Masses of Formal Philosophy ‘Interview’

Alan Hájek

Why were you initially drawn to formal methods?

I came to philosophy as a refugee from mathematics and statistics. I was impressed by their power at codifying and precisifying antecedently understood but rather nebulous concepts, and at clarifying and exploring their interrelations. I enjoyed learning many of the great theorems of probability theory—equations rich in ‘P’s of this and of that. But I wondered *what is this ‘P’?* What do statements of probability *mean?* When I asked one of my professors, he looked at me like I needed medication.

That medication was provided by philosophy, and I found it first during my Masters at the University of Western Ontario, working with Bill Harper, and then during my Ph.D. at Princeton, working with Bas van Fraassen, David Lewis, and Richard Jeffrey—all deft practitioners of formal methods. I found that philosophers had been asking my question about ‘P’ since about 1650, but they were still struggling to find definitive answers. I was also introduced to a host of other philosophical problems, and it became clear to me within nanoseconds of arriving at U.W.O. that I wanted to spend my life pursuing some of them. But I kept being drawn back to the formal methods of mathematics, and in particular of probability theory.

It may be worthwhile to pause for a moment and to ask “What *are* formal methods?” Of course, it’s easy to come up with examples: the use of various logical systems, computational algorithms, causal graphs, information theory, probability theory and mathematics more generally. What do they have in common? They are all abstract representational systems. Sometimes the systems are studied in their own right for their intrinsic interest, but often they are regarded as structurally similar to some target subject matter of interest to us, and they are studied to gain insights about that. They often, but not invariably, have an axiomatic basis; they sometimes have associated soundness and completeness results. There is something of a spectrum of 'formality' here. At the high end, we have, for example, the higher reaches of set theory. At the low end we have rather informal presentations of arguments in English in 'premise, premise … conclusion' form. Higher up we find more formal representations of these arguments, whittled down to schematic letters, quantifiers, connectives, and operator symbols. Near the top we find Euclid's *Elements;* lower down, Spinoza's *Ethics.*

Formal systems typically facilitate the proving of results about some domain. They often provide a safeguard against error: by meticulously following a set of rules prescribed by a given system, we minimize the risk of making illicit inferences. I was struck by how one could start with a rather imprecise philosophical problem stated in English, precisify it, translate it into a formal system, use the inference rules of the system to prove some results about it, then translate back out to a conclusion stated in English. I liked the rigor, the sharpening of questions and their resolution, and the feeling that one was really getting *results.*

I was also impressed by how formal systems could stimulate creativity. Staring at the theorems of a particular system can make one aware of hitherto undiscovered possibilities, or of hitherto unrecognized constraints. (I give an example in the next section.) It can also enable one to discern common structures across different subject matters. For example, Ginzburg and Colyvan (2004) fruitfully emphasize the similarity of the equations of population growth to those of planetary motion (they argue that these systems are governed by similar second-order differential equations).

However, one must be careful not to read too much off a given formalism. It may resemble some target in certain important respects, but it must differ from the target in other important respects. (Compare how a map of a city differs from the city itself in all sorts of ways—it had better do so in order to be of any use!) And one should resist turning formalism into a fetish—as it might be, representing some philosophical problem with triple integrals and tensors, just because one can.

I can’t engage in any such autobiographical reflections without acknowledging the huge intellectual influence of David Lewis on my own research. I found his use of formal methods to be exemplary. I was especially drawn to his work on counterfactuals, on causation, and of course on probability and decision theory. He used such methods sparingly and judiciously, always to illuminate and to make insights easier to come by and to understand. His work serves as a model to me.

What example(s) from your own work illustrates the role formals methods can play in philosophy?

My early philosophical work was on probabilities of conditionals. Conditionals are notoriously recalcitrant beasts, having defied adequate analysis for over two thousand years. I liked Adams’s and (independently) Stalnaker’s idea of looking to probability theory, and in particular its familiar notion of *conditional* probability, for inspiration. They both advanced versions of the thesis that *probabilities of conditionals are conditional probabilities*. Then along came Lewis’s triviality results, which began an industry of showing that various precisifications of the thesis entailed triviality of the probability functions. I liked this industry *a lot*, and I joined in, proving some further triviality results.

I became fascinated with conditional probability in the process. I came to have misgivings about the traditional formula for conditional probability as a ratio of unconditional probabilities:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{(provided } P(B) > 0\text{).}$$

It was well known that this *ratio analysis* runs aground when *P(B)* = 0 (that parenthetical proviso is there for good reason!), yet probability theory admits probability-zero propositions as possible and non-trivial. Worse, the ratio analysis as it stands fails when *P(A ∩ B)* or *P(B)* are *vague* or *undefined*—yet such cases abound, as I argue at length in my (2003b). To be sure, Kolmogorov refined the ratio analysis to handle the *P(B)* = 0 problem, although his refinement still falters on the other problems. And it faces some further problems of its own, as Seidenfeld, Schervish and Kadane (2001) have shown.

I have also utilized formal methods in my work on Pascal's Wager. I have a tendency to be drawn to big things—the Himalayas, the Empire State Building, the Grand Canyon, Elvis's Jungle Room—and according to Pascal, the utility of salvation is as big as can be: *infinite*. His Wager is arguably the most famous argument for theism. This is a perfect example of how formal methods can be brought to bear on a philosophical problem. Think of the problem of whether or not to believe in God as a *decision problem*, and thus one to be solved by decision theory. Put in modern parlance, Pascal assumes that non-belief and God's non-existence both lead to finite utilities, and that you should assign positive probability to God's existence; he concludes that you maximize expected utility by believing in God. Given his assumptions, Pascal was right: the infinite utility of salvation swamps all other terms in the expectation calculation. But as I have argued in my (2003a), it does not follow that you should believe in God. For given Pascal's assumptions, you *also* maximize expected utility in infinitely many other ways. Any mixed strategy between belief and non-belief scores just as well, by Pascalian lights: the infinite utility of salvation *still* swamps all other terms in the expectation calculation, so these mixed strategies all get infinite expectation too. Moreover, arguably *anything* you could do should be regarded as such a mixed strategy, for there is presumably *some* probability that you will wind up believing in God as a result. Now translating back from the formalism to the real world decision problem: far from establishing that you should believe in God, Pascal's premises apparently have the consequence that *everything* you could do is equally rational. You might as well have a beer—or not. I went on to suggest various ways in which the utility of salvation could be understood so as to render Pascal's argument valid.

The St. Petersburg paradox involves another decision problem that apparently has infinite expectation. A fair coin is tossed repeatedly until it lands heads for the first time, and you get exponentially increasing rewards according to how long it takes (starting at, say, $2 if it lands heads immediately). The paradoxical conclusion is usually taken to be that decision theory judges the game to be infinitely good, whereas intuition recoils at this judgment. But I think a more disturbing conclusion is that, once again, as long as you give *any* credence to the possibility of playing the St. Petersburg game, all of your possible actions get infinite expectation. If decision theory is your guide to action, then like Buridan's ass you are paralyzed.

Expected utilities are *sums*; the St. Petersburg game exploits a certain kind of anomalous infinite sum, one which diverges. But we know from real analysis that another kind of anomalous infinite sum is one that is *conditionally convergent*—if we leave it alone, it converges, but if we replace all of its terms by their absolute value, the resulting series diverges. Riemann’s rearrangement theorem tells us that every conditionally convergent series can be reordered so as to sum to any real number; and it can be reordered so as to diverge to infinity and to negative infinity; and it can be reordered so as to simply diverge.

Now let this piece of mathematics guide the creation of a new game, whose expectation series has exactly this property—the formal model thus inspires a new kind of anomaly for rational decision-making. Harris Nover and I (2004, 2006) proposed a St. Petersburg-like game—the *Pasadena game*—in which the pay-offs alternate between rewards and punishment, in such a way that the resulting expectation is conditionally convergent. Decision theory apparently tells us that the desirability of the game is *undefined*, thus falling silent as to its desirability. Worse, the theory falls silent about the desirability of *everything*, as long as you give any credence whatsoever to your playing the Pasadena game—for any mixture of *undefined* and any other quantity is itself undefined. In that case, for example, you can’t rationally choose between pizza and Chinese for dinner, since both have undefined expectation (each being ‘poisoned’ by a positive probability, however tiny, of a subsequent Pasadena game). Thus, once more you are paralyzed—a sin against practical rationality. Yet assigning probability 0, as opposed to extremely tiny positive probability, to the Pasadena game seems excessively dogmatic—a sin against theoretical rationality.

Another body of work in which I appealed to formal methods concerned the so-called *desire-as-belief* thesis. David Lewis canvassed a certain anti-Humean proposal for how desire-like states are reducible to belief-like states: roughly, the desirability of *X* is the probability that *X* is good. He then proved certain triviality results that seemed to refute the proposal. This was another lovely example of how formal methods could serve philosophical ends: this time, a thesis that was born in an informal debate in moral psychology could apparently be expressed decision-theoretically. The decision-theoretic machinery could then be deployed to deliver a formal verdict, which could then be translated back to bear on the informal debate.

I noticed that the probabilities-of-conditionals-are-conditional-probabilities thesis of Adams and Stalnaker looked suspiciously like the desire-as-belief thesis, and that Lewis's triviality results against the former looked suspiciously like his triviality results against the latter. This gave me the idea that the subsequent moves and countermoves that were made in the probabilities-of-conditionals debate could be mimicked in the desire-as-belief debate. I showed that, much as Lewis's original triviality results could be blocked by making the conditional *indexical* in a certain sense, his later triviality results could be blocked by making 'good' indexical in the same sense. Philip Pettit and I (2003) then translated back out of the formalism, suggesting meta-ethical theories that accorded 'goodness' the necessary indexicality. So the trick was to notice a similarity between the formal structures that underpinned the probabilities-of-conditionals and the desire-as-belief debates, something that I could not have noticed about the original debates themselves. After that, it was easy to see how the next stages of the desire-as-belief debate should play out, paralleling the way they did in the probabilities-of-conditionals debate. Two seemingly disparate debates turned out to be closely related.

So I have come full circle back to my original work on probabilities of conditionals. I am in the process of writing a book on the two debates, and their many structural similarities, in a book to be entitled *Arrows and Haloes: Probabilities, Conditionals, Desires, and Beliefs.*

What is the proper role of philosophy in relation to other disciplines?

An ongoing side-project of mine is to gather philosophical heuristics, analogues of ‘castle early and often’ in chess. A useful one is to see the word ‘the’ in neon lights. A locution of the form ‘… the X …’ typically carries with it the presupposition that there is exactly one X, and this presupposition may well be false. Suitably cautioned by the heuristic, I will observe that there are *many* proper roles of philosophy in relation to other disciplines. I would rather not try to order them in importance, or to demarcate their domains.

It’s something of a platitude that philosophy is concerned with foundations, although some platitudes are actually true, especially here in my adopted town of Canberra. So whereas a physicist asks about the chance that a given atom decays in some period of time, or a chemist asks about the chemical properties of some compound, or an astronomer asks what causes black holes to form, a philosophers asks “What is chance?”, “What is a property?”, “What is causation?” If we do our job well, our answers to these philosophical questions will accord with such canonical applications of these concepts in the sciences. That said, the boundaries between philosophy and other disciplines are somewhat permeable. Where, for instance, is the borderline between philosophical and mathematical logic?

Much of philosophy that I find interesting can be characterized by the slogan: “Making our implicit commitments explicit”. These include the commitments of commonsense, familiar to the folk, which almost inevitably infiltrate the sciences to some extent, and even more so the social sciences. They also include the commitments of scientific and social-scientific theories themselves. And philosophy plays a useful role as watchdog of other disciplines: questioning their presuppositions, policing their hasty inferences, clarifying their murky concepts. It teases out unintended and often unwelcome consequences of those presuppositions, provides tools for evaluating those inferences, and offers frameworks for understanding better those concepts. This is especially evident in the various ‘philosophy of __’s. It is somewhat contingent which disciplines get to fill in the blank. Philosophy of physics has long been a respectable field. Philosophy of biology is rather more recent, but it is currently thriving. Philosophy of chemistry is still at a nascent stage, but it is showing promise. Philosophy of geology, of meteorology, of astronomy, and of other ‘special sciences’ are yet to arrive on the scene. Perhaps they are not fundamental enough (as are physics and arguably chemistry), and perhaps they do not raise problems of a distinctive enough kind (as does biology) to merit sustained philosophical attention.

Another significant role for philosophy vis-à-vis other disciplines is to address *prescriptive* questions, where they typically address *descriptive* questions. This distinction is often blurred by the near-homophony of the words ‘idealization’ and ‘ideal’. Indeed, sometimes the words get conflated, as when chemists speak of the ‘ideal gas law’, suggesting a law about maximally virtuous gases, when really it is an *idealized* gas-law. Physics, for example, is up to its neck in idealization, and so is decision theory. But whereas decision theory attempts to codify *norms* and *evaluates* actions that meet them or not, physics just codifies and unifies regularities without approval or sanction. Decision theory exhorts the *ideal* of maximizing expected utility theory, and criticizes us when we fall short of that ideal. But physics never tells an electron that it is irrational, nor a galaxy that it is badly behaved.

Many good philosophers are like intellectual decathletes, knowing a fair bit of mathematics, science, and social science, with better-than-average writing skills. And every serious discipline has its share of philosophical problems. There is thus much opportunity for cross-fertilization: the other disciplines can offer material for philosophers to sink their teeth into, and the philosophers can offer in return rigorous scrutiny of the disciplines' foundational issues.

What do you consider the most neglected topics and/or contributions in late 20th century philosophy

Some topics are properly neglected, and some are paid undue attention. But an improperly neglected field is surely the *philosophy of statistics.* As an undergraduate I was schooled exclusively in the tradition of classical statistics, à la Fisher and Neyman-Pearson. It wasn't until I became a philosopher that I even heard of Bayesian statistics, and it's still an area that has been neglected by philosophers. To be sure, philosophical issues in statistical inference and estimation have been addressed in important work by philosophers such as Hacking and Kyburg, and statisticians such as Dawid and Lindley—but again, these have not reached the philosophical mainstream.

My statistics professors taught me about type 1 and type 2 errors in hypothesis testing, and that the former were somehow worse than the latter. But nobody ever recommended the obvious way to avoid the former: set the significance level of a hypothesis test to 0! So I gathered that the former were clearly not infinitely worse than the latter. That left me wondering exactly what the terms of trade were between the two kinds of error. Here statistics meets *decision theory*, yet a course on it was unavailable in my four-year statistics degree. I was also taught that rational inference and decision-making look rather different in a context of strategic interaction among multiple players. Here, presumably, statistics meets *game theory*, although I never heard a word about their union. And I was taught that "correlation is not causation". A catchy slogan, to be sure, and a sound warning—but nothing more was said about causation. Here statistics meets *causal modeling*. I thus see the interfaces between statistics and these more philosophical preoccupations as primary areas of neglect—certainly in my own education, and I surmise in the academy more generally. Cutting edge work in these areas, in turn, should inform debates in some other areas of philosophy. For example, philosophers of mind working on mental causation could profit from keeping up to date with the causal modeling literature.

Many of the debates in philosophy of mind and in ethics are couched in purely deductivist terms. They concern questions such as these: Are mental facts identical to, or reducible to, or supervenient on non-mental facts? Are moral facts identical to, or reducible to, or supervenient on non-moral facts? While Hume denied the intelligibility of necessary connections between distinct existences, philosophers in these areas sometimes talk as if *only* necessary connections between seemingly distinct existences could be of philosophical interest: identity, reducibility, supervenience. Similarly, while epistemology has embraced probabilistic methods, the debates in the philosophy of mind and moral psychology are still mainly conducted using the all-or-nothing categories of ‘belief’ and ‘desire’. Looking on as an outsider to these debates (and confessing the outsider’s ignorance), I fear that their protagonists are setting themselves up for failure: they risk stalemating. To be sure, I will cheer as loudly as anyone if their deductivist, all-or-nothingist aspirations are realized. But in the meantime, I think that the debates could use a healthy injection of *probability*.

The concept of probability was a relative latecomer on the intellectual scene—it is entirely absent from the works of the ancient Greeks and medievals. One wonders how such clever chaps could get by without it! But even in the twentieth century we see major philosophical work on probability-laden notions being done without any aid from probability theory—think of Hempel on confirmation theory, or Popper on scientific methodology. We can safely say with the benefit of hindsight that trying to force the round peg of inductivism into the square hole of deductivism is like trying to square the circle. I wonder whether future philosophers will look back on these current debates in ethics and philosophy of mind, conducted as they are entirely in deductivist terms, in a similar way.

Perhaps, then, we would do well to break the philosophy of mind and ethics free of the straitjacket of entailment relations, and seek *probabilistic relations* between physical facts and mental facts, or between physical facts and moral facts. Or said without any pessimism: perhaps we should seek such probabilistic relations first, before shooting for full-blooded entailments. Still more ecumenically, we could productively run both research programs side-by-side. Again, think of how traditional epistemology, trafficking as it does in all-or-nothing concepts such as knowledge and belief, happily co-exists with probabilistic epistemology.[1] Indeed, probabilism offers a way of side-stepping some of traditional epistemology's concerns with skepticism, and of resolving some hoary paradoxes, the lottery and the preface paradoxes among them.[2]

So this is a call to arms to philosophers of mind and ethicists: to arm themselves with probabilistic methods when tackling their traditional problems. Just securing comparative probabilistic supervenience relations would be a good start: "physical basis U makes it more probable that there is a mind than physical basis V does";

[1] Let me put in a plug here for the annual Formal Epistemology Workshop, created by Branden Fitelson and Sahotra Sarkar, which has done much to bring together epistemologists from both traditions. See:
http://socrates.berkeley.edu/~fitelson/few/

[2] That said, we need to understand better the interrelations between the probabilistic and the non-probabilistic concepts. It seems that 'knowledge', 'belief', and 'evidence' defy purely Bayesian analysis. Rather, the Bayesian eschews the first two notions, and takes the third for granted.

"physical basis X makes it more likely that there are moral facts than physical basis Y does", and so on. We could then try to work our way up from there.

Related, philosophers often regard *soundness* of arguments as the only criterion of success, when we ought to know better. Soundness, after all, is neither necessary nor sufficient for an argument being compelling or of value. (Now *there's* something that can be said in purely deductivist terms!) Think, for example, of the time-honored philosophical strategy of *parodying* an argument that one doesn't like, as Gaunillo did to St. Anselm's ontological argument, or as Diderot did to Pascal's Wager. We start with the target argument, show that it is 'just like' another argument with an obviously silly conclusion, and which is thus obviously unsound, and conclude that the target argument must likewise be unsound. Setting aside the manifest unsoundness of *this* line of reasoning—tu quoque!—it mistakenly regards soundness as the touchstone of argument assessment. And it surely proves too much. The sensible argument

<u>All emeralds so far observed have been green</u>

∴ The next emerald will be green

is 'just like' the silly argument

<u>All emeralds so far observed have been grue</u>

∴ The next emerald will be grue

where we define 'grue' so as to make the conclusion obviously false. And indeed, both arguments are unsound, just as the Proves-too-mucher would have it. But again, what is needed is a healthy injection of probability—in this case, a confirmation theory that can account for the obvious difference in inductive strength between the two arguments.

To be sure, the proves-too-much strategy is often a useful heuristic, and it is even more often dialectically and rhetorically effective. I am not above using it myself—I did so, for instance, in my paper “Scotching Dutch Books?”. But while we’re at it, notice that the strategy seems to regard the parody argument, as it were, as a *template* for a kind of bad argument, and the strategy regards any argument that fits the template well enough as automatically bad. Here I am reminded of Massey’s important insight in a number of papers (e.g. Massey 1975) that, trivial cases aside, there is no such thing as invalid argument *form*, the way that there is such a thing as valid argument form—no schema or template for arguments, any instance of which is invalid. These papers by Massey are high on my list of underappreciated contributions to philosophy.

Looking elsewhere, *information theory* is apparently flourishing—witness its popularity in cognitive science, computer science, and engineering. Yet the *philosophy of information theory* is so unflourishing as to be virtually non-existent. And I suspect that information theory, and specifically minimum message length theory, may offer just the right tools to address some old philosophical chestnuts—for example, quantifying the extent to which true scientific theories balance simplicity and strength, and thus the extent to which they can be regarded as codifying the *laws of nature* on the Mill-Ramsey-Lewis analysis (Dowe and Hájek 1997). This is surely fertile, and under-harvested, philosophical ground.

Computer simulations bring with them another set of philosophical problems that have not received sufficient attention. What is the epistemological significance of ‘experiments’ that are performed on one’s laptop? How do simulations relate to thought experiments? To what extent should applied ethics pay attention to simulations—for example, to computer models of global warming or of overpopulation? And closer to my home turf, to what extent are Monte Carlo methods faithful to the probabilistic models that inspire them?

Finally, there are hard questions in meta-philosophy that do not receive sufficient attention. Conceptual analysis is thought by many to have passed its use-by date, although it is regarded as alive and well in my adopted town. When should a concept be taken to be primitive, and when is it an appropriate target of analysis? More generally: what should philosophers be doing, and how should we be doing it? Such topics are occasionally addressed, but usually in a somewhat piecemeal fashion—an article here or there on this or that topic in meta-philosophy. Bravo to Vincent Hendricks and John Symons for stimulating profitable and sustained dialogue addressing such methodological questions in their volumes.

What are the most important open problems in philosophy and what are the prospects for progress?

This very nearly reduces to the question: "What are the most important problems in philosophy …?"—for very few important problems in philosophy are *closed.* Gödel's proof of the incompleteness of arithmetic is taken to be a canonical example, although it is arguably more a piece of mathematical logic than of philosophy. I take philosophy to have definitively settled various *negative* results: almost any significant piece of conceptual analysis has eventually met with decisive counterexamples. It's *positive* philosophy that's really difficult.

I won't pretend to give an exhaustive list of *the* most important open problems (that word again!). But I'll mention some obvious ones and then say rather more about some of the problems that grip *me* the most.

First, some obvious ones: the mind-body problem; the nature of consciousness, and of free will; the rationalist/empiricist debate over innate ideas; the existence of God; Plato's problem of the one and the many; the existence of abstract objects; the status of moral facts; the status of modal facts; the meaning of 'meaning'; the content of 'content'; a true analysis of 'truth'; the problems of induction ("old" and "new"); a proper understanding of a 'just' society; the nature of time, and its relationship to space; and (dare I say it), how we are to live good lives. Take a look at the curriculum of your typical undergraduate philosophy program, and you'll see a menu of such topics.

Much as the holy grail for physicists is a grand unified theory of physics, the holy grail for me is a *grand unified theory of rationality.* This would be nothing less than a fully integrated decision theory and confirmation theory that incorporates the insights of statistics, game theory, and causal modeling, and that explicates rational judgment, preference, and action at both the individual and group level.

Let's start with individual rationality. I imprinted on Bayesianism in my philosophical infancy, but like a rebellious child, there is much that I now question in it. Its central notion is that of *degree of belief,* or *credence,* but I find the main analyses of this notion to be inadequate (the betting interpretation and other forms of operationalism, interpretivist accounts that appeal to a decision-theoretic representation theorem, and so on). I find equally inadequate the main lines of defence of Bayesianism: the Dutch Book argument, the representation theorem argument, calibration, and so on. So, much as I admire the theory for its elegance and for its fruits—its explanatory power, its ability to illuminate various old problems in confirmation theory, and its unification of confirmation and decision theory—I believe that its foundations could use some shoring up.

Then there are questions about the very statement of Bayesianism. It tells us that credences should conform to probability theory—but what does that mean exactly? Should credences be defined on a sigma field, or does a field suffice? Should they be countably additive, or does finite additivity suffice? How exactly should the 'normalization' axiom be formulated? If formulated sententially, it takes the form 'all tautologies receive probability 1'—but tautologies of which logic? If classical logic, do we mean propositional logic, predicate logic, predicate logic with identity, or more? Are non-classical logics permitted? If not, why not, and if so, which ones? Should conditional probability be defined in the usual way as a ratio of unconditional probabilities, or as I prefer, as a primitive? How do we extend our theory of conditional probability to uncountable probability spaces? And so it goes.

All of this presupposes that we know what the arguments of probability functions are in the first place, but here again there is work to be done. For example, a currently burgeoning sub-field of research, prompted by problems of self-location such as the 'Sleeping Beauty' problem, concerns whether the contents of probability assignments should be *centered* propositions, and if so, what the 'centers' should be. Again, there are foundational issues aplenty.

Now take the notion of 'credences' as given, and let's agree that they should conform to the probability calculus (whatever that means exactly), just as the Bayesian says. Demanding though this norm is—*you* try assigning probability 1 to all tautologies!—in other respects it is far too permissive. In particular, we need further reasonable constraints on priors. I have misgivings about the details of some of the main proposals, although I am sympathetic to their spirit. They include:

- *Regularity*, an unevocatively named 'open-mindedness' constraint to assign probability 1 only to tautologies. Since regularity is the converse of the normalization axiom, stating the former exactly is problematic in the same ways as the latter is.

- The *Principal Principle*, the more evocatively named constraint that, roughly, one's credence for a proposition should equal one's subjective expectation of its objective chance. Yet we lack an adequate account of objective chance in its own right.
- The *Reflection Principle*, the equally evocatively named constraint that, roughly, one's credence for a proposition should equal one's subjective expectation of one's *future* credences of it. Yet there are apparently various cases in which this principle fails—e.g., those involving memory loss, or even merely epistemically possible memory loss, and indexical beliefs.
- The *principle of indifference*, and more generally, the *principle of maximum entropy*, which roughly say that one should assign the 'flattest', least informative probabilities that one can, consistent with one's evidence. A major worry is the apparent partition-dependence of applications of these principles.

Here the subjectivist interpretation of probability begins shading off into the logical interpretation. I regard getting right the details of these principles, and of additional such principles, as a high priority. After all, whether we like it or not, our epistemic practices betray our commitment to a quasi-logical notion of probability. We think that it would be irrational to deny that the sun will rise tomorrow, to project 'grue' rather than 'green' in our inductions, and to commit the gambler's *fallacy*. Understanding this irrationality takes us to the very heart of confirmation theory.

Subjective probability plays a key role both in confirmation theory and in decision theory. And we find yet more turmoil in the foundations of decision theory. Firstly, we need to sort out the inter-mural disputes between various rival decision theories—

in particular between so-called ‘evidential’ decision theory and the variants of causal decision theory (which themselves seem far from equivalent to me). Then there are paradoxes to be resolved. High among them for me are the paradoxes associated with the St. Petersburg game and the Pasadena game.

We need to understand better the relationship between decision theory and game theory. Decision theorists tell us that rational action is always, everywhere, and without exception a matter of maximizing (their preferred version of) expected utility. Yet game theorists talk instead of Nash equilibria, sub-game perfect equilibria, proper equilibria, trembling hand equilibria, and so on in contexts of strategic interaction with others agents (not to mention their distinction between games in normal and extensive form). To what extent are these just different ways of saying what the decision theorists say, and to what extent do our theories of rationality bifurcate depending on whether we are playing games against nature, or against other agents? (And where do we draw the line for what count as other agents? When I am interacting with my dog Tilly, am I playing a game against nature, or against another agent? It often seems like a bit of both.)

Matters are complicated further when we take into account the *moral* aspects of our decision-making—as we often must—when we are uncertain about which meta-ethical theory is correct. Kantianism says that Jim must not shoot one Indian to stop Pedro from shooting twenty, while utilitarianism says Jim must. Suppose Jim is 50% confident of the truth of each theory. What should he do? And how do we incorporate the deliverances of such theories into decision theory? For example, if a Kantian categorical imperative can be mapped onto a utility scale at all, it would appear to correspond to a negative infinite utility—and we are saddled again with St.

Petersburg-style paradoxes, and corresponding paralysis. Andy Egan and I (MS) are exploring such issues further.

These problems are only writ larger when it comes to aggregating the opinions and preferences of multiple agents, let alone an entire society. Important relationships between rational individual and cooperative group choice are being explored by authors such as Isaac Levi, Christian List, Robert Nau, Philip Pettit, Brian Skyrms, and the Carnegie Mellon trio of Joseph Kadane, Mark Schervish, and Teddy Seidenfeld. But again, there is much further work to be done.

And if all that could eventually give us a modicum of guidance on how we are to live good lives, all the better—although I'm not holding my breath.[3]

Philosophy Program
Research School of the Social Sciences
Australian National University
Canberra, ACT 0200
Australia

REFERENCES

Dowe, David and Alan Hájek (1997): "A Computational Extension to the Turing Test", *Proceedings of the 4th Conference of the Australasian Cognitive Science Society,* Newcastle, NSW, Australia.

Egan, Andy and Alan Hájek (MS): "Moral Uncertainty".

[3] I thank especially Carrie Jenkins, Aidan Lyon, and Daniel Nolan for helpful discussion.

Ginzburg, Lev R. and Mark Colyvan (2004): *Ecological Orbits: How Planets Move and Populations Grow*, New York: Oxford University Press.

Hájek, Alan (2003a): “Waging War on Pascal’s Wager”, *Philosophical Review*, Vol. 113 (January), 27–56. Reprinted in *The Philosopher’s Annual* 2004, ed. Patrick Grim, www.philosophersannual.org.

Hájek, Alan (2003b): “What Conditional Probability Could Not Be”, *Synthese*, Vol. 137, No. 3 (December), 273-323.

Hájek, Alan and Philip Pettit (2004): "Desire Beyond Belief", *Australasian Journal of Philosophy*, Vol. 82 (March), 77-92 (special issue dedicated to the work of David Lewis). Reprinted in *Lewisian Themes: the Philosophy of David K. Lewis*, eds. Frank Jackson and Graham Priest, Oxford University Press, 2004, 78-93.

Hájek, Alan and Harris Nover (2006): “Perplexing Expectations”, *Mind* 115 (July), 703-720.

Massey, Gerald (1975): "Are There Any Good Arguments that Bad Arguments are Bad?", *Philosophy in Context* 4, 61-77.

Nover, Harris and Alan Hájek (2004): “Vexing Expectations”, *Mind*, Vol. 113 (April) 237-249.

Seidenfeld, Teddy, Mark J. Schervish, and Joseph B. Kadane (2001): “Improper Regular Conditional Distributions”, *The Annals of Probability,* Vol. 29, No. 4 (October), 1612-1624.