to be aware of.

Popper didn't impress me because it simply wasn't apparent to me that he was establishing any meaningful improvements to how we go about reasoning and gaining information. Critical rationalism appeared to me to be a way of looking at how we go about the pursuit of knowledge, but to quote Feynman, "Philosophy of science is about as useful to scientists as ornithology is to birds." It wasn't apparent to me that trying to become more Popperian should improve the work of scientists at all; indeed, in practice it is my observation that those who try to think of theories more in the light of the criticism they have withstood than their probability in light of the available evidence are *more* likely to make significant blunders.

Attempting to become more Bayesian in one's epistemology, on the other hand, had immediately apparent benefits with regards to conducting science well (which are are discussed extensively on this site.)

I had criticisms of Popper's arguments to offer, and could probably refresh my memory of them by revisiting his writings, but the deciding factor which kept me from bothering to read further was that, like other philosophers of science I had encountered, it simply wasn't apparent that he had anything useful to offer, whereas it was immediately clear that Bayesianism did.

curi 07 April 2011 03:47:39AM 2 points

Feynman meant normal philosophers of science. Including, I think, Bayesians. He didn't mean Popper, who he read and appreciated. Feynman himself engaged in philosophy of science, and published it. It's academic philosophers, of the dominant type, that he loathed.

that those who try to think of theories more in the light of the criticism they have withstood than their probability in light of the available evidence

That's not really what Popperian epistemology is about. But also: the concept of evidence for theories is a mistake that doesn't actually make sense, as Popper explained. If you doubt this, do what no one else on this site has yet managed: tell me what "support" means (like in the phrase "supporting evidence") and tell me how support differs from consistency.

The biggest thing Popper has to offer is the solution to justificationism which has plagued almost everyone's thinking since Aristotle. You won't know quite what that is because it's an unconscious bias for most people. In short it is the idea that theories should be supported/justified/verified/proven, or whatever, whether probabilistically or not. A fraction of this is: he solved the problem of induction. Genuinely solved it, rather than simply giving up and accepting regress/foundations/circularly/whatever.

FAWS 07 April 2011 04:13:12AM -1 points

tell me what "support" means (like in the phrase "supporting evidence") and tell me how support differs from consistency.

Support is the same thing as more consistent with that hypothesis than with the alternatives ($P(E|H) > P(E|\sim H)$).

curi 07 April 2011 04:17:51AM 3 points

What is "more consistent"?

Consistent = does not contradict. But you can't not-contradict more. It's a boolean issue.

FAWS 07 April 2011 04:27:28AM 1 point

Then you have your answer: Support is non-boolean. I don't think a boolean concept of consistency of observations with anything makes sense, though (consistent would mean $P(E|H) > 0$, but observations never have a probability of 0 anyway, so every observation would be consistent with everything, or you'd need an arbitrary cut-off. $P(\text{observe black sheep} | \text{all sheep are white}) > 0$, but is very small).

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

- gwern (32)
- ingres (30)
- Lumifer (27)
- turchin (23)
- Habryka (22)
- Yosarian2 (21)
- IlyaShpitser (19)
- SaidAchmiz (19)
- gjm (19)
- fowlertm (17)
- SquirrelInHell (15)
- Dagon (15)
- Elo (14)
- Stuart_Armstrong (13)
- Gunnar_Zarncke (12)

RECENT KARMA AWARDS

curi 07 April 2011 04:29:21AM 3 points

Some theories predict that some things won't happen (0 probability). I consider this kind of theory important.

You say I have my answer, but you have not answered. I don't think you've understood the problem. To try to repeat myself less, check out the discussion here, currently at the bottom:

http://lesswrong.com/lw/54u/bayesian_epistemology_vs_popper/3urr?context=3

FAWS 07 April 2011 04:44:48AM * 2 points

Some theories predict that some things won't happen (0 probability). I consider this kind of theory important.

But they don't predict that you won't hallucinate, or misread the experimental data, or whatever. Some things not happening doesn't mean some things won't be observed.

You say I have my answer, but you have not answered.

You asked how support differed from consistency. Boolean vs real number is a difference. Even if you arbitrarily decide that real numbers are not allowed and only booleans are that doesn't mean that differentiating between their use of real numbers and your use of booleans is inconsistent on part of those who use real numbers.

Desrtopa 07 April 2011 03:59:01AM * 1 point

That's not really what Popperian epistemology is about. But also: the concept of evidence for theories is a mistake that doesn't actually make sense, as Popper explained. If you doubt this, do what no one else on this site has yet managed: tell me what "support" means (like in the phrase "supporting evidence") and tell me how support differs from consistency.

I've read his arguments for this, I simply wasn't convinced that accepting it in any way improved scientific conduct.

"Support" would be data in light of which the subjective likelihood of a hypothesis is increased. If consistency does not meaningfully differ from this with respect to how we respond to data, can you explain why it is more practical to think about data in terms of consistency than support?

I'd also like to add that I do know what justificationism is, and your tendency to openly assume deficiencies in the knowledge of others is rather irritating. I normally wouldn't bother to remark upon it, but given that you posed a *superior* grasp of socially effective debate conduct as evidence of the strength of your epistemology, I feel the need to point out that I don't feel like you're meeting the standards of etiquette I would expect of most members of Less Wrong.

curi 07 April 2011 04:08:17AM 1 point

I've read his arguments for this, I simply wasn't convinced that accepting it in any way improved scientific conduct.

Yet again you disagree with no substantive argument. If you don't have anything to say, why are you posting?

can you explain why it is more practical to think about data in terms of consistency than support?

Well, consistency is good as far as it goes. If we see 10 white swans, we should reject "all swans are black" (yes, even this much depends on some other stuff). Consistency does the job without anything extraneous or misleading.

The support idea claims that sometimes evidence supports one idea it is consistent with more than another. This isn't true, except in special cases which aren't important.

The way Popper improves on this is by noting that there are always many hypotheses consistent with the data. Saying their likelihood increases is pointless. It does not help deal with the problem of differentiating between them. Something else, not support, is needed. This leaves the concept of support with nothing useful to do, except be badly abused in sloppy arguments (I have in mind arguments I've seen elsewhere. Lots of them. What people do is they find some evidence, and some theory it is consistent with, and they say the theory is supported so now they have a strong argument or whatever. And they are totally selective about this. You try to tell them, "well, theory is also consistent with the data. so it's supported just as much. right?" and they say no, theirs fits the data better, so it's supported more. but you ask what the difference is, and they can't tell you because there is no answer. the idea that a theory can fit the data better than another, when both are consistent with the data, is a mistake (again there are some special cases that don't matter in practice).)

Desrtopa 07 April 2011 04:15:52AM 2 points

The support idea claims that sometimes evidence supports one idea it is consistent with more than another. This isn't true, except in special cases which aren't important.

Suppose I ask a woman if she has children. She says no.

This is supporting evidence for the hypothesis that she does not have children; it raises the likelihood from my perspective that she is childless.

It is entirely *consistent* with the hypothesis that she has children; she would simply have to be lying.

So it appears to me that in this case, whatever arguments you might make regarding induction, viewing the data in terms of consistency does not inform my behavior as well.

curi 07 April 2011 04:25:22AM 2 points

This is the standard story. It is nothing but an *appeal to intuition* (and/or unstated background knowledge, unstated explanations, unstated assumptions, etc). There is no argument for it and there never has been one.

Refuting this common mistake is something important Popper did.

Try reading your post again. You simply *assumed* that her having children is more likely. That is not true from the example presented, without some unstated assumptions being added. There is no argument in your post. That makes it very difficult to argue against because there's nothing to engage with.

It could go either way. You know it could go either way. You claim one way fits the data better, but you don't offer any rigorous guidelines (or anything else) for figuring out which way fits better. What are the rules to decide which consistent theories are more supported than others?

Desrtopa 07 April 2011 04:40:58AM * 3 points

Of course it could go either way. But if I behaved in everyday life as if it were *equally likely* to go either way, I would be subjecting myself to disaster. For practical purposes it has always served me better to accept that certain hypotheses that are consistent with the available data are more probable than others, and while I cannot prove that this makes it more likely that it will continue to do so in the future, I'm willing to bet quite heavily that it will.

If Popper's epistemology does not lead to superior *results* to induction, and at best, only reduces to procedures that perform as well, then I do not see why I should regard his refutation of induction as important.



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 08 April 2011 12:37:39AM 0 points

Sure, in the sense it is Turing complete;

No, in the sense that it *directly* applies to all types of knowledge (which any epistemology applies to -- which i think is all of them, but that doesn't matter to universality).

Not in the sense that it's Turing complete so you could, by a roundabout way and using whatever methods, do anything.

I think the basic way we differ is you have despaired of philosophy getting anywhere, and you're trying to get rigor from math. But Popper saved philosophy. (And most people didn't notice.) Example:

With this you could start to answer questions like "Why is X moral in the UK but not in Saudi Arabia?"

You have very limited ambitious. You're trying to focus on small questions b/c you think bigger ones like: what is moral objectively? are too hard and, since you math won't answer them, it's hopeless.

David_Allen 08 April 2011 02:13:42AM * 0 points

No, in the sense that it directly applies to all types of knowledge (which any epistemology applies to -- which i think is all of them, but that doesn't matter to universality).

Perhaps I don't understand some nuance of what you mean here. If you can explain it or link to something that explains this in detail I will read it.

But to respond to what I think you mean... If you have a method that can be applied to all types of knowledge, that implies that it is Turing complete; it is therefore equivalent in capability to other Turing complete systems; that also means it is susceptible to the infinite regresses you dislike in "justificationist epistemologies"... i.e. the [halting problem](#).

Also, just because it can be applied to all types of knowledge does not mean it is the best choice for all types of knowledge, or for all types of operations on that knowledge.

I think the basic way we differ is you have despaired of philosophy getting anywhere, and you're trying to get rigor from math. But Popper saved philosophy. (And most people didn't notice.) Example:

I would not describe my perspective that way; you may have forgotten that I am a *third* party in this argument. I think that there is a lot of historical junk in philosophy and that it is continuing to produce a lot junk -- Popper didn't fix this and neither will Bayesianism, it is more of a people problem -- but philosophy has also produced and is producing a lot of interesting and good ideas.

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on Bayesian Epistemology vs Popper - Less Wrong

-1 curi 06 April 2011 11:50PM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

[Article Navigation](#)

[Comments \(226\)](#)

Tags: [epistemology](#) | [popper](#) | [jaynes](#) [Delete](#)

Comments (226)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

curi 07 April 2011 05:22:01AM 2 points

I don't have any criticism of deductive logic itself. But I do have criticisms of some of the premises i expect you to use. For example, they won't all be deductively argued for themselves. That raises the problem of: how will you sort out good ideas from bad ideas for use as premises? That gets into various proposed solutions to that problem, such as induction or Popperian epistemology. But if you get into that, right in the premises of your supposed proof, then it won't be much of a proof because so much substantive content in the premises will be non-deductive.

endoself 07 April 2011 05:43:25AM -1 points

Do you agree with the premises I have used in the discussion of Dutch books and VNM-utility so far? There it is basically "a decision process that we actually care about must have the following properties" and that's it. I did skim over inferring probabilities from Dutch books and VNM axiom 3 and there may be some hidden premises in the former.

curi 07 April 2011 06:55:17AM 2 points

Do you agree with the premises I have used in the discussion of Dutch books and VNM-utility so far?
I don't think so. You said we have to assign probabilities to avoid getting Dutch Booked. I want an example of that. I got an example where probabilities weren't mentioned, which did not convince me they were needed.

Custom Search

- curi [Create new article](#)
- Karma Score [Submit a new link](#)
- 12 [Add new meetup](#)
- 2 [Preferences](#)
- [Log out](#)

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Create](#)

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

Denver Area LW November Meetup: 07 November 2017 08:00PM

Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS



MAIN ▾ | DISCUSSION ▾

WIKI | SEQUENCES | ABOUT

reply to benelliott about Popper issues

-1 curi 07 April 2011 08:11AM

This is a discussion page because I got the message "Comment too long". Apparently the same formatting magic doesn't work here for quotes :(It is a reply to:

http://lesswrong.com/lw/3ox/bayesianism_versus_critical_rationalism/3uly

> > You can conjecture Bayes' theorem. You can also conjecture all the rest, however some things (such as induction, justificationism, foundationalism) contradict Popper's epistemology. So at least one of them has a mistake to fix. Fixing that may or may not lead to drastic changes, abandonment of the main ideas, etc

> Fully agreed. In principle, if Popper's epistemology is of the second, self-modifying type, there would be nothing wrong with drastic changes. One could argue that something like that is exactly how I arrived at my current beliefs, I wasn't born a Bayesian.

OK great.

If the changes were large enough, to important parts (for example if it lost the ability to self-modify) I wouldn't want to call it Popper's epistemology anymore (unless maybe the changes were made very gradually, with Popper's ideas being valued the whole time, and still valued at the end). It would be departing from his tradition too much, so it would be something else. A minor issue in some ways, but tradition matters.

> I can also see some ways to make induction and foundationalism easier to swallow.

> A discussion post sounds about right for this, if enough people like it you might consider moving it to the main site.

104 comments later it's at 0 karma. There is interest, but not so much liking. I don't think the main site is the right place for me ;-)

> > I think you are claiming that seeing a white swan is positive support for the assertion that all swans are white. (If not, please clarify).

> This is precisely what I am saying.

Based on what you say later, I'm not sure if you mean this in the same way I meant it. I meant: it is positive support for "all swans are white" *over* all theories which assert "all swans are black" (I disagree with that claim). If it doesn't support them *more than those other theories* then I regard it as vacuous. I don't believe the math you offered meets this challenge over supporting "all swans are white" more than various opposites of it. I'm not sure if you intended it to.

> > If so, this gets into important issues. Popper disputed the idea of positive support. The criticism of the concept begins by considering: what is support? And in particular, what is the difference between "X supports Y" and "X is consistent with Y"?

> The beauty of Bayes is how it answers these questions. To distinguish between the two statements we express them each in terms of probabilities.

> "X is consistent with Y" is not really a Bayesian way of putting things, I can see two ways of interpreting it. One is as $P(X \& Y) > 0$, meaning it is at least theoretically possible that both X and Y are true. The other is that $P(X|Y)$ is reasonably large, i.e. that X is plausible if we assume Y.

Consistent means "doesn't contradict". It's the first one. Plausible is definitely not what I wanted.

> "X supports Y" means $P(Y|X) > P(Y)$, X supports Y if and only if Y becomes more plausible when we learn of X. Bayes tells us that this is equivalent to $P(X|Y) > P(X)$, i.e. if Y would suggest that X is more likely that we might think otherwise

Custom Search

<p>curi</p> <p>Karma Score</p> <p>12</p> <p>2</p> <p>✉</p>	<div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 2px; text-align: center;">Create new article</div> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 2px; text-align: center;">Submit a new link</div> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 2px; text-align: center;">Add new meetup</div> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 2px; text-align: center;">Preferences</div> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 2px; text-align: center;">Log out</div>
--	--

Website <http://fallibleideas.com/>

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

Create

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

[Denver Area LW November Meetup: 07 November 2017 08:00PM](#)

[Austin Meetup: 07 November 2020 02:30PM](#)

VIRTUAL STUDY ROOM

Co-work with other rationalists online.
[Less Wrong Study Hall](#)

RECENT COMMENTS

[Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points

[Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points

[Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points

[No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points

[I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

then X is support of Y.

This is true but fairly vaccous, in my view. I don't want to argue over what counts as significant. If you like it, shrug. It is important that, e.g., we reject ideas refuted by evidence. But I don't think this addresses the major problems in epistemology which come after we decide to reject things which are refuted by evidence.

The reason it doesn't is there's always infinitely many things supported by any evidence, in this sense. Infinitely many things which make wildly different predictions about the future, but identical predictions about whatever our evidence covers. If Y is 10 white swans, and X is "all swans are white" then X is supported, by your statement. But also supported are infinitely many different theories claiming that all swans are black, and that you hallucinated. You saw exactly what you would see if any of those theories were true, so they get as much support as anything else. There is nothing (in the concept of support) to differentiate between "all swans are white" and those other theories.

If you do add something else to differentiate, I will say the support concept is useless. The new thing does all the work. And further, the support concept is frequently abused. I have had people tell me that "all swans are black, but tomorrow you will hallucinated 10 white swans" is supported less by seeing 10 white swans tomorrow than "all swans are white" is, even though they made identical predictions (and asserted them with 100% probability, and would both have been definitely refuted by anything else). That kind of stuff is just wrong. I don't know if you think that kind of thing or not. What you said here does clearly disown it, nor advocate it. But that's the kind of thing that concerns me.

> Suppose we make X the statement "the first swan I see today is white" and Y the statement "all swans are white". $P(X|Y)$ is very close to 1, $P(X|\sim Y)$ is less than 1 so $P(X|Y) > P(X)$, so seeing a white swan offers support for the view that all swans are white. Very, very weak support, but support nonetheless.

The problem I have is that it's not supported over infinitely many rivals. So how is that really support? It's useless. The only stuff not being supported is that which contradicts the evidence (like, literally contradicts, with no hallucination claims. e.g. a theory that predicts you will think you saw a green swan tomorrow. but then you don't, just the white ones. that one is refuted). The inconsistent theories are refuted. The theories which make probabalistic predictions are partially supported. And the theories that say "screw probability, 100% every time" for all predictions get maximally supported, and between them support does not differentiate. (BTW I think it's ironic that I score better on support when I just stick 100% in front of every prediction in all theories I mention, while you score lower by putting in other numbers, and so your support concept discourages ever making predictions with under 100% confidence).

> (The above is not meant to be condescending, I apologise if you know all of it already).

It is not condescending. I think (following Popper) that explaining things is important and that nothing is obvious, and that communication is difficult enough without people refusing go over the "basics" in order to better understand each other. Of course this is a case where Popper's idea is not unique. Other people have said similar. But this idea, and others, are integrated into his epistemology closely. There's also *far more detail and precision* available, to explain *why* this stuff is true (e.g. lengthy theories about the nature of communication, also integrated into his epistemology). I don't think ideas about interpreting people's writing in kind ways, and miscommunication being a major hurdle, are so closely integrated with Bayesian approaches with are more math focussed and don't integrate so nicely with explanations.

My reply about support is basic stuff too, to my eye. But maybe not yours. I don't know. I expect not, since if it was you could have addressed it in advance. Oh well. It doesn't matter. Reply as you will. No doubt I'm also failing to address in advance something you regard as important.

> > To show they are correct. Popper's epistemology is different: ideas never have any positive support, confirmation, verification, justification, high probability, etc...

> This is a very tough bullet to bite.

Yes it is tough. Because this stuff has been integral to the Western philosophy tradition since Aristotle until Popper. That's a long time. It became common sense, intuitive, etc...

> > How do we decide which idea is better than the others? We can differentiate ideas by criticism. When we see a mistake in an idea, we criticize it (criticism = explaining a mistake/ flaw). That refutes the idea. We should act on or use non-refuted ideas in preference over refuted ideas.

> One thing I don't like about this is the whole 'one strike and you're out' feel of it. It's very boolean,

Hmm. FYI that is my emphasis more than Popper's. I think it simplifies the theory a bit to regard all changes to theories as new theories. Keep in mind you can always invent a new theory with one thing changed. So the ways it matters have some limits, it's party just a terminology thing (terminology has meaning, and some is better than others. Mine is chosen with Popperian considerations in mind. A lot of Popper's is chosen with considerations in mind of talking with his critics). Popper sometimes emphasized that it's important not to give up on theories too easily, but to look for ways to improve

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend 2017](#)

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

"... as the old saying went: 'Not all by vaultDweller on Rationality Quotes April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

[gwern](#) (32)

[ingres](#) (30)

[Lumifer](#) (27)

[turchin](#) (23)

[Habryka](#) (22)

[Yosarian2](#) (21)

[IlyaShpitser](#) (19)

[SaidAchmiz](#) (19)

[gjm](#) (19)

[fowlertm](#) (17)

[SquirrelInHell](#) (15)

[Dagon](#) (15)

[Elo](#) (14)

[Stuart_Armstrong](#) (13)

[Gunnar_Zarncke](#) (12)

RECENT KARMA AWARDS

them when they are criticized. I agree with that. So, the "one strike you're out" way of expressing this is misleading, and isn't *substantially* implied in my statements (b/c of the possibility of creating new and similar theories). Other terminologies have different problems.

> the real world isn't usually so crisp. Even a correct theory will sometimes have some evidence pointing against it, and in policy debates almost every suggestion will have some kind of downside.

This is a substantive, not terminological, disagreement, I believe. I think it's one of the *advantages* of my terminology that it helped highlight this disagreement.

Note the idea that evidence "points" is the support idea.

In the Popperian scheme of things, evidence does not point. It contradicts, or it doesn't (given some interpretation and explanation, which are often more important than the evidence itself). That's it. Evidence can thus be used in criticisms, but is not itself inherently a criticism or argument.

So let me rephrase what you were saying. "Even a correct theory will sometimes have critical arguments against it".

Part of the Popperian view is that if an idea has one false aspect, it is false. There is a sense in which any flaw must be decisive. We can't just go around admitting mistakes into our ideas on purpose.

One way to explain the issue is: for each criticism, consider it. Judge if it's right or wrong. Do your best and act on the consequence. If you think the criticism is correct, you absolutely must reject the idea it criticizes. If you don't, then you can regard the theory as not having any *true* critical arguments against it, so that's fine.

When you reject an idea for having one false part, you can try to form a new theory to rescue the parts you still value. This runs into dangers of arbitrarily rescuing everything in an ad hoc way. There's two answers to that. The first is: who cares? Popperian epistemology is not about laying out rules to prevent you from thinking badly. It's about offering advice to help you think better. We don't really care very much if you find a way to game the system and do something dumb, such as making a series of 200 ad hoc and silly arguments to try to defend a theory you are attached to. All we'll do is criticize you for it. And we think that is good enough: there are criticisms of bad methodologies, but no formal rules that definitively ban them. Now the second answer, which Deutsch presents in *The Fabric of Reality*, is that when you modify theories you often ruin their explanation. If you don't, then the modification is OK, it's good to consider this new theory, it's worth considering. But if the explanation is ruined, that puts an end to trying to rescue it (unless you can come up with a good idea for a new way to modify it that won't ruin the explanation).

This concept of ruining explanations is important and not simple. Reading the book would be great (it is polished! edited!) but I'll try to explain it briefly. This example is actually from his other book, *The Beginning of Infinity* chapter 1. We'll start with a bad theory: the seasons are caused by Persephone's imprisonment, for 6 months of the year, in the underworld (via her mother Demeter's magic powers which she uses to express her emotions). This theory has a bad explanation in the first place, so it can be easily rescued when it's empirically contradicted. For example this theory predicts the seasons will be the same all over the globe, at the same time. That's false. But you can modify the theory very easily to account for the empirical data. You can say that Demeter only cares about the area where she lives. She makes it cold when Persephone is gone, and hot when she's present. The cold or hot has to go somewhere, so she puts it far away. So, the theory is saved by an ad hoc modification. It's no worse than before. Its substantive content was "Demeter's emotions and magic account for the seasons". And when the facts change, that explanation remains in tact. This is a warning against bad explanations (which can be criticized directly for being bad explanations, so there's no big problem here).

But when you have a good explanation, such as the real explanation for the seasons, based on the Earth orbiting the sun, and the axis being tilted, and so on, ad hoc modifications cause bigger problems. Suppose we found out the seasons are the same all around the world at the same time. That would refute the axis tilt theory of seasons. You could try to save it, but it's hard. If you added magic you would be ruining the axis tilt *explanation* and resorting to a very different explanation. I can't think of any way to save the axis tilt theory from the observation that the whole world has the same seasons at the same time, without contradicting or replacing its explanation. So that's why ad hoc modifications sometimes fail (for good explanatory theories only). In the cases where there is not a failure of this type -- if there is a way to keep a good explanation and still account for new data -- then that new theory is genuinely worth consideration (and if there is some thing wrong with it, you can criticize it).

> There is also the worry that there could be more than one non-refuted idea, which makes it a bit difficult to make decisions.

Yes I know. This is an important problem. I regard it as solved. For discussion of this problem, go to:

http://lesswrong.com/r/discussion/lw/551/popperian_decision_making/

> Bayesianism, on the other hand, when combined with expected utility theory, is perfect for making decisions.

Bayesianism works when you assume a bunch of stuff (e.g. some evidence), and you set up a clean example, and you choose an issue it's good at handling. I don't think it is very helpful in a lot of real world cases. Certainly it helps in some. I regard Bayes' theorem itself as "how not to get probability wrong". That matters to a good amount of stuff. But hard real world scenarios usually have rival explanations of the proper interpretation of the available evidence, they have fallible evidence that is in doubt, they have often many different arguments that are hard to assign any numbers to, and so on. Using solomonoff induction is assign numbers, for example, doesn't work in practice as far as i know (e.g. people don't actually compute the numbers for dozens of political arugments using it). Another assumption being made is *what is a desirable (high utility) outcome* -- Bayesianism doesn't help you figure that out, it just lets you assume it (I see that as entrenching bias and subjectivism in reagrards to morality -- we *can* make objective criticisms of moral values).

Article Navigation

[Comments \(187\)](#)

Tags: [popper](#) | [deutsch](#) | [Delete support](#)

Comments (187)

Sort By: New

Enter a comment here

Comment

Show help

Peterdjones 14 April 2011 04:34:04PM 2 points 

Neither the potential infinity of theories, nor the possibility of error favour Popper over Bayes.

"The reason it doesn't is there's always infinitely many things supported by any evidence, in this sense. Infinitely many things which make wildly different predictions about the future, but identical predictions about whatever our evidence covers. If Y is 10 white swans, and X is "all swans are white" then X is supported, by your statement. But also supported are infinitely many different theories claiming that all swans are black, and that you hallucinated. You saw exactly what you would see if any of those theories were true, so they get as much support as anything else. There is nothing (in the concept of support) to differentiate between "all swans are white" and those other theories."

And that doesn't matter unless they (a) all have equal prior, and (b) will continue to be supported equally by any future evidence. (a) Is never the case, but doesn't help that much since, without a solution to (b), the relative rankings of various theories will never change from their priors, making evidence irrelevant. (b) is also never the case. The claims that 100% of swans are white, 90% are white, 80%, and so on will not all be equally supported by a long series of observations of white swans. A Popperian could argue that the theories that are becoming relatively less supported are becoming partially refuted. But relative refutation of T is relative support for not-T, just because of the meaning of relative. The Popperian can only rescue the situation by showing that there is absolute refutation, but no absolute support.

"If you do add something else to differentiate, I will say the support concept is useless. The new thing does all the work. And further, the support concept is frequently abused. I have had people tell me that "all swans are black, but tomorrow you will hallucinated 10 white swans" is supported less by seeing 10 white swans tomorrow than "all swans are white" is, even though they made identical predictions (and asserted them with 100% probability, and would both have been definitely refuted by anything else)."

They have different priors. The hallucination theory is a skeptical hypothesis, and it is well known that skeptical hypotheses can't be refuted empirically. But we can still give them low priors. Or regard them as bad explanations-- for instance, reject them because they are unfalsifiable or too easy to vary.

[deleted] 07 April 2011 03:17:26PM 5 points 

I want to emphasize this line from the long op, which I think is curi's best argument:

Hard real world scenarios usually have rival explanations of the proper interpretation of available evidence, they have fallible evidence that is in doubt, they have often many different arguments that are hard to assign

any numbers to, and so on.

Therefore Bayesianism does not describe the way that we actually find out true things. I think this is a pretty compelling criticism of Bayes, does anyone have a stock answer?

timtyler 07 April 2011 05:15:32PM * 2 points

Therefore Bayesianism does not describe the way that we actually find out true things. I think this is a pretty compelling criticism of Bayes, does anyone have a stock answer?

It isn't a theory about human psychology in the first place.

curi 07 April 2011 09:28:30PM 1 point

Epistemologies are theories about how knowledge is created.

Humans create knowledge.

If you want to be an epistemology, address the problem of how they do it.

timtyler 08 April 2011 12:50:13PM 1 point

Humans do all kinds of things badly. They are becoming obsolete.

For a neater perspective that is more likely to stand the test of time, it is better to consider how machines create knowledge.

komponisto 07 April 2011 04:39:13PM 2 points

Firstly, "the way we actually do X is by Y" is never a valid criticism of a theory saying "the way we should do X is by Z". (Contemporary philosophers are extremely fond of this mistake, it must be said.) If we're not using Bayes, then maybe we're doing it *wrong*.

Secondly, that the fact that we don't *consciously* think in terms of numbers doesn't mean that our brains aren't running Bayes-like algorithms on a low level not accessible to conscious introspection.

[deleted] 08 April 2011 01:45:53AM 0 points

Firstly, "the way we actually do X is by Y" is never a valid criticism of a theory saying "the way we should do X is by Z". (Contemporary philosophers are extremely fond of this mistake, it must be said.) If we're not using Bayes, then maybe we're doing it wrong.

Let me go further. The way people with a good track record of finding out true things (for instance, komponisto) actually go about finding out true things is by collecting explanations and criticisms of those explanations, not by computing priors and posteriors.

What would it mean to be doing it wrong? I can only think of: believing a lot of false things. So tell me some false things that I could come to believe by Popperian methods, that I wouldn't come to believe by Bayesian methods, or even better show me that the converse happens much more rarely.

Secondly, that the fact that we don't consciously think in terms of numbers doesn't mean that our brains aren't running Bayes-like algorithms on a low level not accessible to conscious introspection.

Sure. For instance there's good evidence that our brains judge what color something is by a Bayesian process. But why should I take advice about epistemology from such an algorithm?

JoshuaZ 07 April 2011 06:31:00PM 2 points

Secondly, that the fact that we don't consciously think in terms of numbers doesn't mean that our brains aren't running Bayes-like algorithms on a low level not accessible to conscious introspection.

Failure to perform correctly on the Monty Hall problem is cross-cultural. I haven't seen the literature in any detail but my impression is that the conjunction fallacy is also nearly universal. Whatever humans are doing it isn't very close to Bayes.

curi 07 April 2011 09:29:51PM 1 point

I actually don't agree with this. Those problems are caused by memes, not hardware.

Cross cultural is caused by the logic of the situation different early cultures were being in, and what mistakes are easy to make, being similar.

JoshuaZ 08 April 2011 01:45:20AM 2 points

I actually don't agree with this. Those problems are caused by memes, not hardware

How would you test this claim? (Note by the way that in the case of Monty Hall, the percentages don't change much from culture to culture. It is consistently between 75-90% refusing to switch in all tested cultures. This is actually one of the things that convinced me that this isn't memetic.)

komponisto 07 April 2011 08:51:01PM 3 points

Emphasis on *low-level*. Thinking "hmm, the probability of this outcome is 1/3" is high-level, conscious cognition. The sense in which we're "Bayesians" is like the sense in which we're good at calculus: catching balls, not (necessarily) passing written tests.

The conjunction fallacy is a closer to being a legitimate counterargument, but I would remind you that "Bayes-like" does not preclude the possibility of deviations from Bayes.

Perhaps some general perspective would be helpful. My point of view is that "inference = Bayes" is basically an analytic truth. That is, "Bayesian updating" is the mathematically precise notion that best corresponds to the vague, confused human idea of "inference". The latter *turns out to mean* Bayesian updating in the same sense that our intuitive idea of "connectedness" *turns out to mean* [this](#). As such, we can make our discourse strictly more informative by replacing talk of "inference" with talk of "Bayesian updating" throughout. We can talk about Bayesian updating done correctly, and done incorrectly. For example, instead of saying "humans don't update according to Bayes", we should rather say, "humans are inconsistent in their probability assignments".

curi 07 April 2011 09:31:06PM -2 points

That is, "Bayesian updating" is the mathematically precise notion that best corresponds to the vague, confused human idea of "inference".

I agree with you that "inference" is a vague and confused notion.

I don't agree that finding some math that somewhat corresponds to a bad idea, makes things better!

Popper's approach to it is to reject the idea and come up with better, non-confused ideas.

komponisto 07 April 2011 09:44:10PM 4 points

An idea becomes "non-confused" when it is turned into math. "Inference" may be a confused notion, but Bayesian updating isn't.

If Popper has better math than Bayes, so much the better. That's not the impression I get from your posts, however. The impression I get from your posts is that you meant to say "Hey! Check out this great heuristic that Karl Popper came up with for generating more accurate probabilities!" but instead it came out as "Bayes sucks! Go Popper!"

curi 07 April 2011 09:47:59PM 0 points

If the math is non-confused, and the idea is confused, then what's going on is not that the idea became non-confused but the math doesn't correspond to reality.

| If Popper has better math than Bayes, so much the better.

He doesn't have a lot of math.

No matter how much math you have, you always face problems of considering issues like whether some mathematical objects correspond to some real life things, or not. And you can't settle those issues with math.

| "Bayes sucks! Go Popper!"

You guys are struggling with problems, such as justificationism, which Popper solved. Also with instrumentalism, lack of appreciation for explanatory knowledge, foundationalism, etc

komponisto 07 April 2011 10:03:15PM 1 point



| If the math is non-confused, and the idea is confused, then what's going on is not that the idea became non-confused but the math doesn't correspond to reality.

What? Only confused ideas correspond to reality? That makes no sense.

| No matter how much math you have, you always face problems of considering issues like whether some mathematical objects correspond to some real life things, or not. And you can't settle those issues with math.

You settle those issues by experiment.

| You guys are struggling with problems, such as justificationism, which Popper solved. Also with instrumentalism, lack of appreciation for explanatory knowledge, foundationalism, etc

I'm not sure I see the problem, frankly. As far as I can tell this would be like me telling you that you're "struggling with the problem of Popperianism".

curi 07 April 2011 10:15:41PM 1 point



If you take a confused idea, X. And you take some non-confused math, Y. Then they do not correspond precisely.

| No matter how much math you have, you always face problems of considering issues like whether some mathematical objects correspond to some real life things, or not. And you can't settle those issues with math.

| You settle those issues by experiment.

Can't be done. When you try to set up an experiment you always have to have philosophical theories. For example if you want to measure something, you need a theory about the nature of your measuring device. e.g. you'll want to come up with some mathematical properties *and know if they correspond to the real physical object*. So you run into the same problem again.

| I'm not sure I see the problem, frankly.

How are theories justified?

How are theories induced? If you say using the solomonoff prior, then are the theories it offers always best? If not, that's a problem, right? If yes, what's the argument for that?

jimrandomh 07 April 2011 12:10:09PM 11 points

At this point, I have to conclude that you just plain don't understand Bayesian epistemology well enough to criticize it. I also suspect that you have become too strongly opinionated on this topic to be able to learn, at least until you get some distance.

The principle difference between Bayesian and Popperian epistemology is that Bayesianism is precise; it puts all necessary ambiguity in the prior, and assumes only noncontroversial, well-tested mathematical mathematical axioms, and everything thereafter is deductively sound math. In Popper, the ambiguity (which is still necessary) is in the definitions and spreads through the whole system, making its predictions much less concrete and thus making it hard to falsify.

To make progress in epistemology beyond Popper, you must switch from English to math. It takes a lot of work and a lot of time to rebuild a fuzzy English understanding of epistemology as a precise math understanding, but you will find that the precise math reproduces the same predictions, and many more predictions that the fuzzy English could never have made.

curi 07 April 2011 09:14:31PM 1 point

At this point, I have to conclude that you just plain don't understand Bayesian epistemology well enough to criticize it.

What you should do is say *specifically what I got wrong* (just one thing is fine). Then you'll be making a substantive statement!

making its predictions much less concrete and thus making it hard to falsify.

What predictions? It is a philosophical theory.

To make progress in epistemology beyond Popper, you must switch from English to math.

Your conception of epistemology is different than ours. We seek things like explanations that help us to understand the world.

jimrandomh 07 April 2011 09:29:44PM 3 points

What you should do is say *specifically what I got wrong* (just one thing is fine). Then you'll be making a substantive statement!

Ok, here's one. You criticize Bayesian updating for invoking infinitely many hypotheses, as a fundamental problem. In fact, the problem of infinite sets *is* an issue, but it's resolved in Jaynes' book by a set of rules in which one never deals with infinities directly, but rather with convergent limiting expressions, which are mathematically well-behaved in ways that infinities aren't. This ensures, among other things, that any set of hypotheses (whether finite or infinite) has only finite total plausibility, and lets us compute plausibilities for whole sets at once (ideally, picking out one element and giving it a high probability, and assigning a low total probability to the infinitely many other hypotheses).

What predictions? It is a philosophical theory.

Both theories make predictions about the validity of models using evidence - that is, they predict whether future observations will agree with the model.

Your conception of epistemology is different than ours. We seek things like explanations that help us to understand the world.

No, our conceptions of epistemology are the same. Math *does* help us understand the world, in ways that natural language can't.

curi 07 April 2011 09:42:01PM -2 points

Ok, here's one. You criticize Bayesian updating for invoking infinitely many hypotheses, as a fundamental problem.

No, I didn't say that. *I* invoked them, because they matter. You then claims Jaynes' deals with the problem.

Yet Yudkowsky concedes it is a problem. I don't think you understood me, rather than vice versa.

Both theories make predictions about the validity of models using evidence

Popper never made a prediction like that. And this rather misses some points. Some models for using evidence (e.g. induction) are *literally incapable of making predictions* (therefore people who do make predictions must be doing something else). Here Popper was not making a prediction, and also was pointing out prediction isn't the right way to judge some theories.

No, our conceptions of epistemology are the same. Math does help us understand the world, in ways that natural language can't.

Can you write philosophical explanations in math? Of course math helps for some stuff, but not everything.

jimrandomh 07 April 2011 09:52:57PM 2 points

Ok, here's one. You criticize Bayesian updating for invoking infinitely many hypotheses, as a fundamental problem.

No, I didn't say that. I invoked them, because they matter. You then claims Jaynes' deals with the problem. Yet Yudkowsky concedes it is a problem

Here's where you've really gone astray. You're trying to figure out math by reading what people are saying about it. *That doesn't work*. In order to understand math, you have to look at the math itself. I'm not sure what statement by Yudkowsky you're referring to, but I'll bet it was something subtly different.

Both theories make predictions about the validity of models using evidence

Popper never made a prediction like that.

Uh, wait a second. Did you really just say that Popper doesn't provide a method for using evidence to decide whether models are valid? There must be some sort of misunderstanding here.

timtyler 08 April 2011 12:41:17PM * 1 point

I'm not sure what statement by Yudkowsky you're referring to, but I'll bet it was something subtly different.

I am pretty sure it was this one - where: [Yudkowsky goes loopy](#).

curi 07 April 2011 10:17:18PM -1 points

The only way evidence is used is that criticisms may refer to it.

I'm not trying to figure out math, I'm trying to discuss the philosophical issues.

JoshuaZ 09 April 2011 05:03:14PM 2 points

The only way evidence is used is that criticisms may refer to it.

Please reread what Jim wrote. You seem to be in agreement with his statement that evidence is used.

I'm not trying to figure out math, I'm trying to discuss the philosophical issues.

Unfortunately, they are interrelated. There's a general pattern here: some people (such as Jaynes and Yudkowsky) are using math as part of their philosophy. In the process of that they are making natural language summaries and interpretations of those claims. You are taking those natural language statements as if that was all they had to say and then trying to apply your intuition of on ill-defined natural language statements rather than read those natural language statements in the context of the formalisms and math they care about. You can't divorce the math from the philosophy.

[deleted] 07 April 2011 03:19:23PM 6 points

The principle difference between Bayesian and Popperian epistemology is that Bayesianism is precise; it puts all necessary ambiguity in the prior, and assumes only noncontroversial, well-tested mathematical mathematical axioms, and everything thereafter is deductively sound math.

I think you're overselling it. Here are two big weaknesses of Bayesian epistemology as I understand it:

1. it cannot handle uncertainty about unproved mathematical truths.
2. It does not describe the way any existing intelligence actually operates, or even could operate in principle. (That last clause is the problem of writing an AI.)

I have *never seen* on this website any argument resolved or even approached on semirigorous Bayesian lines, except a couple of not-so-successful times between Yudkowsky and Hanson. Popper, or the second-hand accounts of him that I understand, seems to describe the way that I (and I think you!) actually think about things: we collect a big database of explanations and of criticisms of those explanations, and we decide the merits of those explanations and criticisms using our messy judgement.

In cases that satisfy some mild assumptions (but not so mild as to handle the important problem 1.!) this might be equivalent to Bayesianism. But equivalences go both ways, and Popper seems to be what we actually practice -- what's the problem?

timtyler 07 April 2011 05:02:00PM 0 points

Here are two big weaknesses of Bayesian epistemology as I understand it:

it cannot handle uncertainty about unproved mathematical truths.

Why do you think that?

It does not describe the way any existing intelligence actually operates, or even could operate in principle. (That last clause is the problem of writing an AI.)

Solomonoff induction is uncomputable? So: use a computable approximation.

[deleted] 08 April 2011 02:03:09AM 3 points

Why do you think that?

If you think I'm mistaken, please say so and elaborate.

Solomonoff induction is uncomputable? So: use a computable approximation.

It's hard for me to believe that you haven't thought of this, but it's difficult to "approximate" an uncomputable function. Think of any enormous computable function $f(n)$ you like. Any putative "approximation" of the busy beaver function is off by a factor larger than $f(n)$. I bet Solomonoff is similarly impossible to approximate -- am I wrong?

timtyler 08 April 2011 03:11:48AM 1 point

I am not aware of any particular issues regarding Bayesian epistemology handling uncertainty about unproved mathematical truths. How is that different from other cases where there is uncertainty?

Using a computable approximation of Solomonoff induction is a standard approach. If you don't have a perfect compressor, you just use the best one you have. It is the same with Solomonoff induction.

[deleted] 08 April 2011 02:01:34PM 0 points

I am not aware of any particular issues regarding Bayesian epistemology handling uncertainty about unproved mathematical truths. How is that different from other cases where there is uncertainty?

I outlined the problem with mathematical uncertainty [here](#). The only reason I believe that this is an open problem is on Yudkowsky's say-so in his reply.

Using a computable approximation of Solomonoff induction is a standard approach. If you don't have a perfect compressor, you just use the best one you have. It is the same with Solomonoff induction.

Standard approach to what? I don't know what a "compressor" or "perfect compressor" is, if those are technical terms.

To me, the question is whether an approximation to Solomonoff induction has approximately the same behavior as Solomonoff induction. I think it can't, for instance because no approximation of the busy beaver function (even the "best compressor you have") behaves anything like the busy beaver function. If you think this is a misleading way of looking at it please tell me.

timtyler 08 April 2011 04:29:55PM * 0 points

I outlined the problem with mathematical uncertainty [here](#).

Don't agree with the premise there. As to what Yudkowsky is talking about by saying "Logical uncertainty is an open problem" - it beats me. There's really only uncertainty. Uncertainty about mathematics is much the same as other kinds of uncertainty.

[deleted] 08 April 2011 04:32:53PM 0 points

What premise?

timtyler 08 April 2011 04:53:45PM 0 points

The first 5 lines of the post - this bit:

Any prior P , of any agent, has at least one of the following three properties:

[deleted] 08 April 2011 10:49:05PM 0 points

I can prove that to you, unless I made a mistake. Are you saying you can defeat it a priori by telling me a prior that doesn't have any of those three properties?

[continue this thread »](#)

JoshuaZ 08 April 2011 02:12:25PM 1 point

To me, the question is whether an approximation to Solomonoff induction has approximately the same behavior as Solomonoff induction. I think it can't, for instance because no approximation of the busy beaver function (even the "best compressor you have") behaves anything like the busy beaver function. If you think this is a misleading way of looking at it please tell me.

Solomonoff induction can't handle the Busy Beaver function because Busy Beaver is non-computable. So it isn't an issue for approximations of Solomonoff (except in so far as they can't handle it either).

[deleted] 08 April 2011 02:16:18PM 0 points

I am not saying that "Solomonoff can't handle Busy Beaver." (I'm not even sure I know what you

mean.) I am saying that Solomonoff is analogous to Busy Beaver, for instance because they are both noncomputable functions. Busy Beaver is non-approximatable in a strong sense, and so I think that Solomonoff might also be non-approximatable in an equally strong sense.

timtyler 08 April 2011 05:14:29PM 0 points

[Kolmogorov complexity](#) is uncomputable, but you can usefully approximate Kolmogorov complexity for many applications using PKZIP. The same goes for Solomonoff induction. Its prior is based on Kolmogorov complexity.

curi 07 April 2011 07:38:33PM 5 points

Solomonoff induction is uncomputable? So: use a computable approximation.

What is the argument that the approximation you use is good? What I mean is, when you approximate you are making changes. Some possible changes you could make would create massive errors. Others -- the type you are aiming for -- only create small errors that don't spread all over. What is your method of creating an approximation of the second type?

JoshuaZ 08 April 2011 03:19:11AM 0 points

There's a large amount of math behind this sort of thing, and frankly, given your other comments I'm not sure that you have enough background. It might help to just read up on Bayesian machine learning which needs to deal with just this sort of issue. Then keep in mind that there are theorems that given some fairly weak conditions to rule out pathological cases one approximate any distribution by a computable distribution to arbitrary accuracy. You need to be careful about what metric you are using but it turns out to be true for a variety of different notions of approximating and different metrics. While this is far from my area of expertise, so I'm by no means an expert on this, my impression is that the theorems are essentially of the same flavor as the theorems one would see in a real analysis course about approximating functions with continuous functions or polynomial functions.

timtyler 07 April 2011 09:02:45PM 0 points

Making computable approximations of Solomonoff induction is a challenging field which it seems inappropriate to try and cram into a blog comment. Probably the short answer is "by using stochastic testing".

komponisto 07 April 2011 04:59:18PM 0 points

Here are two big weaknesses of Bayesian epistemology as I understand it:

1. it cannot handle uncertainty about unproved mathematical truths.

Why not? You just use an appropriate formalization of mathematics, and treat it as uncertainty about the behavior of a proof-searching machine.

I have never seen on this website any argument resolved or even approached on semirigorous Bayesian lines, except a couple of not-so-successful times between Yudkowsky and Hanson

I can think of at least one [concrete example](#). But I'm guessing you were familiar with that example (and numerous other smaller-scale ones) and rejected it, so you must mean something different than I do by "argument approached on semirigorous Bayesian lines".

equivalences go both ways, and Popper seems to be what we actually practice -- what's the problem?

Perhaps there isn't any, except insofar as the poster is claiming that Bayes is wrong because it isn't Popper.

JoshuaZ 07 April 2011 06:29:10PM 0 points

Why not? You just use an appropriate formalization of mathematics, and treat it as uncertainty about the behavior of a proof-searching machine.

Unfortunately this isn't helpful. Consider for example a Turing machine that seems to halt on all inputs, and we know that when this one halts it halts with either a 0 or a 1. Does this machine represent a computable sequence (hence should have non-zero probability assigned if one is using a Solomonoff prior)? If we haven't resolved that question we don't know. But in order to use any form of prior over computable sequences we need to assume that we have access to what actually represents a computable hypothesis and what doesn't. There are other problems as well.

komponisto 07 April 2011 09:15:36PM 0 points

I'm having trouble parsing your third (I don't know what it means for a Turing machine to [fail to] "represent a computable sequence", especially since I thought that a "computable sequence" was by definition the output of a Turing machine) and fourth (we don't know *what?*) sentences, but if your general point is what I think it is ("after formalizing logical uncertainty, we'll still have *meta*-logical uncertainty left unformalized!"), that's simply a mathematical fact, and not an argument against the possibility of formalizing logical uncertainty in the first place.

JoshuaZ 08 April 2011 01:52:30AM * 1 point

I'm having trouble parsing your third (I don't know what it means for a Turing machine to [fail to] "represent a computable sequence", especially since I thought that a "computable sequence" was by definition the output of a Turing machine)

A sequence $f(n)$ is computable if there's a Turing machine T that given input n halts with output $f(n)$. But, not all Turing machines halt on all inputs. It isn't hard to make Turing machines that go into trivial infinite loops, and what is worse, Turing machines can fail to halt in much more ugly and non-obvious ways to the point where the question "Does the Turing machine M halt on input n " is not in general decidable. This is known as the [Halting theorem](#). So if I'm using some form of Solomonoff prior I can't even in general tell whether a machine describes a point in my hypothesis space.

komponisto 08 April 2011 02:15:18AM 0 points

What I don't understand is your argument that there is a *specific* problem with logical uncertainty that doesn't apply to implementing Solomonoff induction in general. Yes, the halting problem is undecidable, so you can't decide if a sequence is computable; but *assuming* you've already got a Solomonoff-induction machine that can say "my probability that it will rain tomorrow is 50%", why can't it *also* say "my probability that the Riemann Hypothesis is true is 50%"?

JoshuaZ 08 April 2011 02:23:04AM 1 point

but assuming you've already got a Solomonoff-induction machine that can say "my probability that it will rain tomorrow is 50%", why can't it also say "my probability that the Riemann Hypothesis is true is 50%"?

That's actually a really good example. It isn't that difficult to make a Turing machine that halts if and only if the Riemann hypothesis is true. So a system using Solomonoff induction has to recognize for starters whether or not that Turing machine halts. Essentially, in the standard version of Solomonoff induction, you need to assume that you have access to indefinitely large computing power. You can try making models about what happens when you have limited computational power in your entity (In some sense AIXI implementations and implementations of Bayesian reasoning need to do close to this). But if one doesn't assume that one has indefinite computing power then a lot of the results about how different priors behave no longer hold (or at least the proofs don't obviously go through). For more detail on that sort of thing I'd recommend talking to cousin_it or jimrandomh since they've thought and know a lot more about these issues than I do.

komponisto 08 April 2011 03:19:17PM * 0 points

It isn't that difficult to make a Turing machine that halts if and only if the Riemann hypothesis is true. So a system using Solomonoff induction has to recognize for starters whether or not that Turing machine halts.

Only in the sense that a human trying to solve the Riemann Hypothesis also has to recognize whether the same Turing machine halts.

When I talk about "going meta", I really mean it: when the Solomonoff machine that I have in mind is considering "whether this sequence is computable" or "whether the Riemann Hypothesis is true" or more generally "whether this Turing machine halts", it is going to be doing so the same way a human does: by using a *model* of the mathematical object in question that *isn't actually equivalent to that same mathematical object*. It won't be answering the question "natively", the way a computer typically adds 3+5 (i.e. by specific addition algorithms built into it); instead, it will be more closely analogous to a computer being programmed to simulate three apples being combined with five apples on its display screen, and then count the apples by recognizing their visual representation.

So the upshot is that to be able to give an answer the question "what is your probability that this Turing machine halts?", the Solomonoff AI does *not* need to solve anything equivalent to the halting problem. It just needs to examine the properties of some internal model corresponding to the label "Turing machine", which *need not be an actual Turing machine*. It is in this way that uncertainty about mathematical truths is handled.

It should go without saying that this isn't directly of use in building such an AI, because it doesn't tell you anything about how to construct the low-level algorithms that actually run it. But this thread isn't about how to build a Bayesian AI; rather, it's about whether a Bayesian AI is something that it makes sense to build. And my point here is that "Well, if you had a Bayesian AI, it wouldn't be able to give you probability estimates concerning the truth of mathematical statements" is not a valid argument on the latter question.

[deleted] 08 April 2011 03:56:57PM 0 points

So the upshot is that to be able to give an answer the question "what is your probability that this Turing machine halts?", the Solomonoff AI does not need to solve anything equivalent to the halting problem.

By "Solomonoff AI" do you mean "some computable approximation of a Solomonoff AI"? My impression is that the Solomonoff prior just does solve the halting problem, and that this is a standard proof that it is uncomputable.

when the Solomonoff machine that I have in mind is considering "whether this sequence is computable" or "whether the Riemann Hypothesis is true" or more generally "whether this Turing machine halts", it is going to be doing so the same way a human does.

Humans are bad at this. Is there some reason to think that a "the Solomonoff machine you have in mind" will be better at it?

[continue this thread »](#)

khafra 07 April 2011 03:36:13PM 0 points

My thinking seems to me more qualitatively bayesian than popperian. I don't have a good enough memory to keep all the criticisms I've ever heard of each theory I provisionally accept in mind. Instead, when I encounter a criticism that seems worth considering, I decrease my belief in the theory by an amount corresponding to the strength of the criticism. If I then go on to find evidence that weakens the criticism, strengthens the

original theory, or weakens all possible alternate theories, I increase my belief in the original theory again.

curi 07 April 2011 09:18:04PM 0 points 

You raise an interesting issue which is: what is the strength of a criticism? How is that determined?

For example, your post is itself a criticism of Popperian epistemology. What is the strength of your post?

By not using strengths of arguments, I don't have this problem. Strengths of arguments remind me of proportional representation voting where every side gets a say. PR voting makes a mess of things, not just in practice but also in terms of rigorous math (e.g. Arrow's Theorem)

Sniffnoy 08 April 2011 04:58:43AM 1 point 

What does Arrow's theorem have to do with proportional representation? Arrow's theorem deals with single-winner ordinal voting systems. Is there some generalization that covers proportional representation as well?

curi 08 April 2011 09:37:10AM * -1 points 

For one thing, all elections have a single overall outcome that wins.

Sniffnoy 08 April 2011 10:12:14AM 1 point 

Indeed, but "single-winner" has a technical meaning here that's rather more restrictive than that. Unless each voter could choose their vote arbitrarily from among the set of those overall outcomes, it's not single-winner.

curi 08 April 2011 04:58:21PM -4 points 

can you give the meaning of "single winner" and the reason you think not directly voting for the single winner will remove the problems?

in the US presidential elections, no voters can directly vote for his desired overall outcome. we have a primary system. are you saying that the primary system makes arrow's theorem's problems go away for us?

Sniffnoy 08 April 2011 05:53:28PM 1 point 

This appears to be very confused. Arrow's theorem is a theorem, a logical necessity, not a causal influence. It does not go around causing problems, that can be then be warded off by modifying your system to avoid its preconditions. It's just a fact that your voting system can't satisfy all its desiderata simultaneously. If you're not talking about a single-winner voting system it's simply an inapplicable statement. Perhaps the system meets all the desiderata, perhaps it doesn't. Arrow's theorem simply has nothing to say about it. But if you take a system that meets the preconditions, and simply modify it to not do so, then there's no reason to expect that it suddenly *will* start meeting the desiderata. For instance, though I think it's safe to say that US presidential elections are single winner, even if they somehow didn't count as such they'd fail to satisfy IIA. My point is just don't bring up theorems where they don't apply.

If you want a formal statement of Arrow's theorem... well it can take several forms, so take a look at the definitions I made in [this comment](#) about their equivalence. Then with that set of definitions, any voting system satisfying IIA and unanimity (even in just the "weak" form where if everyone votes for the same linear order, the overall winner is the winner of that linear order) with a finite set of voters must be a dictatorship. (There's a version for infinite sets of voters too but that's not really relevant here.)

curi 08 April 2011 06:03:49PM * -7 points 

You're very condescending and you're arguing with my way of speaking while refusing to actually provide a substantive argument.

it does apply. you're treating yourself as an authority, and me a beginner. and just assuming that knowing nothing about me.

you made a mistake which i explained in my second paragraph. your response: to ignore it, and say you were right anyway and i should just read more.

[continue this thread »](#)

JoshuaZ 08 April 2011 01:47:17AM 1 point 

By not using strengths of arguments, I don't have this problem.

Do you intend to treat all criticism equally?

[deleted] 07 April 2011 03:48:19PM * -1 points 

I'm suspicious of the notion of "increasing" and "decreasing" a belief. Did you pick those words exactly because you didn't want to use the word "updating"? Why not?

My guess is that having a bad memory is as much a disadvantage for Bayesians as for Popperians.

khafra 07 April 2011 04:06:25PM 3 points 

I'm suspicious of your suspicion. Is it purely because of the terms I used, or do you really have no beliefs you hold more tenuously than others?

If I ask you whether the sun rose this morning, will you examine a series of criticisms of that idea for validity and strength?

If I ask you whether it's dangerous to swim after a heavy meal, you'll probably check snopes to verify your suspicion that it's an old wives' tale, and possibly check any sources cited at snopes. But will you really store the details of all those arguments in your memory as metadata next to "danger of swimming after a heavy meal," or just mark it as "almost certainly not dangerous"?

[deleted] 07 April 2011 11:18:40PM 1 point 

I don't think the shortcuts I take on easy questions are very demonstrative of anything.

I'm suspicious of your suspicion. Is it purely because of the terms I used, or do you really have no beliefs you hold more tenuously than others?

I know what it means to think an outcome is likely or not likely. I don't know what it means for a belief to be tenuous or not tenuous.

curi 07 April 2011 09:19:36PM 3 points 

To save on storage, learn powerful explanations. Sometimes ideas can be integrated into better ideas that elegantly cover a lot of ground. Finding connections between fields is an important part of learning.

Learning easy to remember rules of thumb -- and improving them with criticism when they cause problems -- is also valuable for some applications.

David_Gerard 07 April 2011 12:21:00PM 4 points

At this point, I have to conclude that you just plain don't understand Bayesian epistemology well enough to criticize it.

LessWrong is seriously lacking a proper explanation of Bayesian epistemology (as opposed to the theorem itself). Do you have one handy?

timtyler 07 April 2011 05:12:41PM 0 points

<http://yudkowsky.net/rational/bayes> has a section on Bayesian epistemology that compares it to Popper's ideas.

Bayesian epistemology boils down to: use probabilities to represent your confidence in your beliefs, use Bayes's theorem to update your confidences - and try to choose a sensible prior.

curi 07 April 2011 07:39:50PM 2 points

Of course I've read that.

It first of all is focussed on Bayes' theorem without a ton of epistemology.

It second of all does not discuss Popper's ideas but only nasty myths about them. See the original post here:

http://lesswrong.com/lw/54u/bayesian_epistemology_vs_popper/

[deleted] 07 April 2011 12:50:49PM * -2 points

Yes, <http://wiki.lesswrong.com/wiki/Sequences> . Specifically the first four 'sequences' there. Many people on LW will say "Read the sequences!" as an answer to almost everything, like they're some kind of Holy Writ, which can be offputting, but in this case Yudkowsky really *does* answer most of the objections you've been raising and *is* the simplest explanation I know of.

ETA - That's weird. When I posted this reply I could have *sworn* that the username on the comment above wasn't DavidGerard *but one I didn't recognise. I wouldn't point DG* to the sequences because I know he's read them, but would point a newbie to them. Apologies for the brainfart,

[deleted] 07 April 2011 12:59:33PM 3 points

I've read the Sequences, and I don't know what you are referring to--the first four Sequences do explain Bayes-related ideas and how to apply them in everyday life, but they don't address all of the criticism that curi and others have pointed out. Did you have a specific post or series of posts in mind?

[deleted] 07 April 2011 01:19:51PM 0 points

The criticisms here that haven't been just noise have mostly boiled down to the question of choosing one hypothesis from an infinite sample space. I'm not sure exactly *where* in the sequences EY covered this, but I know he did, and did in one of those four, because I reread them recently. Sorry I can't be more help.

David_Gerard 07 April 2011 01:27:18PM * 4 points

What I'm meaning to point out is the absence of something with a title along the lines of "Bayesian epistemology" that explains Bayesian epistemology specifically.

What LW has is "here's the theorem" and "everything works by this theorem", without the second one being explained in coherent detail. I mean, Bayes structure *is* everywhere. But just noting that is not an explanation.

There's potential here for an enormously popular front-page post that gets linked all over the net forever, because the SEP article is so awful ...

benelliott 07 April 2011 11:09:10AM * 3 points 

Damn, I had a reply but it took so long to type that I lost internet connection.

Basically, with your point about supporting infinitely many theories, I refer you back to my comment that started this whole discussion.

As for the 'one strike and you're out' approach to criticism, I have three big problems with it:

The first is that if 'I don't understand' counts as a criticism, and you have claimed it does, then we need to reject every scientific theory we currently have since someone, somewhere, doesn't understand it.

Second, you accused Jaynes and Yudkowsky of being unscholarly when they said that Popper believed falsification is possible. Popper clearly believes refutation is possible, what is the difference between this and falsification?

Third, it leaves no room for weak criticisms.

Imagine I have a coin, which I think might be double-headed, although I'm not sure, it could also just be an ordinary coin. For some reason I am not allowed to examine both sides, all I can do is flip it and examine whichever side comes on top. If I was a Popperian how would I reason about this?

If tails comes on top then I have criticism of the 'double-headed' theory strong enough to refute it.

If heads comes on top then I do not have a strong enough criticism to refute either theory. This apparently means I do not have any criticism at all.

After 1000 heads I still don't have any criticism. At this point I get tired of flipping so I decide to make my decision. I still have two theories, so I criticise both of them for not refuting the other and reject them both. I sit down and try to come up with a better theory.

To me, this seems nuts. After 1000 heads, the 'ordinary coin' theory should have been thoroughly refuted. I would happily bet my life for a penny on it. If you wish to claim that is has been, then you must tell me where exactly, between 1 and 1000, you draw the line between refuted and not refuted.

curi 07 April 2011 09:27:32PM 1 point 

The first is that if 'I don't understand' counts as a criticism, and you have claimed it does, then we need to reject every scientific theory we currently have since someone, somewhere, doesn't understand it.

Sort of. It's not perfect. As far as scientific progress, it should be improved on. Indefinitely.

In the mean time we have to make decisions. For making decisions, we never directly use canonical scientific theories. Instead, you make a conjecture like, "QM isn't perfect. But I can use it for building this space ship, and my spaceship will fly". This conjecture is itself open to criticism. It could be a bad idea, depending on the details of the scenario. But it is not open to criticism by some guy in Africa not understanding QM, which doesn't matter.

Second, you accused Jaynes and Yudkowsky of being unscholarly when they said that Popper believed falsification is possible.

That's not what they said. Check out the actual quotes, e.g. Yudkowsky said "Karl Popper's idea that theories can be definitely falsified". That is not Popper's idea. Ideas cannot be "definitely falsified" but only fallibly/conjecturally/tentatively falsified.

Third, it leaves no room for weak criticisms.

That's a merit!

Well, it does leave room for them as ideas which you might remember and try to improve on later.

If you wish to claim that it has been, then you must tell me where exactly, between 1 and 1000, you draw the line between refuted and not refuted.

I would conjecture an explanation of when to stop and why. It would depend on what my goal was with the coin flipping. If that explanation wasn't refuted by criticism, I would use it.

For example, I might be flipping coins to choose which to use for the coin flipping olympics. my goal might be to keep my job. So I might stop after 5 flips all the same and just move on to the next coin. That would work fine for my purposes.

benelliott 07 April 2011 10:16:04PM 1 point

In the mean time we have to make decisions. For making decisions, we never directly use canonical scientific theories. Instead, you make a conjecture like, "QM isn't perfect. But I can use it for building this space ship, and my spaceship will fly". This conjecture is itself open to criticism. It could be a bad idea, depending on the details of the scenario. But it is not open to criticism by some guy in Africa not understanding QM, which doesn't matter.

This, kind of feels like cheating. You use all its predictions but never give it credit for making them.

Besides, are you really suggesting that 'someone doesn't understand this' is a legitimate criticism. If this was correct it would mean that scientific truth is partly dependant on human minds, and that the laws of physics themselves change with our understanding of them.

That's not what they said. Check out the actual quotes, e.g. Yudkowsky said "Karl Popper's idea that theories can be definitely falsified". That is not Popper's idea. Ideas cannot be "definitely falsified" but only fallibly/conjecturally/tentatively falsified.

Yudkowsky doesn't use 'definitely' to mean 'with certainty'.

That's a merit!

No its not!

Weak criticisms are important, because they add up to strong ones and sometimes they are all we have to decide by.

I would conjecture an explanation of when to stop and why. It would depend on what my goal was with the coin flipping. If that explanation wasn't refuted by criticism, I would use it.

For example, I might be flipping coins to choose which to use for the coin flipping olympics. my goal might be to keep my job. So I might stop after 5 flips all the same and just move on to the next coin. That would work fine for my purposes.

You're moving the goal posts.

I wasn't asking what you would do in pragmatic terms, I was asking at which point would you consider the theory refuted. You have claimed your thinking process is based on examining criticisms and refuting theories when they are valid, so when is it valid?

curi 07 April 2011 10:23:19PM 1 point

This, kind of feels like cheating. You use all its predictions but never give it credit for making them.

I don't know what you mean by "give credit". I'm happy to hand out all sorts of credit.

If I don't have a criticism of using all QMs predictions, then I'll use them. That someone doesn't understand QM isn't a criticism of this. That's only a criticism of explanations of QM.

If this was correct it would mean that scientific truth is partly dependant on human minds,

It would mean that what ideas are valuable is partly dependent on what people exist to care about them.

Yudkowsky doesn't use 'definitely' to mean 'with certainty'.

What does it mean?

He shouldn't write stuff that, using the dictionary definitions, is a myth about Popper, and not clarify. Even if you're right he isn't excused.

Weak criticisms are important, because they add up to strong ones and sometimes they are all we have to decide by.

I think they can't and don't. I think this is a big like saying 3 wrong answers add up to a right answer.

If an argument is false, why should it count for anything? Why would you ever want a large number of false arguments (false as best you can judge them) to trump one true argument (true as best you judge it)?

I wasn't asking what you would do in pragmatic terms, I was asking at which point would you consider the theory refuted.

I would *tentatively, fallibly* consider the theory "it is a fair coin" refuted after, say, 20 flips. Why 20? I conjectured 20 and don't have a criticism of it. For coin flipping in particular, if I had any rigorous needs, I would use some math in accordance with them.

benelliott 07 April 2011 10:40:41PM * 2 points

It would mean that what ideas are valuable is partly dependent on what people exist to care about them.

Be careful with arguing that an idea's value is a different thing to its truth, you're starting to sound like an apologist.

If I don't have a criticism of using all QMs predictions, then I'll use them. That someone doesn't understand QM isn't a criticism of this. That's only a criticism of explanations of QM.

If those explanations are helpful to some people they shouldn't be rejected simply because they are not helpful to others. After all, without them we would never have the predictions.

He shouldn't write stuff that, using the dictionary definitions, is a myth about Popper, and not clarify. Even if you're right he isn't excused.

I don't know about the dictionary definitions, but in everyday conversation 'definitely' doesn't mean 'with certainty'. As Wittgenstein pointed out, these words are frequently used in contexts where the speaker might be wrong for dozens of reasons, and knows it. For instance, "I definitely left my keys by the microwave" is frequently false, and is generally only said by people who are feeling uncertain about it.

I would tentatively, fallibly consider the theory "it is a fair coin" refuted after, say, 20 flips. Why 20? I conjectured 20 and don't have a criticism of it. For coin flipping in particular, if I had any rigorous needs, I would use some math in accordance with them.

I conjecture 21, you don't have any criticism of that either. I now have a criticism of 20, which is that it fails to explain why my conjecture is wrong.

I think they can't and don't. I think this is a big like saying 3 wrong answers add up to a right answer.

If an argument is false, why should it count for anything? Why would you ever want a large number of false arguments (false as best you can judge them) to trump one true argument (true as best you judge it)?

A weak criticism is not the same as an invalid criticism. It just means a criticism that slightly erodes a position, without single-handedly bringing the whole thing crashing down.

The coin-flip thing was intended as an example.

curi 08 April 2011 12:19:23AM 1 point

If those explanations are helpful to some people they shouldn't be rejected simply because they are not helpful to others

You are taking rejection as a bigger deal than it is. The theory that "X is the perfect explanation" for X that confuses some people is false.

So we reject it.

We can accept other theories, e.g. that X is flawed but, for some particular purpose, is appropriate to use.

I don't know about the dictionary definitions,

It means "without doubt". Saying things like "I have no doubt that X" when there is doubt is just dumb.

I conjecture 21, you don't have any criticism of that either. I now have a criticism of 20, which is that it fails to explain why my conjecture is wrong.

The problem situation is under specified. When you ask ambiguous questions like what should I do in [under specified situation] then you get multiple possible answers and it's hard to do much in the way of criticizing.

In real world situations (which have rich context), it's not so hard to decide. But when I gave an example like that you objected.

I can't criticize 20 vs 21 unless I have some goal in mind, some *problem we are trying to solve*. (If there is no problem to be solved, I won't flip at all.) If the problem is figuring out if the coin is fair, with certainty, that is not solvable, so I won't flip at all. If it is figuring it out with a particular probability, given a few reasonable background assumptions, then I will look up the right math to use. If it's something else, what?

A weak criticism is not the same as an invalid criticism. It just means a criticism that slightly erodes a position, without single-handedly bringing the whole thing crashing down.

This is an important issue. I think your statement here is imprecise.

A criticism might demolish *one single idea* which is *part of* a bigger idea.

If it demolishes zero individual ideas, then where is the erosion?

If it demolishes one little idea, then that idea is refuted. And the big idea needs to replace it with something else which is not refuted, or find a way to do without.

benelliott 08 April 2011 07:32:33AM 2 points



You are taking rejection as a bigger deal than it is. The theory that "X is the perfect explanation" for X that confuses some people is false.

Maybe QM is exactly right, and maybe it is just too complicated for some people to understand. There is no need to be so harsh in your criticism process, why not just admit that a theory can be right without being perfect in every other respect.

It means "without doubt". Saying things like "I have no doubt that X" when there is doubt is just dumb.

Yet everyone does it. Language is a convention, not a science. If you are using a word differently from everyone else then you are wrong, the dictionary has no authority on the matter.

The problem situation is under specified. When you ask ambiguous questions like what should I do in [under specified situation] then you get multiple possible answers and it's hard to do much in the way of criticizing.

This is a flaw. Bayes can handle any level of information.

I can't criticize 20 vs 21 unless I have some goal in mind, some problem we are trying to solve. (If there is no problem to be solved, I won't flip at all.) If the problem is figuring out if the coin is fair, with certainty, that is not solvable, so I won't flip at all. If it is figuring it out with a particular probability, given a few reasonable background assumptions, then I will look up the right math to use. If it's something else, what?

Can you really not see why the above is moving the goal posts. Earlier, you said that you think by coming up with conjectures, and criticising them, and only then make decisions. Now you are putting the decision making process in the driving seat and saying that everything is based on that.

So is Popperianism purely pragmatic? Is the whole conjecture and criticism thing not really the important part, and in fact its all based on decision strategies. Or do you use the conjecture-criticism thing to try and reach the correct answer, as you have previously stated, and then use that for decision making.

| If it demolishes zero individual ideas, then where is the erosion?

It makes the idea less likely, less plausible, by a small amount. The coin flip is intended to illustrate it. Saying that you will use Bayes in the coin flip example and nowhere else is like saying you believe Newton's laws work 'inside the laboratory' but you're going to keep using Aristotle outside.

GuySrinivasan 07 April 2011 05:16:59PM 0 points [-]

Following curi's steps, we'd lower our standards. How do you feel about the theory "I don't want to spend more time on this and getting 1000 heads if it's double-headed is 2^{1000} more likely than getting 1000 heads if it's ordinary so I'll make the same decisions I'd make if I knew it were double-headed unless I get a rough estimate of at least a factor of 2^{990} difference in how much I care about the outcome of one of those decisions".

benelliott 07 April 2011 05:24:43PM 1 point [-]

What you appear to be suggesting amounts to Bayesian epistemology done wrong.

curi 07 April 2011 09:20:44PM 2 points [-]

For coin flipping analysis, use Bayes' theorem (not Bayesian epistemology).

timtyler 08 April 2011 12:53:42PM 0 points [-]

Whether you have a double-headed coin or not is still a form of knowledge.

The Bayes' theorem:good, Bayesian epistemology:bad perspective won't wash.

benelliott 07 April 2011 10:08:27PM 1 point [-]

If Bayes generates the right answer here, whereas naive Popperian reasoning without it goes spectacularly wrong, maybe this should be suggesting something. Also it ignores my main point that Poppers theory does not admit weak criticisms, of which the coin coming up heads is just one example.

[deleted] 07 April 2011 09:32:51AM 2 points [-]

"I see that as entrenching bias and subjectivism in regards to morality -- we *can* make objective criticisms of moral values."

You keep asserting that. You keep failing to provide a shred of evidence.

"BTW I think it's ironic that I score better on support when I just stick 100% in front of every prediction in all theories I mention, while you score lower by putting in other numbers, and so your support concept discourages ever making predictions with under 100% confidence"

That's true *right up until you see the first black swan*. All else being equal, simpler explanations are always to be favoured over more complex ones. Look up Kolmogrov complexity and minimum message length.

curi 07 April 2011 09:46:31AM 2 points [-]

You keep asserting that. You keep failing to provide a shred of evidence.

I posted arguments. What did you not like about them? Post a criticism of something I said. "No evidence" is just a request for justification which I regard as impossible, but I did give arguments.

That's true right up until you see the first black swan.

At which point infinitely many of my 100% theories will be refuted. And infinitely many will remain. You can never win at that game using finite evidence. For any finite set of evidence, infinitely many 100% type theories predict all of it perfectly.

prase 07 April 2011 02:14:05PM 1 point

At which point infinitely many of my 100% theories will be refuted. And infinitely many will remain. You can never win at that game using finite evidence. For any finite set of evidence, infinitely many 100% type theories predict all of it perfectly.

It seems that your objection is basically that if I toss a coin seventeen times and it ends up in a sequence of HTTTHTHHHTHTHTTH, there is a specific theory T1 (namely, that the physical laws cause the sequence to be HTTTHTHHHTHTHTTH) which scores higher than the clearly correct explanation T2 (i.e. the probability of each sequence is the same $2^{(-17)}$). But this is precisely why priors depend on the Kolmogorov complexity of hypotheses: with such a prior, the posterior of T2 will be higher than the posterior of T1.

And, after all, you don't have infinitely many theories. Theories live in brains, not in an infinite Platonic space of ideas. Why should we care whether there are infinitely many ways to formulate a theory so absurd that nobody would think of it but still compatible with the evidence? Solomonoff induction tells you to ignore them, which agrees with the common sense.

curi 07 April 2011 07:22:13PM 1 point

Selectively ignoring theories, even when we're aware of them, is just bias, isn't it?

I'm a bit surprised that someone here is saying to me "OK so mathematically, abstractly, we're screwed, but in practice it's not a big deal, proceed anyway". Most people here respect math and abstract thinking, and don't dismiss problems merely for involving substantial amounts of theory.

Of course a prior can arbitrarily tell you which theories to prefer over others. But why those? You're getting into problems of arbitrary foundations.

prase 07 April 2011 08:39:26PM 1 point

Bias is a systematic *error* in judgement, something which yields bad results. It is incorrect to apply that label to heuristics which are working well.

I haven't told you that we are abstractly screwed, but it's no big deal. We are not screwed, on the contrary, the Solomonoff induction is a consistent algorithm which works well in practice. It is as arbitrary as any axioms are arbitrary. You can't do any better if you want to have any axioms at all, or any method at all. If your epistemology isn't completely empty, it can be criticised for being arbitrary without regard to its actual details. And after all, what ultimately matters is whether it works practically, not some perceived lack of arbitrariness.

calef 07 April 2011 08:32:35PM 1 point

We're fundamentally incapable of making statements about reality without starting on some sort of arbitrary foundation.

And I think describing it as "selectively ignoring" is doing it an injustice. We're deductively excluding, and if there were some evidence to appear that would contradict that exclusion, those theories would no longer be excluded.

I'm actually have trouble finding a situation in which a fallibilist would accept/reject a proposition, and a Bayesian would do the opposite of the fallibilist. And I don't mean epistemological disagreements, I

mean disagreements of the form "Theory Blah is not false."

curi 07 April 2011 08:36:54PM 2 points 

We're fundamentally incapable of making statements about reality without starting on some sort of arbitrary foundation.

This is something Popper disputes. He says you can start in the middle, or anywhere. Why can't that be done?

And I think describing it as "selectively ignoring" is doing it an injustice. We're deductively excluding

I was talking about the theories that can't be deductively excluded b/c they make identical predictions for all available evidence.

thakil 07 April 2011 10:28:03AM 3 points 

What arguments? I had a look through what you've written, and found this "There are objective facts about how to live, call them what you will. Or, maybe you'll say there aren't. If there are, then it's not objectively wrong to be a mass murderer. Do you really want to go there into full blown relativism and subjectivism?"

This is hardly an argument for the truth content of a statement. Just because the consequences of a theory of moral behaviour make us feel bad doesn't mean they are not true- we should be interested in whether the statements conflict with how the universe seems to work. The notion of morality independent to sentient beings has always seemed fundamentally absurd to me, and I have yet to find a decent argument in its favour.

The worry of slipping into moral relativism is that we are trapped in a position where we can't punish mass murderers. But theres lots of sensible reasons for mass murderers to punish mass murder, and not indulge in it themselves. One would have to get exceptional utility out of murdering to counteract all the downsides with performing such an action.

The problem here is, as often happens with moral discussion, that wrong is not well defined. You say wrong to mean a grand moral force of the universe, but it could mean "is this a sensible action for this being to take, given their goals and desires". It might turn out that given all that said person IS benefited most by mass murder

curi 07 April 2011 10:43:05AM * 2 points 

As I recall I gave a citation where to find Popper discussing morality (it is The World of Parmenides).

And I explained that moral knowledge is created using the same method as any other kind of knowledge. And i said that that method is (conjectures and refutations).

questioning if people want to advocate strong relativism or subjectivism is an argument, too. if you aren't aware of the already existing arguments against relativism or subjectivism, then it's incomplete for you. you could always ask.

you haven't understood my view. i didn't say it's a moral force. the issue of "what is the right action, given my goals and desires?" is 100% objective, and it is a moral issue. i don't know why you expected me to disagree about that. there is a fact of the matter about it. that is one of the major parts of morality. but there is also a second part: the issue of what are good goals and desires to have?

how can that be objective, you wonder? well for example some sets of goals contradict each other. that allows for a type of objective moral argument, about what goals/values/preferences/utility-functions to have, against contradictory goals.

there's others. to start with, read: <http://www.curi.us/1169-morality>

benelliott 07 April 2011 12:30:08PM 5 points

"what is the right action, given my goals and desires?" is 100% objective

Bayes, combined with Von Neuman Mortenson utility theory answers this, at least in principle.

You keep acting as if it is a flaw that Bayes only predicts. Is it a flaw that Newton's laws of motion do not explain the price of gold? Narrowness is a virtue, attempting to spread your theory as wide as possible ends up forcing it into places where it doesn't belong.

curi 07 April 2011 06:19:56PM 1 point

If bayes wants to be an *epistemology* then it must do more than predict. Same for Newton.

If you want to have math which doesn't dethrone Popper, but is orthogonal, you're welcome to do that and i'd stop complaining (much). However Yudkowsky says Bayesian Epistemology dethrones and replaces Popper. He regards it as a rival theory to Popper's. Do you think Yudkowsky was wrong about that?

timtyler 07 April 2011 07:40:07PM * 1 point

Yudkowsky says Bayesian Epistemology dethrones and replaces Popper. He regards it as a rival theory to Popper's. Do you think Yudkowsky was wrong about that?

It replaces Popperian epistemology where their domains overlap - namely: building models from observations and using them to predict the future. It won't *alone* tell you what experiments to perform in order to gather more data - there are other puzzle pieces for dealing with that.

curi 07 April 2011 08:23:29PM 1 point

There's no overlap there b/c Popperian epistemology doesn't provide the specific details of how to do that. Popperian epistemology is fully compatible with, and can use, Bayes' theorem and any other *pure* math or logic insights.

Popperian epistemology contradicts your "other puzzle pieces". And without them, Bayes' theorem alone isn't epistemology.

timtyler 07 April 2011 08:54:30PM * 2 points

It replaces Popperian epistemology where their domains overlap - namely: building models from observations and using them to predict the future.

There's no overlap there b/c Popperian epistemology doesn't provide the specific details of how to do that.

Except for the advice on induction? Or has induction merely been rechristened as corroboration? Popper enthusiasts usually seem to deny doing that.

curi 07 April 2011 08:55:50PM 1 point

Induction doesn't work.

building models from observations and using them to predict the future

I thought you were referring to things you can do with Bayes' theorem and some input. If you meant something more, provide the details of what you are proposing.

[continue this thread »](#)

benelliott 07 April 2011 07:28:42PM 1 point

The most common point of Popper's philosophy that I hear (including from my Popperian philosophy teacher) is the whole "black swan white swan" thing, which Bayes does directly contradict, and dethrone (though personally I'm not a big fan of that terminology).

The stuff you talked about with conjectures and criticisms does not directly contradict Bayes and if the [serious problems](#) with 'one strike and you're out' criticisms are fixed it I may be persuaded to accept both it and Bayes.

Bayes is not meant to be an epistemology all on its own. It only starts becoming one when you put it together with Solomonoff Induction, Expected Utility Theory, Cognitive Science and probably a few other pieces of the puzzle that haven't been found yet. I presume the reason it is referred to as Bayesian rather than Solomonoffian or anything else is that Bayes is the both most frequently used and the oldest part.

curi 07 April 2011 07:33:14PM * 1 point

The black swan thing is not that important to Popper's ideas, it is merely a criticism of some of Popper's opponents.

How does Bayes dethrone it? By asserting that white swans support "all swans are white"? I've addressed that at length (still going through overnight replies, if someone answered my points I'll try to find it).

| Solomonoff Induction, Expected Utility Theory, Cognitive Science

Well I don't have a problem with Bayes' theorem itself, of course (pretty much no one does, right? i hope not lol). It's these surrounding ideas that make an epistemology that I think are mistaken, and all of which Popper's epistemology contradicts. (I mean the take on cognitive science popular here, not the idea of doing cognitive science).

benelliott 07 April 2011 07:58:17PM 1 point

| (still going through overnight replies, if someone answered my points I'll try to find it)

I think I answered your points a few days ago with my first comment of this discussion.

In short, yes, there are infinitely many hypotheses whose probabilities are raised by the white swan, and yes those include both "all swans are white" and "all swans are black and I am hallucinating" but the former has a higher prior, at least for me, so it remains more probable by several orders of magnitude. For evidence to support X it doesn't have to only support X. All that is required is that X does better at predicting than the weighted average of all alternatives.

| I have had people tell me that "all swans are black, but tomorrow you will hallucinated 10 white swans" is supported less by seeing 10 white swans tomorrow than "all swans are white" is, even though they made identical predictions (and asserted them with 100% probability, and would both have been definitely refuted by anything else).

Just to be clear I am happy to say those people were completely wrong. It would be nice if nobody ever invented a poor argument to defend a good conclusion but sadly we do not live in that world.

curi 07 April 2011 08:21:58PM 1 point

| I think I answered your points a few days ago with my first comment of this discussion.

But then I answered your answer, right? If I missed one that isn't pretty new, let me know.

| but the former has a higher prior

so support is vacuous and priors do all the real work. right?

and priors have their own problems (why that prior?).

Just to be clear I am happy to say those people were completely wrong. It would be nice if nobody ever invented a poor argument to defend a good conclusion but sadly we do not live in that world.

OK. I think your conception of support is unsubstantive but not technically wrong.

[continue this thread »](#)

thakil 07 April 2011 11:05:32AM 0 points



Mm, I'm not sure I entirely agree with that link- we might accept that most long term goals on maximisation will lead to what most people might recognise as morality, but I don't know if all goals are long term. It also makes no argument as to one goal being "better" than another. There's sensible reasons for me to discourage people having the desire to kill me, for example, but I don't see that one could argue that I'm right and he's wrong. If someone is born with just one innate desire, that of killing me, it's in her interest to pursue that goal. Now she might well act morally elsewhere while engaging fully in her training towards killing me, but at some point where she is confident that she will be able to kill me, she should drop everything else and kill me. Of course after this her life is empty, but she only had that one desire, and she had to fulfill it at some point- she got absolutely no value from everything else.

Now was she wrong to pursue that goal? I don't see how I can condemn her. I obviously will do everything in my power to stop her, and I would hope others in society would have goals which are interrupted by my untimely demise, but I don't see where condemnation comes in here. We had conflicting goals, and mine seem "nicer" from an intuitive argument, but if I lived in a world where everyone had a strong desire to see me dead then I imagine it would feel "nicer" to them for me to die, and "nasty" for me to survive.

curi 07 April 2011 06:17:07PM 2 points



Not all goals are long term.

One of the purposes of the dialog is to explain that the *foundations are not very important*. That means you don't have to figure out the correct foundations or starting place to have objective morality. You can start wherever you want, because rather little depends on the starting place.

Once you do make a ton of progress, when you're much wiser, if your starting place was squirrels you'd be able to reconsider it because it's so silly. The same holds for any other particularly dumb starting place.

The ones that will be harder to change later are specifically the ones that you don't see as bad -- that you don't want to change. The ones that are either correct or you don't yet have enough knowledge to see the problems with them.

If someone is born with just one innate desire

Innate desires aren't morality. It's a bad argument "I was born this way therefore I *should* be this way". That's getting and ought from an is.

Moving on, one way to move past the squirrel scenario, which enables you to criticize the squirrel starting point and many others, is you consider other scenarios. Drop the squirrels and put in something else, like minimizing bison. Put in a way variety of stuff. It's not too important what it is. Any kind of value, taken seriously, and which has something to say long term. Even wanting to kill someone will work if you also want them to stay dead forever (if you really want to make sure to destroy all the information that could be used to resurrect them later with advanced technology, and you want to know what kind of remains *could* be used for that and what would violate the laws of physics, then you will need advanced knowledge).

So, you try the same thought experiment with bison-minimizing, or killing-forever.

You find that some of the conclusions are the same, and some are different.

Take only the ones that are the same for thousands of starting points.

Those are the non-parochial ones. They are the ones that don't depend on your culture and biases. There is the objective content.

The parts that vary by starting points are wrong.

That's what I think. And I think this argument isn't bogged down in being totally subjective from the start. Maybe it's not perfect, but that just means we could make an even better argument in the future.

Getting back to your original point, this shared content across many starting points does say stuff about what to do short term. It doesn't give complete arbitrary freedom of action in the short term. It's also not totally restrictive, but that's good (it might get a lot more restrictive if made more precise. but that would also be good. knowing how to live well in high detail would be a good thing even if it gave you less non-immoral options. as long as we don't jump the gun and create very precise rules before we understand how to work them out well, we'll be ok.)

thakil 07 April 2011 07:26:11PM 1 point



Sorry, but I just do not see how you can claim desires are not morality when you have yet to provide a basis for what it is! I see no reason to believe that those bases with common conclusions are somehow better. They might feel better, but thats not good enough

curi 07 April 2011 07:28:57PM * 2 points



you have yet to provide a basis

I've argued that morality is at least largely, if not entirely, *independent of basis*. So asking me for a basis isn't the right question.

Can you give an example of a starting point you think *avoids* the common conclusions such as liberalism?

thakil 07 April 2011 08:32:06PM * 0 points



You have shown that an argument can be made that given a number of seemingly dissimilar, long term goals, e can make arguments which convincingly argue that to achieve them one should act in a manner people would generally consider moral. I am not convinced squirrel morality gives me an answer on specific moral questions (abortion say) but I can see how one might manage it. You have yet to convince me that short term bases will do the same: I am reasonably confident that many will not. To claim theses bases as inferior seems to be begging the question to me.

As to your specific question: how about a basis of wanting to prevent liberalism? It would certainly be difficult to achieve and counter productive, but to claim that those respective properties are bad begs the question: you need morality to condemn purposes which are going to cause nothing but pain for all involved.

curi 07 April 2011 08:52:02PM 2 points



how about a basis of wanting to prevent liberalism?

If you were just to destroy the world, or build a static society and die of a meteor strike one day b/c your science never advanced, then life could evolve on another planet.

You need enough science and other things to be able to affect the whole universe. And for that you need liberalism temporarily. Then at the very very end, when you're powerful enough to *easily* do whatever you want to the whole universe (needs to be well within your power, not at the limits of your power, or it's too risky, you might fail) then finally you can destroy or control everything.

So that goal leads straight to billions of years of liberalism. And that does mean freedom of abortion: ruining people's lives to punish them for having sex does not make society wealthier, does not promote progress, etc... But does increase the risk of everyone dying of meteor before you advance enough to deal with such a problem.

| short term bases

Accomplish short term things, in general, depends on principles. Suppose I want a promotion at work within the next few years. It's important to have the right kind of philosophy. I'll have a better shot at it if I think well. So I'll end up engaging with some big ideas. Not every single short term basis will lead somewhere interesting. If it's really short, it's not so important. Also consider this: we can conjecture that life is nice. People cannot use short term bases, which don't connect to big ideas, to criticize this. If they want to criticize it, they will have to engage with some big ideas, so then we get liberalism again.

[continue this thread »](#)

[deleted] 07 April 2011 10:10:55AM 3 points

The problem is, nobody *else* here (or very few people here) regards justification as impossible, You're essentially saying you refuse to engage by the same evidentiary rules as anyone else here. You're not going to change anyone's mind *without* providing justification.

"At which point infinitely many of my 100% theories will be refuted. And infinitely many will remain. "

Like I said, *look up Kolmogrov Complexity and minimum message length*. At any given time, the simplest of those 'theories' consistent with all data is the one with the highest probability.

curi 07 April 2011 10:25:50AM 2 points

Can you tell how ideas are justified, without creating a regress or other severe problem? Tell me the type of justificationism that works, *then* I will accept it.



MAIN | DISCUSSION

WIKI | SEQUENCES | ABOUT

curi comments on reply to benelliott about Popper issues - Less Wrong

Custom Search

-1 curi 07 April 2011 08:11AM

You are viewing a comment permalink. [View the original post](#) to see all comments and the full post content.

curi [Create new article](#)

Karma Score [Submit a new link](#)

12 [Add new meetup](#)

2 [Preferences](#)

[Log out](#)

[Article Navigation](#)

[Comments \(187\)](#)

Tags: [popper](#) | [deutsch](#) | [support](#) [Delete](#)

Website <http://fallibleideas.com/>

Comments (187)

Sort By: New

You are viewing a single comment's thread. [Show more comments above.](#)

There is no Less Wrong Wiki account associated with curi. You can create one now by supplying your password:

[Subscribe to RSS Feed](#)

NEAREST MEETUPS

- Denver Area LW November Meetup: 07 November 2017 08:00PM
- Austin Meetup: 07 November 2020 02:30PM

VIRTUAL STUDY ROOM

Co-work with other rationalists online. [Less Wrong Study Hall](#)

RECENT COMMENTS

- [Yeah, that's a bug. I am planning to](#) by [Habryka](#) on LW 2.0 Open Beta Live | 0 points
- [Woo! Also if anyone else gets a](#) by [korin43](#) on LW 2.0 Open Beta Live | 0 points
- [Running through this to check that](#) by [Multipartite](#) on Sleeping Beauty gets counterfactually mugged | 0 points
- [No, you don't need update you](#) by [trickster](#) on What Evidence Filtered Evidence? | 0 points
- [I think that the core of religion—that](#) by [adjuant](#) on Religion's Claim to be Non-Disprovable | 1 point

RECENT POSTS

curi 08 April 2011 06:03:49PM * -7 points [\[-\]](#)

You're very condescending and you're arguing with my way of speaking while refusing to actually provide a substantive argument.

it does apply. you're treating yourself as an authority, and me a beginner. and just assuming that knowing nothing about me.

you made a mistake which i explained in my second paragraph. your response: to ignore it, and say you were right anyway and i should just read more.

Sniffnoy 08 April 2011 06:13:19PM * 2 points [\[-\]](#)

and just assuming that knowing nothing about me.

I am, in fact, inferring that based on what you wrote.

you made a mistake which i explained in my second paragraph. your response: to ignore it, and say you were right anyway and i should just read more.

Then I must ask that you point out this mistake more explicitly, because nothing in that second paragraph contradicts anything I said.

curi 08 April 2011 06:15:19PM * -1 points [\[-\]](#)

Your claims implied Arrow's Theorem doesn't apply to the US election system. Or pretty much any other. You also defined "single-winner" so that the single US president who wins the election doesn't qualify.

OK, you're right I didn't actually contradict you. But don't you think that's a mistake? I think it does apply to real life voting systems that people use.

Sniffnoy 08 April 2011 06:18:29PM * 3 points [\[-\]](#)

Perhaps you misunderstand what is meant when people say a theorem "applies". It means that the preconditions are met (and that therefore the conclusions hold), not simply that the conclusions hold. The conclusions of a theorem can hold without the theorem being at all applicable.

curi 08 April 2011 06:19:55PM * -1 points [\[-\]](#)

If you meant "Arrow's theorem is not applicable to any voting system on Earth" why did you object to my statements about PR? The PR issue is irrelevant, is it not?

[LW 2.0 Open Beta Live](#)

by Vaniver | 21v (33c)

[In support of yak shaving part 2](#)

by Elo | 4v (7c)

[Project Hufflepuff: Planting the Flag](#)

by Raemon | 41v (106c)

[European Community Weekend](#)

2017

by DreamFlasher | 16v (16c)

["Flinching away from truth" is often about "protecting" the epistemology](#)

by AnnaSalamon | 73v (53c)

[Further discussion of CFAR's focus on AI safety, and the good things folks wanted from "cause neutrality"](#)

by AnnaSalamon | 36v (43c)

[Be secretly wrong](#)

by Benquo | 32v (47c)

[CFAR's new focus, and AI Safety](#)

by AnnaSalamon | 30v (88c)

[Fact Posts: How and Why](#)

by sarahconstantin | 76v (32c)

[Double Crux — A Strategy for Resolving Disagreement](#)

by Duncan_Sabien | 61v (102c)

LATEST RATIONALITY QUOTE

["... as the old saying went: 'Not all](#)

by vaultDweller on Rationality Quotes

April - June 2017 | 0 points

RECENT WIKI EDITS

RECENT ON RATIONALITY BLOGS

TOP CONTRIBUTORS, 30 DAYS

[gwern](#) (32)

[ingres](#) (30)

[Lumifer](#) (27)

[turchin](#) (23)

[Habryka](#) (22)

[Yosarian2](#) (21)

[IlyaShpitser](#) (19)

[SaidAchmiz](#) (19)

[gjm](#) (19)

[fowlertm](#) (17)

[SquirrelInHell](#) (15)

[Dagon](#) (15)

[Elo](#) (14)

[Stuart_Armstrong](#) (13)

[Gunnar_Zarncke](#) (12)

RECENT KARMA AWARDS