Alan Hájek
Professor of Philosophy
 Philosophy Program
Research School of Social Sciences
Australian National University, Australia

Why were you initially drawn to epistemology (and what keeps you interested)?

If there’s any truth to the old saw ‘Show me the child, and I’ll show you the man’, then I suppose my philosophical infancy has a lot to answer for. Let me begin by explaining how I was initially drawn to philosophy.

I was always torn between the sciences and the humanities, but I finished up doing a science degree at Melbourne University, majoring in mathematics and statistics. I loved probability theory, but I wanted to know more about that ‘P’ that kept appearing in its equations and theorems. When I asked one of my professors about it he was unimpressed. He remarked, somewhat sneeringly, that some people regarded it as a subjective notion. It was clear to me that he did not encourage this line of enquiry. So I continued learning more theorems. I got the impression that all the low fruit had already been picked—it seemed that every significant result had either been proven already, or would be prohibitively difficult to prove. One day in the university café, I was poring over an article on stochastic monotonicity. Like Jules in Pulp Fiction, I had a moment of clarity: I realized my future lay elsewhere.

But where? Faced with this intellectual crisis, like a good Australian I decided to travel around the world, to ‘find myself’ on the road (as they used to say). The remarkable thing is that I did. My second moment of clarity occurred in London, Ontario, of all places. I was visiting a friend of mine, another science undergraduate from Melbourne Uni, who had seen the light first and was
now studying philosophy at the University of Western Ontario. I looked at what he was studying, and I thought ‘Wow! I wish I were doing that!’ The penny dropped so loudly it could be heard several provinces away. I picked up the application form for their graduate program that day. Later that year they were kind enough to admit me despite my complete lack of training in philosophy. (All I knew about philosophy was the Monty Python Philosophers Song.) And before I knew it, I was a philosopher.

It turned out that Bill Harper also wanted to know more about that ‘P’ that kept appearing in the equations and theorems of probability theory, and he was my first philosophy mentor. I completed an MA at Western. I went on to do my PhD at Princeton, under the supervision of Bas van Fraassen and David Lewis, and the further influence of Dick Jeffrey and various terrific students. I wrote a dissertation on probabilities of conditionals and conditional probabilities. I imprinted on Bayesian epistemology, and I never looked back. I liked the idea that belief comes in degrees, and I was gripped by the Bayesian view that those degrees should conform to the probability calculus. (My training in probability theory wasn’t wasted after all!) A formal theory searching for an interpretation met a notion of graded belief searching for a formal theory. A blissful marriage!

Or so it seemed to me at the time. More recently, I have questioned various aspects of Bayesian epistemology, and this is an ongoing line of research in mine now. Along the way I have also become fascinated by Moore’s paradox. This brings me to the next question ...

What do you see as being your main contributions to epistemology?

Williamson (2000) is well known for advocating a ‘knowledge first’ epistemology. I am attracted to an ‘uncertainty first’ epistemology, with that uncertainty codified by probability in the Bayesian tradition of Ramsey, de Finetti, Jeffrey, and others. And much as Williamson takes the concept of knowledge to be unanalyzable, so I take the concept of degrees of belief to be unanalyzable. (I spell this out in more detail in part of my contribution to Erickson and Hájek 2007.) But taking a concept to be primitive in this way does not preclude us from saying many illuminating things about it. For starters, the axioms of probability theory already say a lot about it, the theorems still more. The theory is wonderfully simple, yet extraordinarily powerful. Indeed, it achieves such a balance of simplicity and strength that, by analogy to Lewis’s analysis of ‘laws of nature’, I am tempted to say that probability theory codifies the laws of epistemology—or at least, the laws of ‘uncertainty first’ epistemology, where this uncertainty is rationally constrained.

A garden of philosophical delights awaits you once you embrace Bayesian epistemology. Old chestnuts in confirmation theory, such as the ravens paradox, the grue paradox, and the Quine-Duhem problem, suddenly come into relief and can be given a unified treatment. (See Hájek and Joyce 2008.) And unlike traditional epistemology, trafficking as it does in the all-or-nothing notions of knowledge and belief, Bayesian epistemology naturally underwrites decision theory. In the next section I will discuss how well Bayesianism serves various areas of philosophy.

And yet I have become something of a Trojan Horse. I come to Bayesianism as a friend, but I have misgivings about some of the most important arguments that degrees of belief – ‘credences’ – are rationally required to be probabilities. (See especially my 2009.)

Take, for example, the Dutch Book argument. We identify your credences with your betting rates, and then appeal to the Dutch Book theorem: if your credences violate probability theory, then there exists a set of bets, each acceptable according to your betting rates, which collectively guarantee your loss. This is often made vivid with the image of a bookie, who for some reason is supposed to be Dutch, buying or selling bets to you at prices that you find acceptable, and thereby draining you of your money. Thus, the argument concludes, your credences should obey probability theory.

But the argument faces a dilemma. Either we take this talk of ‘guaranteed loss’ literally, or not. The former, rather flat-footed interpretation of the argument, is easily rebutted: as various authors have pointed out, you can simply refuse the bets. As I would put the point, you can mask your disposition to accept the bets with a stronger disposition to do something else—e.g. to walk away when you see the Dutch guy coming, or to mimic having different betting dispositions. Surely Skyrms (1984, 1987) is right that we should not take the cautionary tale literally. He goes on to

---

1 According to Lewis (1973), a law of nature is a theorem of the best systematization of the universe—the true theory that best balances simplicity and strength.
argue that it merely dramatizes a deeper defect: an *inconsistency* in your evaluations. (Compare Ramsey’s 1990/1931 remark that probability theory provides a *logic* for partial belief.)

However, this interpretation of the Dutch Book argument *doesn’t* seem right to me. For starters, if there is an inconsistency at all, it seems to be some sort of ‘pragmatic’ inconsistency—perhaps troubling all the same, but surely not on all fours with logical inconsistency. But it is also not clear that there need be even a pragmatic inconsistency. For the Dutch Book argument presupposes a *package principle* that requires one to value a set of bets at the sum of the values of the bets taken individually, or less specifically, to regard a set of bets as fair if one regards each bet individually as fair. The package principle seems especially problematic when there are interference effects between the bets in a package—e.g. the placement of one bet is correlated with the outcome of another—and when the package is infinite (see Arntzenius, Elga, and Hawthorne 2004). So the whole is not just the sum of the parts when it comes to betting packages. Moreover, McGee (1999) offers an ‘airtight’ Dutch Book against any agent who has an unbounded utility function and who assigns positive probability to infinitely many possibilities, but it is hard to see how there is any inconsistency in that—or if this is some kind of ‘pragmatic’ inconsistency, then there doesn’t seem to be anything irrational about it. And Briggs (forthcoming) shows how you are Dutch Bookable if you are unsure of your own credences, or if your credences are Moore paradoxical in a certain way. Neither seems to be a case of logical inconsistency, even in some extended sense. I have pursued these themes further in my 2005, 2008b, and 2009.

This gives me the smoothest segue that I can manage to a completely different strand of my work in epistemology: Moore’s paradox (Hájek and Stoljar 2001, Chalmers and Hájek 2007, Hájek 2007, Williams and Hájek MS). G. E. Moore famously observed (1942) that it would be “absurd” to assert a sentence of the form ‘p and I don’t believe that p’—e.g., ‘it is raining and I don’t believe that it is raining’. Since then, a vast literature has appeared, attempting to explain just what this “absurdity” consists in. Here is my top-ten list of reasons why the paradox fascinates me (although I don’t expect to see it on the *David Letterman Show* any time soon):

1. Moore sentences can obviously be true—e.g. it really might be raining, and yet I really don’t believe that it is. They are thus not semantically inconsistent. Yet equally obviously, there is typically something defective in asserting or believing them. In particular, it seems that the very act of asserting or believing them should somehow render them false. This suggests that there is another kind of inconsistency—call it pragmatic inconsistency.

2. In putting constraints on reasonable assertion, Moore’s paradox has consequences for any theory of assertion.

3. In putting constraints on reasonable belief, Moore’s paradox has consequences for any theory of belief.

4. The paradox beautifully brings out an important difference between what philosophers call ‘first person’ and ‘third person’ perspectives. Note that there is nothing defective about my asserting or believing, say, ‘It is raining and Renée believes it is not raining’. It is only in Moore sentences about oneself that things go awry.

5. Moore sentences involve ‘higher-order’ thoughts—e.g. beliefs about one’s own beliefs. Some philosophers (e.g. Rosenthal 2005) think that such thoughts are the basis of consciousness.

6. There is apparently something *irrational* about asserting or believing Moore sentences. They thus shed some light on the nature of rationality.

7. They provide instances of what Sorensen (1988) calls ‘blind-spots’: sentences that can be true, but that cannot rationally be believed. It’s remarkable that Moore sentences seem to involve *individual* blindspots, each involving exactly one person. For example, ‘It is raining and I believe it is not raining’ could be true, and it could rationally be believed by everyone except me. I am the only possible agent ‘blind’ to its truth.

8. Moore sentences are subtly implicated in other philosophical problems and paradoxes—e.g. the surprise exam paradox, according to Sorensen and others.

9. Moore’s paradox provides philosophers with a new weapon in their philosophical arsenal: a new kind of *reductio ad absurdum*. In a usual reductio, some assumption is shown to lead to inconsistency. But it is also salutary if an assumption can be shown to lead to *pragmatic* inconsistency, à la
Moore's paradox. David Chalmers and I deployed this technique in our "Ramsey + Moore = God" (2007). We argued that the Ramsey test for the acceptability of an indicative conditional, as it is usually understood, leads to Moorean absurdity. It commits you to accepting all instances of conditionals of the form 'if p, then I believe that p', which commits you to accepting that you are omniscient, and all instances of the form 'if I believe that p, then p', which commits you to accepting that you are infallible—on pain of Moorean absurdity. Since Moorean absurdity is, well, absurd, as is the extreme hubris of attributing to yourself God-like epistemic powers, this is a reductio of the Ramsey test, so understood.

10. And yet certain philosophical positions may be able to embrace their own Moorean reductio. For example, eliminativists about beliefs (e.g., Churchland 1981, Stich 1983) argue that 'belief' is part of a suspect folk psychology, ultimately to be discarded by science. Thus, Churchland and Stich are committed to uttering sentences of the form: "it is raining and I don't believe it is raining (and neither do you, nor anybody else, for 'belief' is part of a suspect theory of the mental)". Now, Churchland and Stich would not be fazed by this gambit—not by the putative reductio that 'they offer a philosophical position that they don't really believe!' The right thing for them to say, of course, is that their mental state regarding the rain, or their own position, is whatever the fully developed psychological theory postulates it to be. Still, their commitment to Moore sentences is genuine, as is the puzzlement that it may induce in many of the rest of us.

In my (2007) I canvas a series of philosophical positions that apparently have such Moorean commitments. In some cases, the commitments seem to be fatal to the positions; in others, the commitments seem not so troubling after all. Either way, the situation is interesting. And to the extent that the commitments are not so troubling, it seems that pragmatic inconsistency cannot be so readily assimilated after all to logical inconsistency, which really is troubling, Priest (1987) notwithstanding! (Compare my remarks above about the Ramsey/Skyrms interpretation of the Dutch Book argument.)

\[ P(A \rightarrow B) = P(B|A) \quad \text{(provided } P(A) > 0) \]

But let's return to probability theory and Bayesian epistemology, since I'm back at work on that. Recently I have been completing a book manuscript entitled Arrows and Haloes: Probabilities, Conditionals, Desires and Beliefs. I argue that two seemingly disparate debates are strikingly similar. The first concerns the thesis, associated with Stalnaker (1970) and Adams (1975), that probabilities of conditionals are conditional probabilities:

\[ V(A) = P(A^n) \quad \text{(provided } P(A) > 0) \]

Here, \( A \rightarrow B \) should be interpreted as the conditional 'if \( A \) then \( B \)'. The second concerns the thesis, associated with certain anti-Humeans, that desires are beliefs (as it might be, my desire for Tilly to win is a belief that her winning would be good). The thesis was formulated by Lewis as follows, and then attacked by him:

\[ P(A \rightarrow B) = P(B|A) \quad \text{(provided } P(A) > 0) \]

Here, \( V \) is expected value as computed by evidential decision theory, and \( A^n \) is naturally interpreted as '\( A \) is good'.

Notice that both theses employ Bayesian resources. Indeed, Stalnaker and Adams specifically intended the 'P' in the first equation to be interpreted subjectively. Likewise, the 'P' in the second equation should obviously be interpreted subjectively: it is a 'desire as belief' thesis, after all. Each thesis, if true, would tell us something interesting about rational mental states. In the first case, a rational agent's attitude to conditionals is constrained by her conditional probabilities, which in turn constrain her updating dispositions. (Shades of the Ramsey test here.) In the second case, desire-like states appear to be reducible to belief-like states. In an inversion of Hume's famous dictum, the passions are the slave of reason.

Lewis (1976, 1986) famously offered 'triviality' results against these theses. We have yet another kind of reductio ad absurdum argument—the theses don't quite lead to contradiction, but they putatively lead to the unpalatable conclusion that the P's (and V's) that can sustain the equations are somehow radically impoverished, unworthy of your typical rational agent. Notice that I put this point a little cagily—for it seems that a rational agent could have a radically impoverished probability function. For example, the functions that sustain the first equation are trivial because they are at most four-valued. But an ideally rational agent could have such a function. Indeed, an ideally rational omniscient God
presumably has a two-valued function—he or she assigns probability 1 to the actual world, and 0 to all others! The point is that the equation is supposed to govern the credence functions of all rational agents, and rationality surely permits non-trivial functions. Various other authors since Lewis have piled on further triviality results against the equation; mine can be found in my (1989), (1994), (1996) and Hájek and Hall (1994). But as usual, there are ways of fighting back (e.g. van Fraassen 1976)—this is philosophy, after all.

A few years ago I noticed some structural parallels between the two theses, and between Lewis’s triviality results against them, which I found striking. Then, to an extent that surprised me, I realized that one could mimic the moves and countermoves in one of the debates that had already been made in the other. I have found this to be illuminating about the debates themselves, and it has also provided a handy heuristic for generating new results. Watch out for my book (forthcoming) when it hits a bookstore near you—avoid the lines, avoid disappointment and order your copy early!

And so here I am, returning full circle to work I began in my PhD dissertation. Talk about regressing to one’s (philosophical) childhood!

What do you think is the proper role of epistemology in relation to other areas of philosophy and other academic disciplines?

Picture philosophical topics like a big subway map. For me, probability is Grand Central Station. From there, I can get almost anywhere I want to go (well, maybe with a transfer or two). So I will answer a more specific version of this question: what I think is the proper role of Bayesian epistemology in relation to other areas of philosophy and other academic disciplines.

It’s hard to know where to draw the line between the use of Bayesian methods, and Bayesian epistemology per se. If the methods count as part of the epistemology itself, then I might as well list every science in my list of other academic disciplines informed and shaped to some extent by the epistemology. But let me confine myself to some of the roles Bayesian epistemology plays in philosophy, where its employment is especially self-conscious and distinctive. Again, it makes its way into almost every major branch of philosophy. In epistemology, philosophy of mind, and cognitive science, subjective probability functions model states of opinion. Since subjective probabilities are found at the heart of decision theory (and to some extent, game theory), they have ramifications for ethics and political philosophy. Subjective probability appears again in the philosophy of science in the analysis of confirmation of theories, in statistical inference, and in the rational reconstruction of various episodes in the history of science; and in the philosophy of specific scientific theories, such as quantum mechanics, and evolutionary biology. Bayesian methodology can play a central role in metaphysics, the philosophy of logic, the philosophy of language, and even the philosophy of religion. Just to take the last case, I have written extensively on Pascal’s Wager (1997, 2000, 2003, 2008c), and on Hume’s miracles arguments (1995, 2008a) from a Bayesian perspective.

Now there’s a philosophical Grand Central station for you!

What do you consider to be the most neglected topics and/or contributions in contemporary epistemology?

And yet traditional epistemology and Bayesian epistemology seem to be on completely separate tracks. Think of some of the time-honored debates in traditional epistemology: skepticism, Gettierology, reliability, internalism vs externalism, foundationalism vs coherenceism. Think of some of the currently hot topics: contextualism, subject-sensitive invariantism, relativism, luminosity, ‘knowledge how’ (as opposed to ‘knowledge that’), knowledge ‘wh—’ (who, where, when), ... Where are the counterpart debates in Bayesian epistemology? Going in the other direction, think of some of the time-honored debates in Bayesian epistemology: constraints on priors, updating rules, the extension of subjective probabilities to infinite spaces. And think of some currently hot topics: credences about chances (as codified in Lewis’s Principal Principle, 1980), credences about one’s future credences (as codified in van Fraassen’s Reflection Principle, 1984), updating credences on ‘centered’ or ‘indexical’ propositions, ... Where are the counterpart debates in traditional epistemology? It’s not much of an exaggeration to say: wherever we have a debate in one of the epistemologies but not the other, we have a neglected topic.

A topic that has received some attention, but that deserves more, is the relationship between subjective probability, and the staples of traditional epistemology—knowledge, belief, truth, justification, and maybe some kind of ‘fourth condition’ that yields
knowledge when added to justified true belief. For example, there has been a recent resurgence of interest in versions of the so-called ‘Lockean thesis’ that belief (or perhaps rational belief) can be reduced to credence—e.g. belief is credence above some threshold, perhaps contextually determined. Jeffrey thought that belief had passed its use-by date some time ago: “[I am not] disturbed by the fact that our ordinary notion of belief is only vestigially present in the notion of degree of belief. I am inclined to think that Ramsey sucked the marrow out of the ordinary notion, and used it to nourish a more adequate view.” (1970, 171-2). He is referring to Ramsey’s decision-theoretic approach to credence—“belief qua basis of action”, which is ultimately preference-based.

But credences, preferences, and actions are certainly separable in thought, and sometimes in practice. Imagine a Zen Buddhist monk who has credences but no preferences, or a chronic apathetic who has credences but no inclination to action. Moreover, credences explain other aspects of behavior besides those that are preference-related, and they explain much more than behavior. (See Christensen 2004 and Eriksson and Hájek 2007.) Offhand, I would have thought that belief’s primary job description is belief qua representing the way things are; action is downstream, both causally and conceptually. Furthermore, the ordinary notion of belief plays a vital role in psychological explanations, which is not obviously played by credence, however high. For example, we may say that a small child is reaching in the box simply because she believes her toy is there; to attribute a high subjective probability to her seems to over-intellectualize her psychological state.

And what about knowledge, truth, justification, and the American way? For starters, subjective probability theory seems to offer nothing corresponding to the factivity of knowledge. The Bayesian similarly lacks a notion of ‘justification’—or to the extent that she has one, it is too permissive. Any prior is a suitable starting point for a Bayesian odyssey—yet mere conformity to the probability calculus is scant justification. Now, the Bayesian will be quick to appeal to various convergence theorems. For example, Gaiman and Snir (1982) show essentially that for each suitably open-minded agent, there is a data set sufficiently rich to force her arbitrarily close to assigning probability 1 to the true member of a partition of hypotheses. And she is presumably justified in doing so, for it was her evidence that drove her there.

This is a striking theorem, but one should not overstate its epistemological significance (not that Gaiman and Snir do). It is glass half-full theorems, but a simple alternation of the quantifiers turns it into a ‘glass half-empty’ theorem. For each data set, there is a suitably open-minded agent whose prior is sufficiently perverse to thwart such convergence. And strong assumptions underlie the innocent-sounding phrases “suitably open-minded agent” and “sufficiently rich data set”. No data set, however rich, will drive a dogmatic agent anywhere at all. Worse, an agent with a wacky enough prior will be driven away from the truth. Consider someone who starts by giving low probability to being a brain in a vat, but whose prior regards all the evidence that she actually gets as confirming that she is. And we can always come up with rival hypotheses that no courses of evidence can discriminate between—think of the irresolvable conflict between an atheist and a creationist who sees God’s handiwork in everything. Finally, I don’t see how the convergence theorems help one iota in addressing simple skeptical challenges, such as how do I know, right now, that I have a hand?

As for the so-called ‘fourth condition’ on knowledge: various reliabilist and anti-luck epistemologists advocate versions of safety as a condition on knowledge—roughly, at the closest worlds in which a given agent believes p, p is true. Others advocate sensitivity—roughly, at the closest worlds in which p is false, the agent does not believe p. (And some advocate both.) But note well: the closest worlds. Here we find another disjuncture between traditional and Bayesian epistemology: nothing in the standard Bayesian apparatus reflects the notion of ‘similarity’ of worlds that has recently taken center-stage in the analysis of knowledge.

Bayesian epistemologists have become alive to such concerns. Indeed, a large part of what has come to be known as formal epistemology seeks to build bridges between traditional and Bayesian epistemology, and it is a movement that is building momentum. (I’ve tried to build a bridge—or perhaps a tiny platform—between the two epistemologies by offering a Bayesian analysis of ‘agnosticism’ in my 1998.) So I see hope that the long-neglected topics of connecting the traditional and the Bayesian concepts are now receiving their due attention. This makes me sanguine that there will eventually be a rapprochement between these approaches to epistemology.

What do you think the future of epistemology will (or should) hold?
Rapprochement.
References


Christensen, David (2004): Putting Logic In Its Place, New York: Oxford University Press.


Hájek, Alan (forthcoming): Arrows and Haloes: Probabilities, Conditionals, Desires and Beliefs, Oxford University Press.


Williams, John and Alan Hájek (MS): “‘p, and I Have Absolutely No Justification For Believing that p’; the Necessary Falsehood of Orthodox Bayesianism”.


I thank Elle Benjamin, Ralph Miles, and Declan Smithies for helpful feedback.