Philosophical Heuristics and Philosophical Methodology

Alan Hájek[1]

Introduction

Metaphilosophy is all the rage nowadays. Philosophers are becoming increasingly self-conscious about their methodology, as this volume showcases. And rightly so—it is part of our job description to put thinking under the microscope, and that obviously should include our *philosophical* thinking. Much as we want philosophers of physics, of biology, of mathematics, and so on to scrutinize those respective disciplines, so we want philosophers of philosophy to scrutinize *ours*.

When I look at the ways that much philosophy gets done, I see certain recurring patterns of thought. The best philosophers repeatedly deploy a wealth of philosophical techniques. I see them used in papers, books, talks, Q & A sessions, and philosophical discussion more broadly. Since I was a graduate student I have observed them, and when they have seemed especially fecund to me, I have recorded them on an ongoing list and I have tried to internalize them. After all, they are part of the philosopher's intellectual armoury, much as logic is. We may call them *philosophical heuristics*.

But unlike logic, they are not studied or taught. We just *use* them, but we typically do not do so self-consciously; indeed, we may often use them unconsciously, unaware that we are even doing so. Despite the importance of such heuristics to philosophical methodology, they have been surprisingly neglected by philosophers.

[1] Thanks especially to John Barker, Paul Bartha, Bert Baumgaertner, Rachael Briggs, Mark Colyvan, Patrick Greenough, Robbie Hirsch, Ben Levinstein, Bernard Molyneux, Daniel Nolan, Wolfgang Schwarz, Jonathan Simon, Nick Smith, … [this list will grow] for helpful discussion and comments.

(Nozick 1993 is a notable exception, but even his discussion of some of them is confined to just over 8 pages.)

Practitioners of various other disciplines have not been so remiss regarding their own heuristics. Mathematicians, for example, often teach their students heuristics for approaching problems, and there are numerous books (e.g. Polya's 1957 classic *How to Solve It)* detailing them. Here's one that I learned from a mathematics professor: if you are struggling to prove something because it seems obvious, try reductio ad absurdum. Or again, music professors teach their students various "rules" of harmony and counterpoint, which really are rules of thumb: avoid consecutive fifths and octaves, and what have you. No mathematician or musician pretends that once you have mastered such heuristics, you will be the next Gauss or Beethoven; but it would be absurd to question their utility as strategies for guiding one's thinking or creativity.

Almost any complex activity that requires skill has its heuristics—cooking, rock-climbing, photography, gymnastics, playing the xylophone, salsa dancing, taxi driving, taxidermy (I assume)... Indeed, it would be odd if philosophy, one of the most complex activities of which we are capable, *didn't* have heuristics. And it obviously does—hundreds of them that I've identified so far. Some of them are *exclusively* philosophical heuristics, while others are more general-purpose heuristics for conducting one's thoughts that one finds employed elsewhere. And still others are not so much "heuristics" in a narrow sense, but fruitful patterns of thinking. Despite this chapter's title, I don't want to put too much weight on either the word "philosophical" or "heuristics"; as long as they are philosophically fertile strategies, I am happy to count them for my purposes, however one might best label them.

It would be strange counsel to remain silent about them: [whispered] "Yes, they're out there (unfortunately), but don't tell anyone else about them!" And yet I have encountered some critics who view my project with the sort of suspicion visited upon a magician who breaks ranks with his fellow magicians by revealing some of the secrets of their trade. Well, I'll let you in on a few of the secrets of *our* trade. But then, I bet that you tacitly know some or all of them anyway; I'm just making them explicit.

I have already presented a first batch of heuristics in my (forthcoming). There I discussed such techniques as:

- *Check extreme cases, and near-extreme cases.* These are especially promising places to look first for counterexamples, and doing so reduces your search space.
- *Death by diagonalization: reflexivity/self-reference.* Plug into a function itself as its own argument, and more generally, appeal to self-referential cases. It worked for Cantor, Russell, and Gödel; it may work for you.
- *Self-undermining views.* Relatedly, check whether a philosophical position or argument that falls in the very domain that it purports to cover treats itself consistently.
- *Begetting new arguments out of old.* Good arguments are often easily transformed from one domain to another. Exploit analogies between, e.g., space, time, and modality; and between rationality and morality.
- *Trial and error.* When ingenuity fails you, sometimes you can just run systematically through the relevant cases until you find one that meets your needs.

I stressed that they are not sure-fire recipes to philosophising—there are none—but rather defeasible guides that tend to work.

Here comes a new batch. I will focus on a cluster of heuristics that have some interesting family resemblances. Indeed, *resemblance* is a theme that unites several of them. For instance, there may be some role that we want played—by a concept, or a choice, or a solution to a problem—but there are multiple candidates that are very similar in relevant respects, and as a result appear to play the role equally well. This idea soon fans out to encompass a number of heuristics related to handling arbitrariness, continuity reasoning, the analysis of concepts that come in degrees, and so on, but it is a good starting point.

Away we go.

1.a See definite descriptions in neon lights

A philosophical thesis or an analysis that involves a definite description '... *the* F...' typically presupposes that there is *exactly one* F. Ask:

i) Are there, or could there be, *multiple* F's?

ii) Are there, or could there be, *no* F's?

In short, check to see if there are *any* or *many* F's. The hard task of looking for problem cases has been broken down into two easier sub-tasks. i) and ii) correspond to tests for *uniqueness* and *existence*, respectively, familiar in mathematics.

Examples:

Counterfactuals

Stalnaker (1968) analyses counterfactuals[2] along the following lines:

'if it were the case that X, it would be the case that Y' is true

iff

Y is true at THE nearest X-world'.

[2] In fact, his analysis is of conditionals in general, but Lewis's reply, to which I turn, concerns counterfactuals more specifically.

(Since I do not have a neon-lights font at my disposal, the capital letters will have to do.) The "THE" indicates the assumption that for any X, there is a *unique* nearest X-world. Lewis (1973a) has two main objections:

i) There may be *multiple* nearest X-worlds. 'If Bizet and Verdi were compatriots, then …' Would they both be Italian, or both be French? It seems that nothing favours one nationality over the other. So there are at least *two* nearest worlds in which they are compatriots. In general, Stalnaker's analysis appears to founder on cases in which there are ties for the nearest X-world***s***.

ii) There may be *no* nearest X-worlds. 'If I were taller than 7 feet, then …' How tall would I be? 7 feet 1 inch? 7 feet ½ inch? 7 feet ¼ inch? …' We apparently have an infinite sequence of ever-closer worlds in which I am taller than 7 feet, with none closest. In general, Stalnaker's analysis appears to founder on cases in which there are *no* nearest X-worlds.

The problem of evil

A well-known argument against various forms of theism goes like this:

> If an omnipotent, omniscient, omnibenevolent God existed, then He would have created THE best of all possible worlds. But the actual world is not the best of all possible worlds; therefore such a God does not exist.

i) Perhaps there are many worlds tied for first place in the relevant 'goodness' ordering. In that case the argument needs to be recast: God would have created (at least?) one of them, and the actual world is not one of them.

ii) Perhaps worlds get better and better without end. Imagine an infinite sequence of worlds in which there are successively more happy people, or more happy rabbits …, each world better than its predecessor.

Functionalism

Functionalist accounts of theoretical and folk-theoretical terms often invoke locutions such as ‘the X role’ (neon lights), or better still for my purposes, ‘the occupant of the X role’ (double neon lights). This presupposes that there is exactly one such role, and exactly one such occupant. Functionalists speak, for instance, of ‘the occupant of the pain role’ (which is ritually taken to be the firing of c-fibres). Lycan (2009) argues that the presupposition of a single occupant begs the question against the dualist view according to which pain behavior is causally overdetermined, being caused both by physical neural events and non-physical pain events. One might also question whether there is an occupant of the role at all; and whether there is a single pain role—e.g., perhaps different folk theories assign it different roles.

So far, the examples have been familiar. Now, let me use the heuristic to make a less familiar point—to do some new philosophical work.

Velocity as the time-derivative of position

Nearly every physics textbook analyses velocity as the time derivative of position:

$$v = dx/dt.$$

THE time derivative. Let’s subject this analysis to our heuristic.

i) I see no problem with there being multiple time-derivatives of position. To be sure, there may be many time variables, corresponding to many frames of reference; but in that case there are correspondingly many velocities, one for each time variable. The heuristic came up empty-handed here.

ii) I *do* see a problem with the time-derivative not existing. There are functions that are not differentiable everywhere; in fact, there are functions that differentiable *nowhere*; in fact, there are functions that are *continuous everywhere and differentiable nowhere*. Weierstrass discovered such pathological 'saw-tooth' functions. Now think of a particle whose position function of time is such a function. This particle moves highly erratically, darting this way, then that, in a wild manner. Yet it traces a continuous path, so it is implausible that it goes in and out of existence. In short, we apparently have an example of *motion without velocity*. This may not be an outright counterexample to the physicists' analysis of velocity, but it is surely odd. The heuristic helps to lead us to a puzzling case.

*   *   *   *   *   *

In response to the problem of multiple candidates for F in the definite description 'the F', one may settle instead for the *indefinite* description, '*a(n)* F'. Then any of the candidates should do; all that matters is their existence, not their uniqueness. But we may still be able to mount our assault from the other side: even *indefinite* descriptions are in trouble when their existence presuppositions fail. This brings me to a closely related heuristic.

1.b See indefinite descriptions in neon lights

A philosophical thesis, or an analysis, that involves an indefinite description '... *a(n)* F...' presupposes that there is *at least one* F. Ask whether there are, or could be, *no* F's. All of the cases above faced a challenge of this kind, so even replacing their definite descriptions with indefinite descriptions would not get them out of the woods.

Exposing a definite or indefinite description by paraphrase

Sometimes a problematic definite or indefinite description can come in disguise, and when it does, it is harder to see any neon lights, because one does not actually see the word ‘the’ or ‘a(n)’. Still, it may be lurking in the background, to be revealed in its neon glory by paraphrase.

Example. To *maximize* a quantity means to achieve THE greatest amount of some quantity. The paraphrase makes the definite description explicit. But we can also speak of a function that has multiple maxima, many points at which it achieves its maximum value. Think of y = cos x—while in a sense it has one maximum (namely, 1), in another sense it has many maxima (at $x = \pm 2n\pi$, $n = 0, 1, 2, \ldots$). In the latter sense, each of these points is A maximum, the indefinite description made explicit. The same can be said for *minimizing* a quantity, and *minima*, mutatis mutandis. But whether we speak of the single maximal/minimal value attained by a quantity, or the many ways in which it may be attained, one cannot maximize/minimize a quantity that has no maximum/minimum.

For instance, according to decision theory, rationality requires you to *maximize* expected utility. It does not seem problematic for decision theory if there are multiple ways to do so in a given situation. Consider Buridan’s Ass, who can maximize expected utility by eating either of two equidistant hay bails; either way, it will achieve THE maximum amount of expected utility. On the other hand, cases in which there is an infinite sequence of actions of ever-greater expected utility are problematic. Pollock’s (1983) Ever-better wine provides an example: the longer you wait to open the bottle of wine, the better it gets. When should you open it? We can specify the case so that any time is too soon; yet never opening it is the worst option of all. There is no sense in which you can maximize expected utility here.

Similarly, a naïve version of consequentialism says that one is morally required to perform an action that has maximally good consequences (in some sense or other); but one cannot if there are actions that yield better and better consequences without end. Moving to a higher level of abstraction, some functions have no maximum at all—those that increase towards an asymptote without ever reaching it, or that have no bound, or whose range is an open set.

Indeterminacy

So far, we have looked at cases where 'the F' does not have a unique referent, because there is determinately more than one F, or determinately none; and cases where 'a(n) F' does not have a referent, because there are determinately none. A different kind of problem arises when it is *indeterminate* how many F's there are, and in particular, it is indeterminate whether there is *one*, or indeterminate whether there is *at least one*. Similarly, the word 'the' suggests that there is a determinate answer to a relevant question, or a fact of the matter of the bearer of some relevant property. But there might be indeterminacy regarding these things.

Example. Philosophers of biology ask questions such as 'What is *the* function of the frog's eye?' And they consider answers such as 'to detect flies', 'to detect dark spots', and 'to detect food'. But perhaps there is no determinate answer to the question—no determinate function of the frog's eye. Perhaps it is even indeterminate whether there is any function of it at all.

Example. What is *the* right thing to do when one faces a moral dilemma? Parfit (1984) argues that often it is indeterminate. It may not even be determinate that there is *a* right thing to do.

We will encounter indeterminacy again in our next heuristic, which has a number of points of contact with the previous ones.

2. Arbitrariness, and how to respond when faced with it

When a philosophical position has the form '… the F …', or '… a(n) F …' and there are multiple F's, it may be permissible just to pick one of them. For example, if there are many best of all possible worlds, maybe God can create any one of them. But perhaps choosing one F out of the many candidates will be *arbitrary* in an unacceptable way. God's creating one out of multiple worlds tied for first place in the goodness ordering would violate the principle of sufficient reason. For that reason, when Leibniz thought that God created *the* best of all possible worlds, he really meant *the*, not *a*!

Onwards to the next heuristic, or better, set of heuristics. They begin with problems arising out of arbitrariness, and then offer a number of ways of responding to these problems.

The sorites paradox furnishes a classic example:

> A grain of sand is not a heap; two grains of sand is not a heap; for all n, if n grains of sand is not a heap, then neither is n+1 grains; therefore there are no heaps.

Sure, you can arbitrarily stipulate that there is a particular grain of sand at which a heap suddenly comes into being—say, the 17th. But that's exactly the problem: it's arbitrary. Why that choice, rather than the 16th, or 18th, or other nearby choices?

Arbitrary choices are familiar from daily life. Where should we set speed limits, or the voting age, or the drinking age? A sign that these choices are arbitrary is that different societies set them differently, and sometimes a given society will revise its settings over time. But we are typically not troubled by their arbitrariness, for it is *not* arbitrary that *some* setting be made, as opposed to none.

On the other hand, arbitrary choices are often regarded as fatal to philosophical positions that have to make them. The classical interpretation of probability, with its notorious principle of indifference, and Carnap's (1952) logical interpretation of probability, have been widely thought to be killed, or at least seriously wounded, by the arbitrary choices that they are forced to make—the suitable partition of 'equipossible' events in the former case, or the setting of an index of inductive caution, λ, in the latter.

More broadly, arbitrariness can be a sign of a flaw in a philosophical position. It is forced to make a choice; but why that choice, when it apparently could just as easily have made another one? The choice would seem to have no force—normative, or semantic, or otherwise. How could it be binding—rationally, or morally, or semantically, or what have you? And it's unlikely that it is lining up with a joint of nature, or of metaphysics, or of semantics, or what have you.

A good case study, which will showcase several techniques, involves Lewis's (1973a) analysis of *laws of nature*. Start with all of the true theories of the world. Some are very simple, but not informative—e.g. the theory whose sole axiom is that everything is self-identical. Some are very informative, but not simple—e.g. the collective (true) contents of *Wikipedia*. Some achieve a better balance of simplicity and informativeness than others. According to Lewis, the laws of nature are the theorems of the true theory of the world that best balances simplicity and informativeness—for short, *the* best system. But he acknowledges the possibility that there may be more than one reasonable way to trade off simplicity against informativeness. Different standards for balancing simplicity and informativeness may yield different theories as the winner of the Lewisian competition. What, then, are the laws? We could simply choose one set of standards, and insist that *it* dictates what the laws are. But why that set, rather than another set? This choice threatens to be arbitrary.

More generally, a problem arises for a philosophical position when there are multiple candidates for some job description appealed to by that position, all apparently equally good, and choosing any one of them over the others seems arbitrary.

But there are various possible responses. I will classify them into three kinds:

1) Symmetry-breaking responses (playing favourites): all candidates are equal, but some are more equal than others.

2) Symmetry-preserving responses (even-handed): all candidates really are equal, but we can deal with that.

3) Hybrid responses: first some symmetry-breaking; then, symmetry-preserving among the candidates that remain.

<u>1) Symmetry-breaking responses</u>

1)i. *One of the candidates is salient, or privileged*

The first response is to insist that one of the candidates stands out after all, so choosing it over the others is not arbitrary after all. I do not have an account of what makes a candidate salient or privileged, but I do have some rules of thumb. Extremal cases are usually salient. So too are points of symmetry, and they can provide symmetry-breakers! In doing so, they can answer charges of arbitrariness. One candidate stands out: the symmetrically placed one. Consider our dividing a pie that you and I both want—if we don't agree on the 50/50 split, what could we agree on?

Or one candidate may be privileged in virtue of being more fundamental than its rivals—e.g. the latter being reducible to the former, or supervening on the former, but not vice versa. Then it may be appropriate to thump the table in favour of that candidate.

Lewis imagines such a response for his account of lawhood: there are privileged standards for balancing simplicity and informativeness. He does not say more about what privileges such standards, but we could begin to flesh out his idea—e.g. by considering the shortest descriptions of all that is true of the world, couched in some canonical language in which all predicates correspond to natural properties, and all names correspond to a natural division of entities. Some authors have similarly attempted to rehabilitate the principle of indifference by regarding certain partitions to be privileged—e.g. those that are maximally fine grained (Elga's 2004 "predicaments"), or that are invariant under certain fundamental transformations (Jaynes 2003).

1)ii. *Go subjectivist/pragmatist*

The next response runs: "*You* get to choose the candidate that you want—it's *your* interests that you want to serve!" This response may be plausible when the job description involves *subjects*, for example rational agents. Subjective Bayesians about probability, for example, are often untroubled by the seeming arbitrariness in the choice of priors. But the response is surely implausible for the laws of nature.

1)iii. *Go conventionalist*

"*We* get to stipulate the winning candidate; having done so, we agree on it thereafter." (Note that arbitrariness of a choice was a key part of Lewis's (1969) definition of that choice being a matter of *convention*.) This response works for some of the laws of society, up to a point—e.g. the setting of speed limits, the voting age, and drinking age. Again, this response is not so promising for the laws of nature—we don't get to decide what they are. Some philosophical problems should not be farmed out to sociologists.

1)iv. '*Nature is kind*'

"The multiplicity problem is not really a problem, because however we reasonably make the arbitrary choice, there will be the same clear winner. While there might in principle be disagreement among the multiple best candidates for some job description, *in fact* such disagreement will not arise. Think of this as an empirical bet that nature is kind to us, and will see to it that all these candidates agree." Lewis favoured this response regarding the laws of nature. He thought that on any reasonable choice of the best system, certain regularities will keep appearing—e.g. the law of gravitation (or perhaps its relativistic correction). This response is perhaps not so effective for an *analysis* of lawhood that is intended to hold in other possible worlds, including those in which nature is unkind; but perhaps our concept of lawhood would not apply in such a world.

2) Symmetry-preserving responses

2)i. *Pluralism*

"All of the candidates are right. Each provides a legitimate meaning, a reasonable explication of an inexact concept or a reasonable precisification of a vague concept." This may be a good response when the arbitrariness at issue is purely semantic.

At the other end of the spectrum, we have

2)ii. *Eliminativism*

"None of the candidates are right. The multiplicity of candidates serves to show that the original concept is incoherent, and should be eliminated." This may be a good way to bring out problems with a concept that was already in dispute, as 'law of nature' is to some extent. But there is a danger of this response 'proving too much'. Most of our concepts are vague, and susceptible to sorites reasoning. We may be left with very few of them if we wield this response too enthusiastically. To be sure, however, even ordinary objects are under threat from such reasoning by eliminativists such as Unger (1979) and van Inwagen (1990).

Somewhere between the extremes of this response and the previous one, we have

2)iii. *It's a terminological matter*

"It's not that any given candidate is right or wrong. Proponents of different candidates are merely taking different stands on a terminological issue." Again, this may be a good response when the arbitrariness at issue is purely semantic. But of course the setting of speed limits, the drinking age, or voting age is not a terminological matter.

2)iv. *It's indeterminate*

"There is no fact of the matter of what the right candidate is. Rather, it is indeterminate." This could be a good explanation of why the leading candidates are tied; or their being tied could explain why the matter is indeterminate.

2)v. *Supervaluate*

"What's true on *all* ways of making the arbitrary choice is determinately true. What's false on *all* ways is determinately false. Everything else is indeterminate." Lewis suggests that if nature is not kind, and different candidates for the best theories disagree on what the laws are, then the laws are those theorems *common to all* of them. On this approach, there is the threat that there will be little or no overlap between the best theories, in which case there will be few laws or none. We might hope that nature is at least a bit kind, guaranteeing a decent amount of overlap.

Recall the problem for Stalnaker's account of counterfactuals that there may be ties for which antecedent-world is nearest. He later (1981) responds by supervaluating over the candidates.

2)vi. *Subvaluate*

"What's true/false on *all* ways of making the arbitrary choice is determinately true/false. Everything else is true *and* false." On this approach, there is the threat that there will be too many laws. And some of them may be highly disjunctive, piecing together regularities favoured by one best system with those favoured by another, and yielding something that is not simple by any standard.

Here, and in following pluralism by eliminativism above, we encounter another mini-heuristic: *do the opposite*. Take some approach to a problem, and follow the opposite approach, dual approach, or complementary approach. (Suitably cautioned by my first heuristic, this assumes there is exactly one such approach!) For example, replace universal quantifiers by existential quantifiers, necessities by possibilities, conjunctions by disjunctions, or replace parameter values by values that are in some sense complementary. In this case, subvaluationism can be regarded as the opposite approach to supervaluationism.

And even here we face some arbitrariness, now at the meta-level. How do we justify super-valuating and subvaluating *over* other ways of meta-valuating, quantifying over the ways of choosing candidates? To be sure, supervaluating and subvaluating are *salient*, in virtue of being extreme, the two endpoints of a spectrum of ways of valuating: what's true on at least one way of making the choice; what's true on at least two ways; what's true on at least three ways; … what's true on all ways of making the choice. But salient too is 'majority-rules-valuating', which appeals to a point of symmetry, the midpoint: what's true on *more than half* of the ways of making the arbitrary choice is true; what's false on *more than half* of the ways is false.[3] (What's true on exactly half the ways we might treat as indeterminate, or we might treat as true and false.) Moreover, how are we to choose *between* supervaluating and subvaluating, when they are equally salient? *This* choice threatens to be arbitrary. A big debate begins here, turning on considerations that militate in favour of one approach or the other, and this is not the place—or even *a* place—to enter it.

3) Hybrid responses

These responses combine some symmetry-breaking response (to cull some of the candidates while leaving others live) with some symmetry-preserving response (to treat those that remain even-handedly). One might insist, for instance, that *some* (but not all) of the candidates are privileged; then supervaluate over those that remain. Or one might consider all of the candidates that are subjectively chosen by some agent

[3] If there are only finitely many candidates, it's clear what 'more than half' of them means—we just count them all and divide by 2! But if there are infinitely many candidates, it's less clear. We will need some sort of *measure* over the candidates. But which? Arbitrariness—dare I say it—looms!

or other, and subvaluate over those. And so on. This yields many hybrid responses, mixing and matching techniques from the previous two categories.

I have presented many ways of responding to arbitrariness. Which should be used in a given case? Is that *arbitrary*?! How should we *respond*?! Different ways are appropriate for different cases, and I don't have a heuristic for deciding *that* (yet). I suggest that where possible, one should look to the symmetry-breaking strategies first, and only when they fail to leave just one candidate standing, look to the symmetry-preserving strategies over those that remain. But this still leaves open many possible responses, which will need to be handled on a case-by-case basis. Fear not—I have no aspirations to turn students into philosophical automata. There will always be an important place for good philosophical judgment, and here's one such place.

So much for arbitrariness. It is continuous with the next set of closely related heuristics, which I will put under one big heading.

3. Continuity

3.1 Continuity reasoning

Let's revisit the sorites paradox from a different perspective. You can draw a line in the sand, so to speak, and claim that there is a particular grain at which a heap suddenly comes into being—say, the 17th. But it seems absurd that such a small change in one respect, the difference between 16 and 17 grains, should result in such a large difference in another respect, the non-existence or existence of a heap. Note that the problem here is not that of arbitrariness, justifying why the line should be drawn there rather than elsewhere. Rather, the problem is that *there shouldn't be a line at all*.

This is an example of *continuity reasoning*. Roughly, the pattern is that one variable is a function of another, and small changes in the former should lead to small changes in the latter. Such reasoning is often part of commonsense. We are surprised, for instance, when we are told that we share 98.4% of our DNA with chimpanzees. How can such a small change in genotype lead to such a large change in phenotype? More generally, discontinuities may induce some of the discomfort that arbitrariness may cause us. Why should cases that are similar in some relevant respect give rise to such dissimilarity in another respect? And where there is a discontinuous 'jump' in some function of interest or importance to us, both the placement and the size of the jump may seem disconcertingly arbitrary.

It should be stressed how much continuity reasoning underwrites inductive inference. We don't just think that the unobserved resembles the observed (in suitable respects); we also think that the *nearby* unobserved *closely* resembles the observed, and typically the more nearby, the closer the resemblance (other things being equal). The world mostly does not deliver abrupt changes; properties tend to change gradually over space and time. Consider how painting restoration, or computer programs for reducing noise on photographs, operate on this assumption: where there is information on only particular parts of a picture, the default assumption is that nearby parts will be the same, or similar. Induction would be stymied if things systematically underwent sudden jolts. And when it is, that just adds to our discomfort!

Continuity reasoning is also common and fertile in philosophy. Sider (2002) poses the problem that certain conceptions of Hell are incompatible with a traditional doctrine about God. According to these conceptions, after we die, we either go to Heaven or Hell; some of us go to Heaven and some go to Hell; Heaven is much better than Hell; and God decides who goes where. The problem is that any criterion for His decision will admit of borderline cases. As a result, there will be some people who just make the cut and go to Heaven, and other very similar people who just miss out. Now Sider appeals to a continuity premise: "the proportionality of justice prohibits very unequal treatment of persons who are very similar in relevant respects" (59). He argues that one cannot square such treatment with God being just.

One may argue in a similar way for vegetarianism being morally required. Clearly we should not be *omnivores*—e.g. cannibalism is surely morally prohibited, and most of us would agree that so is eating monkeys.[4] If we are meat-eaters at all, we must have some criterion for deciding which animals it is permissible to eat and which not—as it might be, intelligence or sentience. Again this criterion will admit of borderline cases. Some animals will just make the cut (as it were) for being off limits for eating, and others will just miss out. And now comes the continuity premise: animals that are so similar to each other in relevant respects should not be given such disparate treatment. Conclusion: all animals are off limits.

We could argue along similar lines against the death penalty. Whatever criterion we use for drawing a line on the basis of severity of crime, beyond which perpetrators are executed, their punishment would be strikingly different from that of similar offenders just short of the line. Indeed, there is a whole class of "slippery" slope arguments that have been deployed against euthanasia, abortion, gun control, and so on, which in each case appeal to a continuity intuition that similar cases should be treated similarly.

Of course, there are ways of fighting back. For starters, continuity reasoning can be run in both directions. We can equally 'show' that any collection of grains of

[4] I bracket extreme cases in one's very survival is at stake, which may not be so clear.

sand, however small, is a heap by starting with a paradigm case of a heap, removing grains of sand one by one, and insisting that no single removal could turn a heap into a non-heap. The very same continuity premise that underwrites an argument for vegetarianism could just as well underwrite an argument for omnivorism. And eventually some crackpot is bound to appeal to continuity reasoning to argue for the far more widespread use of the death penalty, perhaps even supplanting parking fines and speeding tickets. Beware of the *slippery slope* that the unbridled use of continuity reasoning could send us down! And if we drive continuity reasoning, as we might say, *both* left-to-right (→) *and* right-to-left (←), we'll arrive at a contradiction (→←).

Moreover, continuity-based arguments must appeal to some sort of *metric*, at least loosely specified, for each relevant variable—some measure of 'distance' according to which we can judge roughly how close entities or cases are to each other. (It need not be numerical, but it must be more than merely a comparative ordering.) Disputes might arise over the choice of metrics, and favouring certain metrics over others might appear to be arbitrary.

And whatever the metrics, sometimes discontinuities are acceptable, and even required. My favourite function is the *Alan Hájek* function. It is the function from people to {0, 1} that is my characteristic function: it assigns 1 to me, and 0 to everyone else. Offhand, this function should be discontinuous, whatever reasonable metric we impose to capture how similar people are to one another.[5] It doesn't matter how close you are to me according to such a metric; you still get a 0, and only I get a 1!

[5] It will trivially be judged continuous according to the *discrete* metric: if $x = y$ then $d(x, y) = 0$; otherwise, $d(x, y) = 1$. But I assume this metric is not reasonable here.

3.2 Drawing inspiration from the mathematics of continuous functions

Let's look to mathematics for a better understanding of continuity, through its treatment of continuous functions. The informal definition of such a function is that small changes in its input value result in correspondingly small changes in its output value. More formally, the function *f*(*x*) is *continuous at c* if

$$\lim_{x \to c} f(x) = f(c).$$

This presupposes an underlying metric, as becomes clear when the limit is given its usual 'ε … δ' definition.

We may generalize this for a function between two topological spaces: a function $f: X \to Y$ is continuous if the pre-image of every open set of *Y* is open in *X*. We may think of this as a kind of 'supervaluating' over all metrics, preserving what is structurally common to all continuous functions, whatever the underlying metric. This may go some way to allaying the concern, raised above, that the choice of any particular metric is arbitrary.[6] Weber and Colyvan (2010) give a topological version of the sorites paradox. Kelly (1996) appeals to the topological definition of continuity to characterize empirical methods, regarding possible data streams as infinite sequences of discrete inputs, and open sets as empirically verifiable propositions that are empirically verifiable by finite sequences of such inputs.

We philosophers can outsource a lot of our problems, and let practitioners of other fields do a lot of the hard work for us. In this case, various beautiful theorems involving continuity that mathematicians have proven can do philosophical work. Start with the *intermediate value theorem:*

> *If f is a real-valued continuous function on the interval [a, b], and u is a number between f(a) and f(b), then there is a c ∈ [a, b] such that f(c) = u.*

[6] Thanks to Mark Colyvan here.

The underlying idea is simple. Think of a continuous function as one that you can trace without your pen ever leaving the page. Start at its value at the beginning of a closed interval, and trace it until you reach its value at the end. Along the way you must have crossed every intermediate value between those two endpoint values—in particular, any designated intermediate value.

We can use the intermediate value theorem to argue that at any time there must be a pair of antipodal points on the Earth that have the same temperature (and pressure too!) More philosophically, Joyce (1998) appeals to this theorem in his argument for probabilism, the thesis that rational credences obey probability theory. He shows that if your credences violate probability theory, then they are *accuracy-dominated*: there exists a probability function that is closer to the truth in every possible world. The intermediate value theorem supports a key step in the argument.

Or consider Brouwer's fixed point theorem:

> *Any continuous function f from a closed interval of the real line to itself has a fixed point—a point x_0 such that*
>
> $$f(x_0) = x_0.$$

Arntzenius and Maudlin (2010) ingeniously invoke this theorem to resolve a paradox concerning time travel. An old-fashioned camera takes a picture of a developed black-and-white film—a 'negative'—that leaves a time machine. The picture is developed, and the negative put in the time machine, sent back to the time at which the picture is to be taken, and leaves the time machine then. The trouble is that a negative has the complementary shades to the object of which it is a picture—a dark grey object has a light grey negative, and so on—so the story appears to be contradictory. But there is a neat solution. Represent the mapping from a shade of grey to its complementary shade as a continuous function on [0, 1], with 0 representing pure black and 1 representing pure white. By Brouwer's theorem, it has a fixed point: a shade of grey that is its own complement. So if the developed film is uniformly that shade, the story is consistent!

I adverted to a famous theorem involving continuity in my argument that there can motion without velocity: the existence of an everywhere-continuous function that is nowhere-differentiable. Interestingly, the Weierstrass function is the limit of an infinite sequence of functions, each of which is *everywhere*-differentiable. The anomalous behaviour of the function kicks in suddenly 'at infinity', in this sense behaving nothing like the functions that approach it.

3.3 Discontinuity at infinity

This is an instance of another philosophically important phenomenon, which we might call *discontinuity at infinity*. This is not the place to characterize this technical notion rigourously, but roughly it involves the failure of a natural extension to the definition of continuity to cases where we can make sense of a function's behaviour at infinity:

$$lim_{x_n \to \infty} f(x_n) = f(\infty)$$

where $<x_n>$ is an increasing sequence. It could be a sequence of ordered *functions*, with *f* representing a binary property that another function could have or not. For example, everywhere-differentiability of a function could be represented by a 1, its failure represented by a 0; then the Weierstrass function scores a 0, even though it is the limit of a sequence of functions each of which scores a 1.

Again, we recognize discontinuity at infinity in various philosophical examples. It figures in certain paradoxical decision problems that involve infinity in some way. Nover and Hájek's (2004) *Pasadena game* has a pathological (indeed, undefined) expected utility that is the limit of a sequence of perfectly well-behaved expected utilities. Arntzenius, Elga and Hawthorne's (2004—same issue!) 'Satan's Apple' involves an infinite sequential choice problem that becomes paradoxical only when the limit at infinity is reached. And Pollock's problem of when to open the Ever-better wine displays a similar discontinuity at infinity: the option of waiting forever and never opening the wine is discontinuously worse than all of the options that approach it.

While I applaud using mathematical theorems to support philosophical points (and have done so myself), we should keep in mind Benacerraf's (1967) caution about deriving philosophical conclusions from mathematical theorems. *When somebody purports to establish a philosophical thesis with a mathematical theorem, they must have assumed some philosophical premises.* In fact, keeping this caution in mind is a good philosophical heuristic in its own right! Examples of apparent violations of Benacerraf's cautionary words include:

- Putnam (1981): the Lowenheim-Skolem theorem shows that realism is false.
- Lucas (1961) and Penrose (1989): Gödel's theorem shows that minds are not machines.
- Various people: the Dutch book theorem shows that credences are rationally required to be probabilities.

None of these theorems establish what they are purported to without the aid of ancillary philosophical premises. Spelling them out clarifies what the associated arguments really should be, and in doing so fixes targets for further debate. I, for example, assumed that a particle could in principle move according to a Weierstrass function.

3.3 Continuity reasoning in philosophical methodology

But let's continue with continuity reasoning, this time applied at a 'meta'-level. Continuity considerations can be applied to philosophical methodology itself. Consider Priest's (1994) *Principle of Uniform Solution:*

> *If two paradoxes are of the same kind, then they should have the same kind of solution.*

Think of paradoxes as situated in 'paradox space', and solutions as situated in 'solution space'. The Principle can be regarded as a requirement that nearby paradoxes should be mapped to nearby solutions.

Again, we may ask: what is the appropriate metric—when are two paradoxes of the same kind? Smith (2000) presses this concern, arguing that two paradoxes may be similar at one level of abstraction, while being dissimilar at another. And is discontinuity sometimes acceptable—two paradoxes of the same kind having quite different solutions? I suggest that it is. Think of a limiting case: the *very same* paradox may get two quite different solutions, both adequate. (They may target different premises, for starters.) All the more, then, it would seem that two nearby paradoxes could have two quite different solutions, both adequate.

Notice how similar the Principle of Uniform Solution is to van Fraassen's (1989) "Symmetry Requirement":

> *Problems which are essentially the same must receive essentially the same solution.*

More generally, I think that the symmetry requirement is kindred with continuity reasoning. Both are useful heuristics. Jaynes (2003), for instance, appeals to the symmetry requirement in his defence of the principle of indifference.

3.4 Continuity and modal induction

Earlier I emphasised the important role that continuity reasoning can play in induction, ampliatively inferring a conclusion about the actual world from a premise about this world. Continuity reasoning can also play an important role in what we might call *modal induction*: ampliatively inferring a conclusion about the space of *possible worlds* from a premise about this space.

In my (forthcoming) I offer various methods for showing that something, call it *X*, is possible. One such method can be regarded as employing continuity reasoning. It follows this schema:

1) *Almost-X* is possible.
2) The small difference between *almost-X* and *X* makes no difference to possibility.

Conclusion: *X* is possible.

Chalmers (1996) argues along these lines that physical-duplicate zombies are possible (beings that have no conscious experiences, but that are physically identical to normal agents that have conscious experiences). Functional-duplicate zombies, he argues, are possible (functionally identical to normal agents); and moving from functional-duplicates to physical-duplicates will not make a difference to what's possible.

4. Mismatch of degrees

It's a problem for a putative analysis of some concept, *C*, if *C* comes in degrees that vary continuously, while the analysans has discontinuous 'jumps', or vice versa. More generally, it's a problem if there is a mismatch of the degrees of the analysandum and the analysans. A notable case of this is when one side of a putative analysis comes in degrees, while the other does not. This in turn may involve a mismatch of *vagueness*, one side admitting borderline cases, the other not.

4.1 The analysandum does not come in degrees, analysans does

Berkeley was fond of saying that *to exist is to be perceived*. (For some reason, he was particularly fond of saying it in Latin.) Now, existence does not come in degrees—it is the ultimate on/off property or attribute. But offhand, being perceived does. Think of perceiving a table in a totally dark room, in which you slowly turn the lights up using a 'dimmer' dial, or think of things on the periphery of your visual field. Moreover, these provide borderline cases of being perceived; but it is hard to make sense of borderline cases of existence.

To be sure, we might reply on behalf of Berkeley that 'being perceived' is a threshold notion—e.g. perceiving the table suddenly begins at a certain minimal light level. But that introduces a disquieting discontinuity, and an apparent arbitrariness in the setting of the threshold. Then again, we might use one of the many responses to arbitrariness that I detailed in §2, so I do not claim this objection is decisive. Berkeley himself replied that everything that exists is perceived by God; and presumably he would add that God's perceptions don't come in degrees.

Personal identity seems to be 'all or nothing', yet various analyses of it involve things that come in degrees (e.g. psychological or bodily continuity—not the same sense of "continuity" as before!). But—another useful philosopher's technique—one man's modus ponens is another's modus tollens. And Parfit (1984) tollenses where others ponens, arguing that personal identity indeed comes in degrees. (If it does, my earlier insouciance about the discontinuity of the *Alan Hájek* function may need revisiting.)

Now turn things around (another philosophical heuristic in its own right!):

4.2 The analysandum comes in degrees, the analysans does not.[7]

Hume (1748/1902) defines a miracle as a violation of a law of nature. But arguably, being a miracle comes in degrees, whereas being a law of nature does not, and thus violating a law of nature does not. Some miracles are more *miraculous* than others. Arguably, a resurrection would genuinely violate a law of nature. (If not, then Hume's definition is in even more trouble, since this about as miraculous an event as any attested to.) But other miracles need not. For example, the Red Sea parting on Moses' command need not. Consistent with the laws, one water molecule after another could spontaneously move this way or that way, collectively constituting the parting of the sea. To be sure, the event is extraordinarily *improbable*, given the laws. Perhaps, then, a miracle should be defined that way. After all, improbability comes in degrees, and is thus fit to match the degrees of the analysandum.

According to the von Mises (1957)/Church (1940) analysis of randomness, a sequence is random iff *every* recursively specified subsequence has the same relative frequency of every attribute. If this analysis is unfamiliar to you, and indeed if even the terminology in the analysans is unfamiliar to you, so much the better—the power of this heuristic will be revealed all the more. For even if the right-hand-side makes you glaze over, you know that *universal quantification does not come in degrees*. (The same is true of existential quantification.) Either it is (entirely) true that *every*

[7] If you or your students have trouble remembering which is which, here is a mnemonic: the analysan*dum* is the *dumb* thing (the concept as it is found in the wild, pre-reflection), and the analys*ans* is the *ans*wer (as provided by some thoughtful philosopher).

blah-blah has yadda-yadda, or it is (entirely) false; the analysans does not come in degrees. But surely randomness comes in degrees. Suppose we toss a coin indefinitely. Consider interspersing the completely random sequence (say)

H H T H T T T T H …

with the obviously *non*-random alternating sequence

H T H T H T H T…

to yield:

H H H T T H H T T H T T T H T T H…

Surely this resulting sequence is *partially* random. But without paying any attention to the details of the von Mises/Church analysis, you know that it cannot deliver that verdict. (In fact, its verdict is that the sequence is *non*-random—entirely so!)

This account of randomness was intended to undergird von Mises' *frequentist* account of (objective) probabilities: they are relative frequencies in random sequences of trials. Philosophers now mostly agree that frequentism provides a bad analysis of probability (see Hájek 1996 and 2008 for many arguments for this conclusion). But *a bad analysis can provide a good heuristic*. It will typically not be a total failure—it will typically get a wide range of central cases right. (If it did not, it presumably would never see the light of day.) While failing to give necessary and sufficient conditions for its analysandum, it may nevertheless be a usually-reliable guide to the analysandum, reliable enough to serve as a useful way to think about the associated concept.

This brings us to the next heuristic—or better, set of heuristics.

5. Replace non-extensional notions with extensional surrogates

Kahneman and Tversky (1982) have famously contended that we are bad at reasoning probabilistically—witness our tendency to neglect base rates or to commit the conjunction fallacy on various questions concerning probability assignments. For example, people are prone to think that if they test positive for a disease when given a test that is 95% reliable—it has a 5% chance of giving false positives and a 5% chance of giving false negatives—the chance that they have the disease is 95%; this neglects the prior probability that they have the disease in the first place, which may render that chance much smaller. Gigerenzer (1991) has found that if such problems are rephrased in terms of frequencies, we fare much better. This is not to say that probabilities *are* frequencies; just that frequentist thinking is a good heuristic for probabilistic thinking. I believe that a key reason for this is that frequencies, unlike probabilities, are *extensional*, and as such are easily pictured. Imagine, say, that you are one of 10,000 people who have taken the test, and that it showed positive. Now suppose you learn that the base rate of the disease is 1%, so only 100 people in the population have the disease, and 9900 don't. 95 of the former group (truly) showed positive, and 495 of the latter group (falsely) showed positive (naively identifying probability with frequency, as I am recommending here!). That is, less than 1/5 of the people in your shoes actually have the disease—that should be comforting. Frequentist thinking makes the point intuitive and vivid.

More generally, I propose the heuristic: *replace intensional notions with extensional surrogates*. (An intensional notion is one for which truth value may fail to be preserved under replacement of co-referential expressions; an extensional notion is one for which truth value is preserved.) I submit that the extensional notions are easier to think about and to deal with; the intensional notions are more *opaque* to us.

Some examples are well entrenched in philosophical thinking:

- Replace talk of *necessity* or *possibility* with talk of *what's true at all* or *some possible worlds*.
- Replace talk of *counterfactuals* with talk of *what's true at the nearest antecedent worlds*.

Here at least, most philosophers think that the extensional surrogates really are equivalent to the original intensional notions. In some other cases, the alleged equivalence is controversial, but the heuristic value is there nonetheless. For example:

- Replace talk of indeterminacy with talk of what's true on some but not all admissible precisifications, and replace talk of what's determinately true/false with talk of what's true/false in all admissible precisifications.
- Replace indicative conditionals with *material* conditionals. (Material conditionals are extensional since they are truth functional, and the extension of a sentence is its truth value.)

Or perhaps better is a two-step replacement:

- Replace indicative conditionals with strict conditionals, and replace *them* with material conditionals true in all possible worlds.

The first step replaces an intensional notion with another intensional notion, but one that has a simple extensional translation.

Then there are analyses whose failure is relatively uncontroversial nowadays, yet still they may be useful ways to think about the relevant concepts, if only as an act of pretence:

- Replace talk of laws of nature with talk of universal generalizations over individual entities.
- Replace talk of causation with talk of constant conjunction of event-types.

Indeed, a large part of the failure in each case stems from the extensional analysans not capturing the *intensionality* of the analysandum![8]

I have so far discussed the handling of intensional notions. But even more recalcitrant to our natural modes of thinking are *hyperintensional* notions—ones for which truth value may fail to be preserved even under replacement of *necessarily* co-referential expressions. Properties, at least when finely individuated, are often taken to be examples—think of *triangular* and *trilateral* as distinct properties, yet necessarily co-present or co-absent. Still, there is an extensional surrogate:

- Replace talk of properties with talk of sets of possible individuals.

We may generalize our heuristic, then, to become: *replace non-extensional notions with extensional surrogates*.[9] This covers both ways for a notion to fail to be extensional: intensional, and hyperintensional. Set-theoretic treatments of concepts are extensional, and they often provide helpful replacements of tricky intensional and hyperintensional concepts.

And so it goes. We might come up with our own ways of unpacking a non-extensional context in an extensional way. I like this one:

- Replace talk of norms with universal quantification over all norm-abiding agents. E.g. replace 'one has a moral obligation to treat others as ends rather

[8] A reason why these extensional analyses clearly fail, where some considered earlier appear to succeed, may be that these quantify over extensional entities, whereas the successful ones quantified over entities that themselves might be regarded as intensional (possible worlds, a modal notion). It is less clear that the analysis of (in)determinacy is successful; but then, perhaps it is less clear whether the notion of a precisification being *admissible* is itself intensional. (If it is tacitly modal, then presumably it is.) To the extent that an analysis traffics in entities that are themselves intensional, perhaps it is not truly an extensional analysis after all. I bracket that concern here, trusting that applications of my heuristic are easy to recognize in any case.

[9] Thanks to Daniel Nolan for this formulation.

> than means' with 'all morality-abiding agents are agents who treat others as ends rather than means'.

The 'replace non-extensional notions by extensional surrogates' heuristic works especially well when the replaceans has the same logic as the replaceandum, sanctioning exactly the same inferences. This is clearly the case when replacing probabilities with frequencies—the latter obey the probability axioms (with finite additivity). It is arguably also the case with the extensional replacements of necessity and possibility (with suitable choices of accessibility relations), and of counterfactuals. But even when it is clearly not the case, the replaceans can serve as a guide to reasoning—defeasible, to be sure. Mill's methods for identifying causation are really methods for identifying constant conjunction; and while causal inference has come a long way since, they still provide good rules of thumb. We should just remember that good rules of thumb are not the last word.

This heuristic is powerful, I think, because it is easier to *picture* things when they're extensional. This brings me to my final heuristic.

6. Draw a picture

This heuristic again draws on a psychological fact about us, this time a familiar one. Conceptual relationships that may be obscure when cloaked in words or symbols often leap out at us when we represent them visually. Think of the power of Venn diagrams to represent set-theoretic relationships. And the surrogates above that can be cast set-theoretically immediately lend themselves to Venn diagrams. The foregoing heuristic, then, works particularly well in tandem with this one: *translate from non-extensional to extensional notions, then diagram the latter*.

Example. If you draw a Venn diagram of the disease test example above, with areas representing classes of individuals roughly proportional to their respective population sizes, the importance of the base rate becomes even more obvious.

Example. Thanks to the Kripke semantics for various systems of modal logic, we can easily diagram how different modal systems, corresponding to different assumptions on an 'accessibility' relation among worlds, validate different inferences.

Example. Having given his own possible worlds analysis of counterfactuals, Lewis (1973a) goes on to present various counterfactual fallacies, such as antecedent-strengthening:

p □→ q

∴ (p & r) □→ q.

It may not be immediately obvious that this is not a valid argument form. But he goes on to draw diagrams that allow one to *see* easily that it is not.

However, the 'draw it' heuristic can stand alone, unaided by the previous heuristic. For example, it helps to diagram causal relationships even without entertaining the fiction that causation is merely constant conjunction—either with Lewis-style (1973b) 'neuron diagrams', or Pearl-style (2009) causal networks.

Physicists know well the power of pictures—for example, drawing a Minkowski space-time diagram rather than performing a deduction using the mathematics of special relativity. When a rod with clocks at each end moves relative to you, the clock at the front runs behind the clock at the back. Showing this is relatively easy with a diagram, relatively hard with the mathematics. Mathematicians are similarly well versed in the heuristic utility of pictorial representations of abstract relationships. Indeed, Nelson (1993) is an entire book of proofs without words or symbols. Again, we philosophers can import some of our heuristics from other disciplines.

7. Some metaphilosophical ruminations

It is another psychological fact about us that we are limited in our ability to see what follows from what. (The Wason selection task famously brings this out.) We are prone to drawing illicit inferences from our premises, and we are often blind to their consequences, which are sometimes unwelcome. Here, logic is our great tool, our consistency and validity policeman. But I must emphasise again that it is by no means our only such tool.

I can characterize a good chunk of the philosophy that I find congenial with the slogan *'making our implicit commitments explicit'*. We collect, deploy and systematize intuitions, analyze arguments, conduct thought experiments, refine definitions, adduce normative constraints, and so on; and when we are doing our job properly, we check for the consistency and for any unwelcome consequences of our products (and celebrate the welcome ones). The heuristics are to a large extent aids to these enterprises.

Of course, you don't need to be a philosopher to make our implicit commitments explicit. Mathematicians do it too, and they remind us what a worthy enterprise it can be. If my slogan sounds like it trivializes philosophy, we should remember that it may be no easy feat. Nobody thinks that it's easy to make explicit how our commitment to certain basic facts about positive integers, addition and exponentiation implicitly commit us to Fermat's Last Theorem.

If we think of logic as an all-purpose tool for checking what follows from what and what is compatible with what at a high level of abstraction, then the heuristics collectively form more of a Swiss army knife. Some of them have rather more specific targets or uses. "See definite descriptions in neon lights" is a bit like a Swiss army knife's scissors: useless for many occasions, but exactly what you need for some. We want heuristics that have some generality, but that also have real bite. Too general, and they become useless: e.g., "make an insightful point!" Too specific, and they become next to useless again: e.g., "when someone argues that space is purely relational, reply with Kant's problem of incongruous counterparts!" The best heuristics lie somewhere in the middle of the spectrum: general enough to be applicable in a wide range of cases, but not so general as to be empty. But of course the heuristics can vary considerably in their generality among themselves, and still earn their keep. The Swiss army knife's scissors may have a greater range of uses than its corkscrew, but on certain occasions the latter is exactly what you need.

Conclusion

I've given a small sampler of some philosophical heuristics. I have chosen them because there are some nice interconnections among them, and I have tried to present some of them in considerable detail to illustrate their fruitfulness from a number of angles. But I could have chosen any number of others.

As well as studying the *content* of our best scientific theories, we study the *methodology* of the best scientists—see e.g. Harper's recent book (2012) on Newton's methodology. Doing so not only tells us something about how science gets done; it also can inspire us as to how to do it well. Similarly, studying the methodology of the best philosophers not only tells us something about how philosophy gets done; it also can inspire us as to how to do it well. I have learned from them many of the philosophical heuristics on my ever-growing list.

To the extent that we fail to pay attention to such heuristics, we miss out on rich resources for philosophical thinking. Yet I think that by and large, philosophers have been singularly unreflective about them. I invite you to reflect on them with me.

REFERENCES

Arntzenius, Frank, Adam Elga, and John Hawthorne (2004): "Bayesianism, Infinite Decisions, and Binding", *Mind* 113 (450), 251-283.

Arntzenius, Frank and Maudlin, Tim (2010): "Time Travel and Modern Physics", *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2010/entries/time-travel-phys/>.

Benacerraf, Paul (1967): "God, the Devil, and Gödel", *The Monist*, 51: 9-32.

Carnap, Rudolf (1952): *The Continuum of Inductive Methods*, Chicago: University of Chicago Press.

Chalmers, David J. (1996): *The Conscious Mind*, Oxford: Oxford University Press.

Church, A. (1940): "On the Concept of a Random Sequence", *Bulletin of the American Mathematical Society*, 46: 130–135.

Elga, Adam (2004): "Defeating Dr. Evil with Self-Locating Belief", *Philosophy and Phenomenological Research* 69:383–396.

Gigerenzer, Gerd (1991): "How to Make Cognitive Illusions Disappear: Beyond 'Heuristics and Biases'" in: W. Stroebe & M. Hewstone (eds.), *European Review of Social Psychology* 2, 83–115.

Hájek, Alan (forthcoming): *Philosophical Heuristics and Philosophical Creativity*, in *The Philosophy of Creativity*, eds. Elliot Paul and Scott Barry Kaufman, Oxford: Oxford University Press.

Harper, William L. (2012): *Isaac Newton's Scientific Method*, Oxford: Oxford University Press.

Hume, David (1748/1902): *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, Second Edition, 1902.

Jaynes, E. T. (2003): *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press.

Joyce, James M. (1998): "A Non-Pragmatic Vindication of Probabilism", *Philosophy of Science* 65, 575-603.

Kahneman, Daniel and Amos Tversky (1982): "Judgment Under Uncertainty: Heuristics and Biases", in *Judgment Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic and Amos Tversky, Cambridge: Cambridge University Press.

Kelly, Kevin (1996): *The Logic of Reliable Inquiry*, Oxford: Oxford University Press.

Lewis, David (1969): *Convention*, Cambridge MA: Harvard University Press.

Lewis, David (1973a): *Counterfactuals*, Oxford: Blackwell.

Lewis, David (1973b): "Causation", *Journal of Philosophy* 70: 556–567.

Lucas, J. R. (1961): "Minds, Machines and Gödel," *Philosophy* 36, 112-127.

Lycan, William (2009): "Giving Dualism Its Due", *Australasian Journal of Philosophy* 87 (4), 551-563.

Nelson, Roger B. (1993): *Proofs Without Words*, The Mathematical Association of America.

Nover, Harris and Alan Hájek (2004): "Vexing Expectations", *Mind* 113 (450), 237-249.

Nozick, Robert (1993): *The Nature of Rationality*, Princeton: Princeton University Press.

Parfit, Derek (1984): *Reasons and Persons*, Oxford: Oxford University Press.

Pearl, Judea (2009): *Causality*, Cambridge: Cambridge University Press, 2nd edition.

Penrose, R. (1989): *The Emperor's New Mind*, Oxford: Oxford University Press.

Priest, Graham (1994): "The Structure of the Paradoxes of Self-Reference", *Mind* 103, 25-34.

Pollock, John (1983): "How Do You Maximize Expectation Value?", *Nous* 17, 409-421.

Polyá G. (1957): *How to Solve It*, 2nd ed ., Princeton University Press.

Priest, Graham (1994): "The Structure of the Paradoxes of Self-Reference", *Mind* 103, 25-34.

Putnam, Hilary (1980): "Models and Reality" *The Journal of Symbolic Logic* 45 (3), 464-482.

Sider, Ted (2002): "Hell and Vagueness", *Faith and Philosophy* 19, 58–68.

Smith, Nicholas J. J. (2000): "The Principle of Uniform Solution (of the Paradoxes of Self-Reference)", *Mind* 109, 117-22

Stalnaker, Robert (1968): "A Theory of Conditionals", *Studies in Logical Theory,* N. Rescher (ed.), Oxford: Oxford University Press, 98-112.

Stalnaker, Robert (1981): "A Defense of Conditional Excluded Middle", in Harper, W.L., Stalnaker, R., and Pearce, G. (eds.), *Ifs,* Reidel, Dordrecht.

Unger, Peter (1979): "There Are No Ordinary Things", *Synthese* 41: 117–154.

van Fraassen (1989): *Laws and Symmetry*, Oxford: Clarendon Press.

van Inwagen (1990): *Material Beings*, Ithaca: Cornell.

von Mises Richard (1957): *Probability, Statistics and Truth*, revised English edition, New York: Macmillan.

Weber, Zach and Mark Colyvan (2010): “A Topological Sorites”, *The Journal of Philosophy* 107, No. 6 (June 2010), 311–25.