I'm Kenny Easwaran, philosopher working on formal epistemology, decision theory, philosophy of mathematics, and social epistemology. AMA.

easwaran [1] and r/Science AMAs[1]

[1]Affiliation not available

April 17, 2023

Abstract

I work in areas of formal epistemology, philosophy of mathematics, decision theory, and am increasingly interested in issues of social epistemology and collective action, both as they relate to my earlier areas and in other ways. I've done work on various paradoxes of the infinite in probability and decision theory, on the foundations of Bayesianism, on the social epistemology of mathematics, and written one weird paper using metaphysics to derive conclusions about physics. Links of Interest: My research website including links and descriptions to most of my papers. My appearance (in 2015) on Julia Galef's "Rationally Speaking" podcast, discussing Newcomb's Paradox, its connection to other issues in decision theory and free will, and what I call a "tragedy of rationality". A discussion (from 2011) with Jonathan Weisberg about the role of accuracy in constraining beliefs and probabilities, and their connection, on Philosophy TV. The idea of this discussion eventually became my Dr. Truthlove paper in Nous (paper available from Philosophers' Annual - 10 Best Papers of 2015) My paper "Decision Theory without Representation Theorems", at the open access journal Philosophers' Imprint. My old blog, Antimeta, which I ran for several years in graduate school, discussing issues in philosophy of mathematics, probability, and occasionally metaphysics. My posts from the period 2005-2009 on Brian Weatherson's blog, Thoughts, Arguments, and Rants.

the WINNOWER

REDDIT

I'm Kenny Easwaran, philosopher working on formal epistemology, decision theory, philosophy of mathematics, and social epistemology. AMA.

EASWARAN R/SCIENCE

I work in areas of formal epistemology, philosophy of mathematics, decision theory, and am increasingly interested in issues of social epistemology and collective action, both as they relate to my earlier areas and in other ways. I've done work on various paradoxes of the infinite in probability and decision theory, on the foundations of Bayesianism, on the social epistemology of mathematics, and written one weird paper using metaphysics to derive conclusions about physics.

Links of Interest:

My research website including links and descriptions to most of my papers.

My appearance (in 2015) on Julia Galef's "Rationally Speaking" podcast, discussing Newcomb's Paradox, its connection to other issues in decision theory and free will, and what I call a "tragedy of rationality".

A discussion (from 2011) with Jonathan Weisberg about the role of accuracy in constraining beliefs and probabilities, and their connection, on Philosophy TV.

The idea of this discussion eventually became my Dr. Truthlove paper in Nous (paper available from Philosophers' Annual - 10 Best Papers of 2015)

My paper "Decision Theory without Representation Theorems", at the open access journal Philosophers' Imprint.

My old blog, Antimeta, which I ran for several years in graduate school, discussing issues in philosophy of mathematics, probability, and occasionally metaphysics.

My posts from the period 2005-2009 on Brian Weatherson's blog, Thoughts, Arguments, and Rants.

CORRESPONDENCE:

DATE RECEIVED:
May 11, 2017

DOI:
10.15200/winn.149443.31476

ARCHIVED:
May 10, 2017

CITATION:
easwaran , r/Science , I'm Kenny Easwaran, philosopher working on formal epistemology, decision theory, philosophy of mathematics, and social epistemology. AMA., *The Winnower* 4:e149443.31476 , 2017 , DOI: 10.15200/winn.149443.31476

My question is on the metaphysical limits of mathematics.

For ancient thinkers, the infinite was considered unknowable. Since Leibniz, this has rapidly changed. Things formerly left to the realm of opinion and confused sense perceptions (such as the layout of a coastline), can now be mapped using various recursive algorithms and fractals.

Are there still elements of nature that "resist" mathematics, in the way that Plato talks about the struggle between intellect and necessity? Or can mathematics provide a complete metaphysical picture? Does modern mathematics have limits, or can it provide a full account of nature? Do any unknowable, indeterminate gaps remain in nature?

If no gaps remain, should we turn towards pure mathematics, rather than traditional logic, as the basis for our metaphysical systems?

If gaps do remain, should we view them as metaphysical gaps, or an epistemological problem? More important to me is whether we're *justified* in viewing such gaps as metaphysical.

iunoionnis

I think in a sense, our best understanding of everything is mathematical. But we don't always have the best mathematics for each thing we might want to understand. The infinite is one area that resisted mathematization for a while, but other areas did too. (Think about how the development of Arrow's theorem and related social choice theory revolutionized its area, and also things like the mathematics of networks, or phase transitions.)

Modern mathematics is still limited, but I don't think there's any principle limits on what can be mathematized by future developments of mathematics, except if there are aspects of reality that are actually unknowable or unspeakable. There may well be such limits, but it's hard to say much useful about what lies beyond them.

Professor Easwaran - thanks for coming, we're delighted to have you.

I was hoping you could say a bit about your interest in social epistemology, qua someone who primarily (?) works in formal epistemology. Not a lot of social epistemology is currently formal. Do you see yourself as someone bringing formalism to social epistemology, or are you happy to just go with the field as is?

One can of course imagine a logician who is completely formal in one half of their life, but also doubles as a completely non-formal ethicist as well. Is that your approach, or how do you see yourself fitting in with both camps?

ADefiniteDescription

My work in philosophy of mathematics (The Role of Axioms; Probabilistic Proofs and Transferability; Rebutting and Undercutting in Mathematics) is actually mostly non-formal social epistemology. The probabilistic proofs paper has one section trying to argue that we might have a way of applying formal Bayesian theory to mathematical claims (I think an interesting newer approach is here) but the main idea is that *even if* we can do that, there's social reasons to be aiming for something different with mathematical arguments (that is, we should care about providing reasons that others can accept as their own, rather than caring about just how confident we should be in our conclusions).

I don't have a clear sense of how these two branches of my work relate to one another, but for now, I mainly think that my familiarity with mathematics is helpful for both, in different ways.

(I've also done some formal social epistemology in the coauthored paper Updating on the Credences of Others, by Easwaran, Fenton-Glynn, Hitchcock, and Velasco. Using the un-hyphenated name, we are the EaGlHiVe.)

/u/rhetoricgirl asked in the announcement thread:

Hi!

Thank you for participating in this AMA session.

I'm a communication studies graduate student whose research is philosophically informed, and I was curious how decision theory accounts for individuals with neurocognitive disorders such as dementia? In particular, I wanted to know how decision theory handles groups who may not meet the theory's standards for rationality.

Dementia patients' decisions are not always based on rational means because due to the disease's effects on the brain (this fact does not discount their decision-making skills, but their decisions may not fit into the criteria set by rational decision making). For instance, say you and a dementia patient are given the choice between a stack of hundred dollar bills or a sack of pennies. Now, you may pick the

stack of bills not because pictures of Benjamin Franklin are aesthetically pleasing to you. Rather you realize the value of the stack to obtain goods and services within society. The dementia patient might pick the sack of pennies because her favorite color is copper and her favorite color makes her happy. The dementia patient may not be viewed as a rational actor because she is acting more on her feelings without accounting for the societal value placed on money.

Thank you for reading this long-winded question (comm people can talk your ear off)!

lapse_of_taste

Thanks! This is an interesting and important question that my work hasn't sufficiently engaged with, but I do have thoughts on the matter.

The first thing that I want to say is that the notion of rationality I am trying to work with has a sort of Humean aspect, saying that what matters is whether your plans and actions are the sort that seem to you to be good at achieving the goals that you have. If your goals involve having shiny copper things rather than drab green pictures, and your goals don't involve social exchange of these objects for other things, then this sort of rationality says go ahead and take the pennies, and leave the pictures of Benjamin for someone else. (For a quick summary of how this comes out of David Hume, the first page of this paper appears to be useful. I haven't read the rest of Setiya's argument in that paper, but I suspect he's going to argue for a slightly more subtle understanding of Hume than the one I'm putting forward.)

That said, there's a further distinction in epistemology that might be relevant to people who want to claim that the dementia patient described here is being irrational. "Internalists" in epistemology (like Conee and Feldman) say that what matters is having internal justification for your beliefs, so that a brain in a vat, or a victim of a Cartesian demon, might be equally rational as a person in the real world. "Externalists" (particularly reliabilists like Alvin Goldman) often say that what matters is reasoning according to abilities that are *in fact* reliable, so that a real person with real senses is justified in her beliefs, while the victim of the Cartesian demon is not. I would say that there is a parallel idea for actions. The person who takes the hundreds has a plan for acting in the world that is in fact more reliable at providing for her future interests than the person who takes the pennies (at least, in the current social setting - maybe not in a post-apocalyptic setting where copper is more important than currency!)

At any rate, I don't know enough about actual cases of dementia (or other situations that we might consider cognitive impairments) to say for sure whether they get in the way of action that is rational by some appropriate internal standard, or whether they just change people's desires and access to knowledge about the world in ways that leave them internally rational but externally less reliable at carrying out long term plans.

If there are impairments that interfere so much with belief, desire, and action that people stop making choices that even internally make sense, then perhaps the person has actually become irrational in the Humean sense that I'm most interested in. But such a person is in some ways no longer an agent at all. I don't mean any moral significance to attach to this - such persons can still presumably feel suffering and joy. But if they don't have a connection between desires and actions, then the sort of rationality I am interested in doesn't apply to them.

Professor Easwaran,

In your "Truthlove" paper, you mention a few times that some of your approach to defending probabilistic thinking is motivated by James' position on the relationship between belief and evidence.

Do you think that James would accept the "use" of Bayesian thinking that you propose in that paper, or

does your argument require rejecting some portion of James' position in "Will To Believe" and elsewhere?

(As an aside - I wonder if you have talked to fellow Aggies Profs. McDermott or Crick about this, given their respective commitments/interests to/in James and Pragmatism more generally.)

mediaisdelicious

One of the last times I presented that paper before it was published was when I was trying to get hired at Texas A&M, and I did get some useful questions and discussion then. I unfortunately haven't yet had a chance to take advantage of the presence of my colleagues to get a substantially deeper understanding of the views of James himself.

In some ways, the point I'm making in the Truthlove paper is that even if the world is *in fact* one way, it might be more *useful* to describe it another way. Even if people just have binary beliefs, it might be more useful to describe them as having probabilities. I think many pragmatists would be very amenable to this sort of point (perhaps without the flourish of using the words "in fact") - we should describe the world in whatever way is most useful for our purposes.

I suspect that in many ways, a view that is more congenial for the pragmatists would be that of Bruno de Finetti (particularly his paper Probabilism), who I believe explicitly conceives of his viewpoint as that of a pragmatist. de Finetti's view is actually much more orthodox for a Bayesian, saying that degree of belief just is whatever guides action, rather than thinking there is some notion of "aiming at the truth" prior to action.

At any rate, the Dr. Truthlove paper is one that I think of as characteristic for my work, in that I'm putting forward a viewpoint that I think probably isn't correct, but probably also has independent interest, and may help us better understand logical space so that we can understand other views in the vicinity better. But I present it in that paper *as if* I accept it.

I'm moving from a physics undergrad program into a philosophy graduate one, and hope to balance both fields. So, how do you keep up with the latest mathematical research while doing philosophical work?

Araraguy

Simple answer: I barely keep up with the latest *philosophical* work in my own field, and I am *definitely* not keeping up with the latest work in other fields!

I'm really not good at reading journals or otherwise actively keeping up with recent work in any field. So I arrange external circumstances to help keep me up to date. I work as an editor at several journals, and accept most referee requests that come in, which gives me a lot of reason to read (anonymized) current papers, many of which are quite interesting. I also go to a lot of conferences, both ones directly in my areas of formal epistemology and decision theory, and somewhat broader ones in philosophy or logic generally, and I usually come out with interesting ideas from talks I've seen there.

But the biggest thing is really just maintaining friendships with people who work in other fields. From graduate school I know a lot of mathematicians, and I also have a few people I stay in touch with through Facebook at least that are physicists, economists, psychologists, and working in other fields as well. From occasionally hearing what they're up to, I get an occasional sense of interesting ideas coming out of their fields.

Hi Kenny!

You were one of my RAs at Canada/USA Mathcamp a number of years ago, so it's crazy seeing your name pop up on my reddit feed. You were quite a mathematics inspiration to me, so I just wanted to stop by and offer a quick thank you!

Chuckzduck

I'm glad to have played a role!

Wow, fascinating and in depth material! I'm completely unfamiliar with epistemology and decision theory, and I was wondering what practical application the the field and the theory have? I tried reading your paper, but I need to delve into the history of traditoinal decision theory as first endorsed by Savage and Jeffrey in order to understand your paper.

belovicha21

I don't know all the practical applications. But most of the field of microeconomics is based on decision theory (Savage was himself an economist). And a major current controversy in statistics (the conflict between Bayesian and frequentist methodology) turns on the relation between epistemology and decision theory.

I would also recommend reading Daniel Kahneman's book "Thinking, Fast and Slow" for an understanding of how theoretical descriptions of rational belief and decision might differ from the way humans actually work.

What is the day to day as a philosopher?

palladists

Day to day I'm doing things like meeting with students, answering student questions about homework and exams, sitting on university committee meetings about assessment of programs, e-mailing colleagues and collaborators about projects we're working on or papers of theirs I've read, reading and commenting on papers for journals, and occasionally actually planning lectures or writing my own papers. On the one hand, everything I'm doing is an interaction with other people. But on the other hand, I'm often sitting in a room alone in front of a computer while doing most of it, or interacting impersonally with people in a lecture or committee setting.

Are people just socially inept? I find that social problems repeat themselves throughout history, yet people seem not to try to solve social problems. What is your proposed solution? Or comment...

ElsyrDeimos

I think people do *try* to solve social problems. But social problems are really hard, because any proposed solution to the problem also changes the way people interact, and thus changes the nature of the problem.

As a simple example, consider the idea of trying to predict the stock market vs trying to predict the path of a comet. Because the stock market is directed by the actions of millions of people that are trying to predict it, any new method for predicting it will *change* how it moves. It's *possible* for a new method of predicting the movement of comets to inspire people to change how the comet moves, but it's much less common. Prediction of social behavior will almost always change the thing being predicted, while prediction of physical systems often doesn't, making social problems much harder to address than physical ones.

Have you seen the paper presenting functional decision theory by Levinstein and Soares (https://intelligence.org/files/DeathInDamascus.pdf)? What do you think about the theory - does it solve the tragedy of rationality?

UmamiSalami

I haven't read that paper yet, but I've talked to proponents of that and related views over the past few years, and I'll see the presentation they make of that paper at the Formal Epistemology Workshop in Seattle in a few weeks.

My overall thought is that where traditional causal decision theorists try to solve the tragedy of rationality by putting off the notion of rationality to the latest possible moment (whether you grab the second box or not), the functional decision theorists want to put it at the earliest possible moment (a sort of fictional self-creation moment of choosing what algorithm to be). I think that both are valid to evaluate, as well as intermediate points on the chain (like the moment of planning, or the period of training and education).

[deleted]

[deleted]

I don't have a fully worked out argument at this point.

But it seems to me that "rationality" is a virtue that a system can have that (at least partially) consists in being effective at achieving its aims. Epistemic systems are rational if they are effective at getting accurate representations of the world; action systems are rational if they are effective at achieving the desired ends.

Extended agents like humans consist of both epistemic and action systems, and have many separate parts that all contribute to this. We have personalities that structure our overall lifestyle. We form habits and intentions that guide our behavior over shorter extended periods. We make plans for specific future events. We perform actions in the moment. All of these systems are dedicated to achieving whatever it is that we value in our many ways.

Some situations, like Newcomb-type problems (including prisoner's dilemmas, Kavka's toxin puzzle, and others) make it so that the habits and virtues that are most effective at promoting one's ends overall lead to particular actions, but other actions are the ones that are most effective in the moment at promoting one's ends. Thus, on my characterization, the rational habits lead to irrational actions, and the rational actions are only promoted by irrational habits.

It seems to me that some of these points are similar to issues that are familiar from rule vs act utilitarianism. And the prisoner's dilemma is also often described as a tragedy (particularly in the phrase, "tragedy of the commons").

EDIT: I wish people wouldn't delete questions that got answered. I believe this one was asking something useful about whether the view of a "tragedy of rationality" I talk about in the Rationally Speaking podcase is defensible (though I think the question also had a bit of harshness to it that might have prompted the take-down).

One box or two box?

SSBMPuffDaddy

I have a really hard time with this, but probably one. I certainly would like to be the kind of person that one boxes (but I'd like to accidentally grab the second box as well).

Do you think we need to develop theories of reasoning under normative uncertainty? If we are uncertain about which normative ethical theory or decision theory is correct, should we apply a meta-decision theory? We could just try to "maximize expected moral value" across different normative theories, but it's not clear how to actually compare moral value across wildly different ethical theories. One proposal put forth by Nick Bostrom and William MacAskill that sort of addresses this is that we should imagine a parliamentary model where representatives of each normative theory get votes in proportion with the credence assigned to them, and the decision is based on the final tally (possibly using something like ranked or range voting).

merelor

I think we do in fact find ourselves in situations of normative uncertainty (though a lot of apparent normative uncertainty is probably in fact empirical uncertainty - a lot of disagreement about euthanasia and the death penalty turns on how frequently people are in fact wrong about who is guilty or whether their medical condition is untreatable). As a result, we do in fact reason under normative uncertainty. Thus, if we want to understand how we do reason, and how we could do it better, then having theories for reasoning under normative uncertainty would be useful.

One problem with a lot of the discussion on this sort of topic though is what level of "should" we're going for. If you want to know what would overall be best to do in cases of normative uncertainty, it's the thing that is recommended by the correct normative theory. But that advice is no help to someone who has normative uncertainty.

Given that the correct normative theory may well be the sort of thing that is knowable a priori (if there is a fact of the matter here), it's going to be hard to avoid this conclusion in an ideal theory.

Instead, we're going to have to follow political scientists in terms of talking about non-ideal theories that still have some normative status. Any such theory is automatically going to have problems, because it allows for people to do things that are wrong while believing that they are right. Good advice here will probably depend on lots of empirical knowledge about the types of scenarios that tend to be affected by this sort of uncertainty, and the things that might go wrong in each. I don't think that a priori reasoning about this is going to be that helpful.

All of this goes equally for the problem of peer disagreement in epistemology. (It's no use asking for what would be ideal in a case where we know at least one person thinks something non-ideal and we don't know which.)

u/MaceWumpus asked in the announcement thread:

Dr. Easwaran, I'm looking for arguments for the use of Bayesian (or more broadly probabilistic) treatments of confirmation, which I've had a surprising amount of trouble finding. Howson and Urbach (for example) seem to argue that the main reason to be a Bayesian is that alternative pictures of confirmation, such as those involved in classical statistics and in the philosophies of Hempel and Popper, are worse. Are there any overviews or particularly good papers / books on the subject that you'd suggest?

lapse_of_taste

I was going to list several papers, but when I googled them, I found that most of them come up on the syllabus for Branden Fitelson's course on confirmation: http://fitelson.org/confirmation/syllabus.html

Branden was my PhD advisor, and a lot of his work has been on different bayesian measures of confirmation, and the way they do and don't respond to the traditional paradoxes from Hempel and Popper and others.

I would particularly recommend the papers he lists under weeks 6 and 7 if you'd like the relation between Hempel and Bayesian views.

To preface, I know nothing about logic or philosophy. Has there been any effort to define or view strict mathematical arguments as something other than arising from a sequence of logical steps? My point is that the idea is to write out contemporary mathematical ideas as a sequence of verifiable steps is an impossible proposition.

I tend to view mathematical arguments as locally true, without knowing what that means. Yesterday I was saying to a colleague that we had forward result and a partial converse, so by the mean value theorem of theorems there must be a characterization in the middle if we change the hypotheses of the two results. While a joke, there is some truth to statements like that.

EnteredInfo

I think in philosophy of mathematics, it's now a fairly widely accepted view that rigorous mathematical arguments are often usefully modeled as a sequence of logical steps, but definitely are not such formal objects. No one does all the cases, or all the steps, in a familiar type of reasoning, and while the contemporary computer-aided proof movement wants fully formalized arguments, no one else does.

Bill Thurston has a classic paper setting out the issue: On Proof and Progress in Mathematics

Don Fallis has some good papers attacking the formal proof understanding of rigorous proofs: Intentional Gaps in Mathematical Proofs; What do Mathematicians Want?

Catarina Dutilh Novaes has also done interesting work on this (I'm not sure if this is the most relevant paper, but it's one)

Jody Azzouni defends a "derivation indicator" view that is perhaps closer to the sequence of logical steps one.

And you might be interested in my papers on probabilistic proofs, and rebutting and undercutting defeat in mathematics.

What do you think of Pascal's wager or infinite utilities in general?

What about using expected value maximization in cases where our credence is very tiny but the utility is even more astronomically huge but still finite? Do you think there are any finite cases where expected utility maximization is inappropriate?

merelor

I'm skeptical of actual infinite utilities for particular outcomes, as in Pascal's wager. Even if there are outcomes with infinite utilities, I think that it's better to think in terms of whole ranges of infinities rather than there being discrete "infinite" value that can't be improved or worsened by a small amount. (Technically speaking, that means I prefer a non-archimedean field rather than the extended reals or Cantorian cardinalities.)

I think "expected value maximization" is something that we implicitly do subconsciously in forming our preferences, rather than a mathematical technique that we should be explicitly imposing to regiment

our preferences. It's hard to figure out what your own actual credences and utilities are, though maybe it's easier to figure out what certain external policies would recommend. There are probably cases in business and politics and other social endeavors where we want to agree on some values and some probabilities and decide what to do collectively, and explicitly doing expected value calculations with externally agreed on credences and utilities *might* often help there. But I suspect that cases where the probability is extremely small and utility is extremely large are cases where apparently small disagreements about the procedure can lead to extremely large disagreements on action, so we have to be very careful.

I don't think expected utility maximization is inappropriate when you have the right probabilities and utilities. But given that those are hard to explicitly come by, there are probably many situations in which certain rules of thumb are better than the explicit calculation. (I'm thinking of things like trolley problems where our intuition says don't push the guy in the way of the trolley because there's a good chance he won't actually stop it, even though the philosopher is telling me for sure that the guy will stop the trolley.)

What meta-ethical and normative ethical theory do you subscribe to?

AronBhalla

I'm not as much of an expert on metaethics and normative ethics as on issues in logic and epistemology, but I do have some views here.

My overall thought is that all value comes from the goals of things with desires or purposes. I'd like to be able to derive some sort of consequentialist desire-satisfaction utilitarian view of ethics, but there are some missing steps.

Hi Kenny.

Can one doubt the truth of the a priori? By extension, can we be perfectly certain of logical and mathematical claims, even tautologies?

captain__sock

I think most people that have ever studied math or logic have had the experience of doubting something that later turned out to be provably true. So one certainly *can* doubt the truth of the a priori.

The deeper question is whether one ever *should* doubt the truth of the a priori. I think a lot of philosophers have for a long time said no - one should in some sense ideally always be certain of the a priori. I think there's room for multiple senses of "should" here - in some logically idealized sense, one should already recognize which configurations are actually logically possible and shouldn't doubt the tautologies and mathematical claims. But in another sense, where one considers possibilities where 1947563473 is prime as well as possibilities where it is not, I think one has no internal pressure to be certain of the one that is in fact logically true.

For the converse question, I think it's important to distinguish certainty as in having no doubt, from certainty as in impossibility of error. One certainly *can* lack all doubt for some logical and mathematical claims (and one can also lack all doubt for some empirical falsehoods too! for a non-trivial example consider an 18th century physicist thinking about Euclidean geometry in space). As for whether one *should*, I'd probably need to think more about the notions of "should" involved.

Professor Easwaran, Thank you for doing this AMA. I just finished reading a book on the Kuhn vs

Popper debate and I was left astounded by how various theories and meanings are relativised and distorted by the (not exclusively) intelectual elite in order for it to strengthen and maintain their position of power.

I would love to get a brief comment from you regarding these two very different approaches on science and epistemology and current philosphical schools of thought that revisit and/or defy the Kuhnian basis of the Example (Paradigm?) of modern science.

Best regards

chaositect

Some people see Popper as telling us what rational science is (ideal scientists instantly give up on theories once they're falsified) while Kuhn tells us about the irrational behavior of people (famous old scientists hold onto their theories and suppress competitors while they're still in power).

However, I think Kuhn's picture is more subtle than that. Our evidence is never definitive at falsifying hypotheses (Newtonian gravitation looked equally falsified by orbital irregularities of Uranus and Mercury, but it was saved from both by postulating gravitational interactions with unobserved planets Neptune and Vulcan, though Vulcan needed to have some odd properties to be compatible with observations). Sometimes, the epicycles needed to maintain an old theory end up being productive, and sometimes they don't. Science works better if some people pursue each version of the theory, so we need the old guard around inspiring some people to defend the modified theory just as we need the new guard trying to push strange alternatives.

Usually, the young revolutionaries are totally wrong, and the old guard working within an established framework is able to develop the theory in interesting and progressive ways. Occasionally they're right, and it takes a generation for science to catch up. This is probably better than many other possible social arrangements for the production of science, at least in terms of discovery and understanding of the world.

There's a rationality community (so to speak), associated with LessWrong, AI risk research, and the Machine Intelligence Research Institute. What do you think of it, and of the general prospects/importance of AI risk research?

studyinglogic

I've found this collection of people really interesting to talk to. I've been invited to MIRI a few times to talk with them, and they also put on a conference with some philosophers at Cambridge on self-prediction in decision theory a few years ago, which I went to and enjoyed.

I'm not as convinced as they are of the importance of the issues in AI risk that they discuss (I had a long and interesting discussion with a bunch of people on Facebook several weeks ago about the ways in which AI risk is similar or different to the risk in other sorts of complex intelligent systems, like the risk that some attribute to neoliberal market capitalism).

But from my own academic perspective, I can say that the set of views around Newcomb-style problems that they've put together are some of the most interesting new ways to justify some intuitions that I've seen in quite a while. And their paper on logical induction is a useful breakthrough for that topic as well, even if it doesn't yet address the problem that motivates them in decision theory.

I have two questions regarding formal representations in philosophy:

- What do you think is the right (if that's the best way to think of it) representation of beliefs and

decisions? (Example: For beliefs, should it be AGM belief revision, or Bayesianism, or Dempster-Shafer theory, or some other model? For decisions, should it be causal decision theory, or evidential decision theory, or prospect theory, or ... ?)

- To what extent do you think these formal models track how things really are? (In whatever way you wish to interpret the phrase "how things really are.")

studyinglogic

I think this question is lurking in a lot of my papers, but not really explicit. I think there's one question here about how humans actually think and act, and another question here about what is the best model for thinking and acting. My general view is that there are probably many different models for thinking and acting that could all work well in their own way. Bayesianism, or some more sophisticated version of AGM, or something Dempster-Shafer-like, could all probably work.

At the moment, Bayesianism is the only one that I know of that hooks up nicely with decision theories (and I think there we need something like Savage or Buchak as the base theory, and all the Newcomb-type problems need to be translated into forms where there is act-state independence, rather than CDT or EDT).

But I wouldn't be surprised if there are multiple formal models that all work well. (One point of my Dr. Truthlove paper was that there might be two superficially very different formal models that come practically to nearly the same thing. And my recent paper in *Res Philosophica* on the Tripartite Role of Belief suggests that a parallel thing might be true for whether accuracy, action, or evidence is the fundamental gold of belief.)

As for what *humans* actually do, I would need to do a lot more empirical work. Maybe we don't quite behave in sophisticated enough ways to match any of these theories, and something more like prospect theory is a more accurate account. More likely, what we do is actually totally different and any theory like these is just an approximation. (Consider the relation of thermodynamics based on caloric to modern statistical mechanics.)

But in any case, I don't think we should postulate structures that require incredibly complex set-theoretic constructions beyond ZF set theory (like hyperreal analysis, or finitely additive functions on the full powerset to deal with issues for conditional probability that Dubins and de Finetti were interested in).

What are the most important areas of mathematics that a formal epistemologist needs to know?

More generally, what should a formal epistemologist know outside of work done in philosophy?

SamuelTXKhoo

The most important general areas of mathematics for formal epistemology are probability and logic. It's also helpful to have measure theory (and thus some real analysis) and also some general topology. But I think any of the mathematics that you'll need to know will be things that you can learn when needed, as long as you get some substantial mathematical training that enables you to follow some abstract proofs about new definitions.

I think the fields outside of philosophy and mathematics that are most relevant are psychology, economics, and computer science. I wish I had actually taken some of those classes while I was a student!

What snack do you get at the movies?

bambooslashbang

It's rare that I go to a theater for movies any more (I've averaged maybe once or twice a year for a while), and when I do go I usually don't get a snack. Popcorn gets too stuck in my teeth, and I'm usually not interested in sweet snacks.

Do you believe numbers and mathematical entities exist in any real way?

Do mathematics emerge from observation of a material world (without anything to count does math mean anything?) or does a material world exist because of mathematical principles?

Bjarki56

This was a question that gripped me deeply during graduate school, and I was convinced of a nominalist view like that of Hartry Field, on which mathematical entities (and other abstract entities) don't really exist. However, after reading more, and attending various metaphysics conferences, I've become less convinced that I even understand what the notion of "existence" here is anyway!

These days I'm still thinking about topics related to the work of Field. But my viewpoint is more that any *application* of mathematics should always be grounded in some purely internal, physical description of the system, and that we should understand *which* properties of the mathematical objects are the ones that represent meaningful facts about the system we're describing, and which don't.

I'm still tempted both by the nominalist view that all there is is physical stuff with mathematics a human invention to describe it, as well as the opposite, quasi-Pythagorean view that everything is fundamentally a big abstract mathematical system that we just experience as physical from our position in it. But I'm not convinced there's anything useful for me to say about these views, so I focus more on the issues of what it means to use mathematics in any particular application.

Do you find induction or deduction more useful in economics?

grasha87

I've never actually done much work in economics myself. But I think for actual reasoning, induction is more useful almost everywhere (even in mathematics!) Deduction is useful once you've found your conclusions and want to write them up in a way that other people will also accept.

In the course of doing research, I've put together some thoughts that could have strong implications for some philosophical problems such as the problem of logical omniscience and Bayesianism on propositional statements. At some point I would like to engage with the philosophical community on it. What are some preliminary steps I can take?

amateurtoss

I think the first thing to do is to read some existing work on this problem, by people like Hacking, Garber, Gaifman to see where your view is different and how it fits in. If you don't already know some people working on these issues within the philosophical community, you should get to know someone and establish conversations with them - I don't know if you're a student or an academic or a researcher without these sorts of affiliations, but your opportunities for getting to know philosophers as people to establish conversations will be different in these circumstances. I think it will be hard to get philosophers (just like members of any intellectual community!) to read your work if you haven't done these steps first, given that there's already too much interesting stuff to read all of it, and it's much

harder to read work that comes from a very different angle if you don't already have some understanding of where the person is coming from.

There's a bit of very important work that I see being developed on this topic by non-philosophers, but even with a lot of discussion, it takes time to get this written up in a way that is fruitful for the cross-community interaction.

Do you ever refer to Taoism or Buddhist philosophy as an example when discussing the duality of humam perception?

Harveythepookah

I haven't thought a whole lot about perception itself in my work. I would like to be more familiar with Taoist and Buddhist ideas about the persistence of the self through time, but I don't know enough yet.

Professor what do you think is better for learning philosophy, studying the history of philosophy starting by Ancient Greece or studying it by topic (epistemology ,metaphysics ..) ??

FrankRuX

I think it depends on what personally motivates you. For me, there were particular topics that I found interesting that brought me in, and then seeing certain names come up repeatedly in many of the discussions gave me interest in learning more about those bits of the history. For other people, the history of ideas might have more interest for them, and particular problems that keep coming up might then come later. And just as different people might be motivated by different topics, I think different versions of history might also work - you could have an intro that does the history of western philosophy jumping from Greeks to Descartes and the early moderns, to analytic and continental philosophy. But you could also have a very different version that traced the history of Buddhist thought and led to thinking about issues in the philosophy of mind and the relation of action and ethics.

What do you think of the Lottery and Preface paradoxes? Do you think they need to be solved - and if so, how should they be solved?

SamuelTXKhoo

I'm sometimes tempted by the view that there's just degrees of belief, and both of these paradoxes go away. But I think there is a way that we often talk about a binary concept of belief, and these two cases illustrate something interesting about it.

They're formally similar: both consist of a collection of independently likely propositions (ticket #38173 won't win; claim 38 on p. 173 of the book is true) and one belief stating that not all of these independently likely propositions are true. However, most people feel that in the lottery case we don't believe the individual propositions, but in the preface case we do. (There are others that say we do believe the lottery propositions.) The fact that we have different intuitive responses to formally similar cases suggests that some other feature is relevant.

My thought is that if there's some interesting notion of belief beyond probability, then it must track some difference between these cases. I think that one important thing is that the evidence we have for the lottery propositions is not sensitive (if ticket #38173 were the winner, I'd still have exactly the same evidence I do now) while the evidence we have for the book propositions often is (if claim 38 on p. 173 were false, I probably would have observed something different in my research).

I don't know of any theory linking probability and full belief that accommodates this idea (though I do know several views that accommodate one or the other).

Every one of our behaviors has a chaotic influence on the environment, which is clear from chaos theory. That is, it's impossible to predict the consequences of our actions. Given that this is the case, what is the point of trying to apply logical reasoning to our behavior when we can't know the consequences of it?

EntropyAnimals

It depends on what you mean by "impossible to predict". It's true if you mean it's "impossible to be extremely certain to an extreme degree of precision". But it's in fact quite possible to predict (to a moderate degree of confidence) the consequences of our actions (to a moderate degree of precision). I am quite confident that if I pour water into the pot and put it on the stove and turn on the gas, the water will boil in a few minutes (even though there might be a gas line outage in the next few seconds, or there might be a nuclear strike). I know that boiling the water will be helpful if I want to make pasta or cook an egg. I think that understanding these sorts of things were major advances for humanity. So I think that trying to extend this sort of knowledge to greater confidence and greater precision is useful, even if we can't get perfect.