Philosophical Heuristics and Philosophical Creativity

Alan Hájek

1. Introduction

They say that anyone of average intelligence and moderate talent can become a strong, competition-level chess player by mastering and internalizing certain heuristics. (Who are "they"? Actually, it was a friend of mine, who said it once. But I believed her.) These are captured by slogans such as "castle early and often", "avoid isolated pawns", and so on. They are *chess heuristics*. Analogously, philosophy has a wealth of heuristics—I will call them *philosophical heuristics*—although they have not been nearly so well documented and studied. This paper is partly an introduction to a larger project of mine of identifying and evaluating philosophical heuristics. I also intend it to be a contribution to the philosophy of creativity: I will argue that such heuristics can enhance one's ability to make creative contributions to philosophy.

2. A working definition of 'creativity'

I do not have anything especially creative to say about what *creativity* consists in,[1] and defining it precisely will not be essential to my achieving my goals here. I will be happy to take as a working definition one that is almost universally adopted by psychologists, which is focused on creative *products*. A product—an idea or artifact—is *creative* to the extent that it is (a) novel and (b) valuable.[2]

Each condition is apparently necessary. The first condition is obviously so, although we may want to follow Boden's (2004) distinction between *H*-creativity, short for *Historical* creativity

---

[1] The ensuing discussion largely follows Gaut (2010).
[2] To be sure, products are not the only things that may be creative—*people* and *acts* may be as well, but they will mostly not be my concern here.

(nobody else has produced the product before), and *P*-creativity, short for *Psychological* creativity (the product is new to a particular individual, although it may have been produced previously by someone else, unbeknownst to the individual). The second condition is also apparently necessary: we hardly deem products that are worthless as creative, however original they might be. To be sure, we may regard some surprising but ultimately unsuccessful leaps of thought as 'creative' in some sense; but the scare quotes seem appropriate, this sense being more of a pejorative one, and not my topic here. The condition may still be questioned, however, since we may want to allow for 'dark' creativity, the ingenious creation of all-too-successful malevolent products. If so, our discussion is becoming somewhat terminological. Let me simply stipulate that my topic is philosophical creativity *in the good sense*; then the necessity of condition (b) comes for free.

Some philosophers contest the sufficiency of conditions (a) and (b) for creativity. (See Gaut 2010.) The main source of concern is that a product that is produced by the wrong sort of *process* might meet conditions (a) and (b), but still not count as creative—for example, if someone produces it by pure luck. In section 11 I will downplay the importance of the psychological processes that lead to a philosophical product in our judgments of philosophical creativity, since those processes are almost always obscure to us in others, and often even in ourselves. And my primary interest is in those products to which we *ascribe* creativity. In any case, the prospect of a wrong sort of process leading to a valuable product in philosophy does not seem to be as live as it is in art, say—a philosophical gem is unlikely to be produced by pure luck, for example. So my focus will be on philosophical *products*—philosophical positions, arguments, analyses, counterexamples, puzzles, paradoxes, and so on—however they were produced. That's what we ultimately value. I will regard them as valuable to the extent that they advance philosophical understanding, although this too will not be essential in what follows.

Identifying philosophical heuristics has been an ongoing project of mine at least in the background, and sometimes in the foreground, for many years. It might help you to understand the project if I tell you some of the things that led me to it. If you are impatient to get to the heuristics themselves, you could skip the following Apologia. Once we have met a few of the heuristics and seen them in action, we will be in a position to discuss their relationship to *philosophical creativity*.

3. Apologia

It all began when I was a graduate student. I had the chance to observe a lot of good philosophers in action, and I wanted to be a good philosopher. Sometimes I was reading their work; sometimes I was seeing them in action in person. Often I had a feeling of wonder at what appeared to be moments of inspiration from these philosophers: "Wow! That was impressive." I began trying to figure out if some of their skills were learnable, and to the extent that they were, I wanted to learn them. Over the years I have found good philosophers repeatedly using certain techniques, perhaps unconsciously much of the time, which *can* be easily learned.

After graduate school, I taught at Caltech for a number of years. I had highly talented students who typically knew little about philosophy, for ten weeks at a time. How was I to convey to my students in such a short time how philosophy is done? Of course, the main thing was to *show* them lots of good instances of philosophizing—to have them read some of the classics, and some of the best of the relevant recent literature—and then to get them to philosophize themselves. But that did not seem enough to me.

For consider some other skill—say, skiing. A skiing instructor does not just say: "You've seen people ski well; now do it yourself! Go on: SKI!" Rather, the instructor gives you *skiing heuristics*, breaking down skiing into manageable bits: "shift your weight to the downhill ski",

"keep your hands forward", and so on. Yet in philosophy we typically just show our students finished pieces of philosophy—the classics, the recent literature—and then effectively say: "You've seen people philosophize well; now do it yourselves! Go on: PHILOSOPHIZE!" I thought we could do better.

So I began to think about identifying manageable techniques in these philosophical showpieces. To be sure, the teaching of such heuristics was a tiny part of my courses—they came more as quick asides as we read through the classics or the recent literature. ("Here Hume is using the 'proves too much' strategy", or what have you.) But pointing out the heuristics as they showed up along the way was certainly valuable. And it was fun.

My introductory course on metaphysics and epistemology began with Descartes, and I drew more of my inspiration from him. In my opinion, he got much of good philosophical methodology right, nearly four hundred years ago, in his *Discourse on Method* (1637). He would start with a hard problem, break it down into smaller parts, perhaps breaking them down still further, until eventually he reduced his original problem to a bunch of smaller problems of manageable size, and for which he had "clear and distinct ideas". (Of course, he had much more to say about his methodology than that!) In my case, the hard problem starts as *doing philosophy*—which is really a set of hard problems: trying to come up with an original philosophical position, or analysis, or argument, or puzzle, or paradox; or trying to come up with a counterexample to someone else's philosophical position or analysis, or trouble for an argument of theirs, or solving a puzzle, or resolving a paradox. (There are other conceptions of philosophy, of course, but I am characterizing the kind of philosophy that will be of most interest to me here.) The heuristics provide some tools for breaking these problems down into easier sub-problems, and the better they are at doing this, the more useful they are.

For *philosophy is hard*. I feel that I could use all the help that I can get, and I suspect you could too. The heuristics help in various ways. The first way has an obvious connection to creativity: the heuristics are useful while you are waiting for inspiration to strike. Suppose that you have been beating your head on a problem, waiting for a lightning bolt of brilliance to strike you. I hope it does. In the meantime, some of the heuristics may help you. Or you may have a hunch that there is something wrong with a philosophical position, but you are struggling to nail what it is. Again, some of the heuristics may help. They are also useful when you want to be self-critical. Suppose you have just written a philosophical paper, or a thesis, or a book. It is curiously difficult to be a good critic of your own work, to see problems that might be obvious to someone else; meanwhile, would-be critics are *lining up* eagerly to point them out to you. By somewhat mechanically running through the heuristics, you are more likely to find trouble-spots, or ways of improving your position: a counterexample to your view here, an illicit inference that you made there, a way of strengthening an argument here, a way of generalizing your point there, or applying it to a new domain.

More generally, the heuristics provide ways of *making one's implicit commitments explicit*. (Here, "one" may be *you*, or it may be someone else with whom you are engaging.) Sometimes these commitments are unwelcome, and they should force one to rethink the views that led to them; sometimes they are welcome, and one should celebrate them! To be sure, formal and informal logic already provide some safeguards against making illicit inferences, and they may help us make one's implicit commitments explicit. But focusing solely on logic is an overly narrow conception of the ways in which philosophers reason. We have a much richer toolbox at our disposal. My project is to identify far more of our tools.

Caveats:

- There are many distinct abilities that go into making a good philosopher, and I do not pretend to provide heuristics for all of them, or even a tenth of them. As I say, and will keep saying, these are merely *heuristics*. They are not guaranteed to work, any more than avoiding isolated pawns is guaranteed to strengthen your position in a chess game. My project is not one of finding foolproof philosophical algorithms, or surefire recipes for philosophical progress—there are none. But the heuristics do tend to work, often enough that I think they earn their keep, and then some.

- The heuristics are no substitute for depth. I am not promising wisdom, or profound insight. But I hasten to add that what I offer is not antithetical to them, either, and indeed I think it helps rather than hinders progress to such wisdom or insight. Again, compare chess. No chess instruction book is crazy enough to say that you will become a grandmaster once you have learned to castle early and to avoid isolated pawns. On the other hand, no chess instruction book apologizes for recommending that you do these things; indeed, if it didn't it would be remiss. Closer to home, mathematics professors explicitly teach their students various techniques for solving problems, and they often unapologetically teach heuristics for proving things (e.g. 'When you want to prove something that seems obvious, try *reductio ad absurdum*'), without compromising their students' potential for deep mathematical insights. Still closer to home, teaching our students formal and informal logic provides no guarantee of leading them to wisdom, or profound insight, but doing so helps rather than hinders progress to such wisdom or insight. We don't apologize for doing so, and indeed if we didn't we would be remiss. Again, I regard the heuristics as more tools from the same toolbox.

- Choosing the right heuristic, or the best heuristic, for a given situation usually requires good judgment—and yes, creativity. And having seized upon a given heuristic, there may be many ways in which it can be applied, and again there is a role for creativity here.

****

Enough talking for now about this exercise in metaphilosophy—enough meta-metaphilosophy (until §11). It is time to bring on a few of the heuristics. The list I present here is just the tip of the tip of the iceberg; I will settle for seven of them, when as I say, I have hundreds. The ones I have chosen are some of my favourites. Moreover, as we will see, some of them are conducive to 'constructive' philosophizing, which might be regarded as especially creative.

Let me begin with an easy heuristic, which I think is quite fruitful.

4. Check extreme and near-extreme cases

4.1 Check extreme cases

Start with a hard problem: someone proposes a philosophical position or analysis and you are looking for trouble for it, because you suspect that there is something wrong with it. (The 'someone' might be you, in which case your job is to find trouble for your own position before someone else generously does it for you.) Try this simpler problem: look for trouble among *extreme cases*—the first, or the last, or the biggest, or the smallest, or the best, or the worst, or the smelliest, or ... It is a snappy way to reduce the search space.

Even if there are no counterexamples lurking at the extreme cases, still they may be informative or suggestive. They may give you insights that you would have missed by focusing on more run-of-the-mill, typical cases.

'Check extreme cases' is a good heuristic, three times over:

- It is often easy to use. Extreme cases of the relevant kind are often easily identified.

- It is fertile. Extreme cases are often troublesome in virtue of being extreme—that is, sometimes problems that could not arise in normal cases can arise at the extremes. (For example, sometimes it is important that normal cases are 'surrounded' by other, similar cases, whereas extreme cases are 'loners', isolated.)

- We are liable to forget or miss extreme cases, since we tend to live our lives in the thick of normal cases.

To be sure, there will not always be extreme cases of the relevant kind. They often correspond to maxima or minima of appropriate functions. But some functions have neither—for example, those that are unbounded in both directions, or those that approach but never reach upper and lower asymptotes. It is not clear, for example, that we can make sense of Leibniz's notion of "the best of all possible worlds"—it seems that we can imagine better and better worlds without end. (Imagine a sequence of worlds in which there are successively more happy people, or more happy rabbits …) Nor can we obviously make sense of "the worst of all possible worlds". So if our topic is the goodness of worlds, we may not be able to find any extreme cases. No matter; while this heuristic is not universally applicable, it is widely applicable.

Pólya (1957) notes the value of this heuristic in mathematics. This reminds us that good philosophical heuristics need not be purely *philosophical* heuristics; indeed, it would be surprising if they always were. Some of them are heuristics for good thinking more generally. But I will confine myself to philosophical examples.

4.1.1 Some philosophers regard 'every event has a cause' to be a necessary truth. At first, one may wonder how to argue against this claim—where should one start? The heuristic guides the search for a counterexample: start with extreme events. For instance, start with *the start*. The *first* event is an extreme event: the big bang. There was no prior event to cause it; it surely did not cause *itself*; and it surely was not retro-caused by some later event—so we have our

counterexample. To be sure, this presupposes that there was exactly one big bang. As far as I know, this is a respectable presupposition by the lights of current cosmology. But it does not matter if it is not. Whether or not there was the big bang is presumably a *contingent* matter. No necessary truth can settle the matter either way.

4.1.2 Lewis (1986b, 213) writes: "Some philosophers wish to believe only in entities that have some causal efficacy." For instance, Armstrong (1989) does not believe in Platonic universals for that reason; some philosophers of mathematics do not believe in numbers, and in abstract objects more generally for that reason. Lewis continues: "either they must reject such totalities as the big event which is the whole of history, or else they should correct their principle." Similarly, if we regard the whole of history as an event (the ultimate one), it provides another counterexample to the dictum that every event has a cause.

4.1.3 What is it for F's to be *ontologically dependent* on G's? Here is a very natural analysis: it is not possible for F's to exist without G's existing. (See Thomasson 1999.) But consider an extreme case of existence: something that exists *necessarily* (not merely contingently). According to this analysis, for any F and for any *necessary* G, F is ontologically dependent on G. If the number 3 is necessary, then we are all ontologically dependent on it. Now *there's* an unexpected argument for the Holy Trinity!

Relatedly, what is it for a theory T to be *ontologically committed* to G's? Here is a very natural analysis: it is impossible for T to be true without there being G's. But then for any theory T and any *necessary* G, T is ontologically committed to G. If God exists necessarily, then atheism is ontologically committed to Him. If numbers exist necessarily, then Field's (1980) nominalism about numbers is ontologically committed to them. Obviously this does not capture what is distinctive about atheism, or Field's nominalism.

4.1.4   According to the von Mises/Church definition (von Mises 1957; Church 1940) of *randomness*, an infinite sequence is *random, with respect to an attribute,* if every recursively selected subsequence has the same limiting relative frequency of the attribute. It is *random* if it is random with respect to every attribute. By this definition, the extreme sequences of outcomes of tossing a coin forever,

heads, heads, heads, …

and

tails, tails, tails, …

are random. After all, *every* subsequence of the first sequence has the same limiting relative frequency of heads: 1 (the limit of the sequence 1/1, 2/2, 3/3, …). And *every* subsequence has the same limiting relative frequency of tails: 0 (the limit of the sequence 0/1, 0/2, 0/3, …). Similarly for the second sequence. But they strike me as maximally *non*-random sequences.

4.1.5   According to decision theory, you should act so as to maximise your expected utility: the sum of products of utility and corresponding probability, for each possible state of the world, associated with your action. A gamble with *infinite* expected utility provides an extreme case. Since expected utility is a sum, we are guided to consider ways in which sums can be infinite, and so to consider gambles whose expectations are series that diverge to infinity. The St. Petersburg game is a celebrated example. A fair coin is tossed until it lands heads for the first time. The longer it takes, the better for you. Your rewards escalate exponentially according to the following schedule:

First heads on toss number	Probability	Pay-off
		2

$$\frac{1}{!^n} \qquad\qquad 4$$

$$\qquad\qquad 3$$

$$\frac{1}{!^n} \qquad\qquad 2^n$$

Let us identify utilities with dollar amounts (or if you prefer, recast the game in terms of utilities directly). The expected utility of the game is

$$(\frac{1}{2} \times 2) + (\frac{1}{4} \times 4) + (\frac{1}{8} \times 8) + \ldots = 1 + 1 + 1 + \ldots = \infty$$

So according to decision theory you should be prepared to pay any finite amount to play the game once. This seems crazy. In fact, $100 seems too much to most people.

4.2 Check near-extreme cases

Extreme cases can sometimes be dismissed: we might say that we lack trustworthy intuitions about them, or that they are pathological, or even that they are not really cases at all of the notion in question. (All three things have been said about the St. Petersburg game, for example.) So it can be useful to move to *near*-extreme cases instead.

Kenny Easwaran has told me that during preparation for a mathematics competition, he was coached to look at near-extreme cases. For example, if you have to prove a statement concerning all triangles, it may be helpful to see how it would look for a triangle with, say, a 179-degree angle and two angles of half a degree. A truly extreme case might be unhelpful, or even impossible—e.g. a 'triangle' with a 180-degree angle and two 0-degree angles is a straight line, and not really a triangle at all. So try a near-extreme case instead.

4.2.1   von Mises might reply to the all-heads and all-tails counterexamples that they are degenerate cases of random sequences—for example, they are what you get from two-headed or two-tailed coins, and they are sequences as random as such coins could possibly produce. Then consider a near-extreme case:

  Tails, Heads, Heads, Heads, …

It still passes the von-Mises/Church test for randomness, but it still looks highly *non*-random to me, and it is not degenerate. I would say the same of sequences with finitely many tails at places specified by some function; or even sequences with infinitely many tails, spaced progressively further apart according to various non-recursive functions. Yet they all pass the test.

4.2.2 A popular defensive move in response to the St. Petersburg game is to insist that utility functions must be *bounded*. In that case, there can be no games of infinite expectation. But then consider a truncated, finite St. Petersburg game, with a very high truncation point—say, the game is called off if the first heads has not been reached by 100 tosses. There is a good sense in which this game is like the 'extreme' St. Petersburg game, just not as extreme. It still seems that decision theory overvalues the game (at $100). But this problem no longer relies on unbounded utility.

So the heuristic here is two-stage: *first look to extreme cases to spot trouble for a philosophical position; then, retreat to a less extreme case, where the same trouble may still lurk*. To be sure, it may lurk to a lesser degree; but it may also be a more plausible case, and one that may resist some of the defensive moves that may work in the extreme case. This heuristic may also be useful in cases where there are no extreme cases of the relevant kind. Still, there may be cases that we regard as near-extreme, and in which trouble arises.

Having identified trouble for a philosophical position, one should try to go on to diagnose its *source*. The trouble will typically be just a symptom of a deeper underlying problem. The St.

Petersburg and truncated St. Petersburg games make vivid the fact that decision theory looks only to the 'location' of the distribution of the random variable associated with a given option. It pays no heed, for example, to the 'spread' of the distribution, as represented by its variance. If variance (and perhaps higher moments of the distribution) are relevant to choiceworthiness, then these gambles not only provide counterexamples to decision theory; they also enhance our philosophical understanding. And so it goes with counterexamples more generally: ideally, they do not merely serve the negative function of refuting a particular philosophical thesis, but they also teach us positive lessons about some philosophical concept or issue.

5. Death by diagonalization: reflexivity/self-reference

You can't bite your own teeth, unless something has gone badly wrong for you, dentally speaking. You can't see your own eyes—not directly, anyway—unless something has gone *very* badly wrong for you, optically speaking.

The next heuristic bids us to take a philosophical thesis, and to make it refer to *itself*, to plug into a function *itself* as its own argument, and more generally, to appeal to *self-referential* cases. This technique is another handy way of cutting down the search space when you are looking for counterexamples.

Let us take our cue from Cantor's 'diagonalization' proof of the uncountability of the reals, or Gödel's proof of the incompleteness of arithmetic, or the halting problem, or Russell's paradox, or the liar paradox. They remind us of the august history of the technique of self-reference; its application can yield profound results.

5.1 I begin with a non-philosophical example that is not remotely profound as a lighthearted warm-up. My friend Alan Baker has two (aptly-named) cooking conjectures:

Baker's first conjecture: Everything tastes better with either chocolate or garlic added.

(There's a putative cooking heuristic for you.) This has much prima facie plausibility. But we don't need to look far for two clear counterexamples: *chocolate*, and *garlic*.

Baker's second conjecture: everything tastes worse with both chocolate and garlic added.

And here we can find another clear counterexample: *a mixture of chocolate and garlic*!

5.2 Sorensen (1996) opens with this sprightly instance of the reflexivity heuristic:

Here is a debate I cannot lose. I argue that it is possible to (at least unwittingly) believe the impossible, say, that there is a largest prime number. The *impossibilist* objects that I am mistaken. Wrong move! By trying to correct me, the impossibilist concedes that I believe a false proposition. The proposition in question (i.e. that impossibilities can be believed), if false, is necessarily false. Thus, the impossibilist would be conceding that an impossibility can be believed. (247)

Sorensen is a master of this heuristic more generally. His "Faking Munchausen's Syndrome" (2000) gets my vote for the funniest title in philosophy. (Munchausen's syndrome is a mental disorder wherein those afflicted fake having a disease that they do not in fact have.)

5.3 Realism is often stated as a thesis of mind-independence: 'to be a realist about X is to be committed to the mind-independence of X. But what about realists about *minds*, as most of us are? They are surely not committed to minds being mind-independent.

5.4 Kim (1982) suggests that an *intrinsic* property is one that can belong to something unaccompanied. Lewis (1999) replies: *unaccompaniment* is not an intrinsic property, yet it can belong to something that is unaccompanied.

5.5 According to the betting interpretation of subjective probability, your degree of belief in a proposition is the price (in cents) at which you are indifferent between buying and selling a bet that pays $1 if the proposition is true, and nothing otherwise. But I have degrees of belief about my own betting behavior—e.g. I am confident that I will not enter into any bets today. This degree of belief cannot be understood in terms of a betting price of mine.

5.6 Homework: This puzzle is due to Raymond Smullyan (2000). You can choose one of the following two offers. Which should you choose?

Offer 1. You get $10.

Offer 2. You get $10 if you say something true, and something other than $10 if you say something false.

Hint: offer 2 is *much* better!

The next heuristic is closely related.

6. Self-undermining views

It is an embarrassment for a philosophical position if it is *self-undermining*. This can happen when the position itself falls in the domain that it purports to cover. The proponent of the position potentially faces a charge of a kind of philosophical hypocrisy (presumably unintended).

6.1 Quine (1951) famously said: "Any statement can be held true ... no statement is immune to revision." Stove (1991) asks: is *this* statement immune to revision? Quine's dictum is itself a statement, and as such falls in the domain of statements over which it quantifies.

6.2 "Truth is relative, not objective" is an oft-heard slogan in these post-modern times. But someone who says it seems to regard *it* as an objective truth.

6.3 There is the story of Putnam, in one of his incarnations, arguing along these lines in a lecture: Either logic is empirical, or it is a priori; but logic is not a priori, since quantum mechanics teaches us that disjunctive syllogism is an invalid argument form; thus, logic is empirical. But as an audience member pointed out, this argument assumes that disjunctive syllogism is valid. (See Jammer 1974.)

6.4 According to the verificationist theory of meaning, the meaning of a sentence is determined by the method by which it is empirically verified, and in particular, a sentence that is unverifiable is meaningless. But it seems that the very statement of this theory is itself unverifiable, and so by verificationist lights is meaningless.

6.5 More generally, consider a skeptic about meaning who claims that there is no fact of the matter of what our words mean. What are we to make of the sentences that he or she uses to state the position?

****

All of the heuristics I have considered so far might be regarded as 'negative'—they are strategies for coming up with counterexamples for other philosophers' analyses, or problems for other philosophers' views or arguments, rather than for constructing such analyses, views or arguments in the first place. Even philosophizing that is 'negative' in this sense can be creative—witness Lewis's counterexample to the thesis that we should believe only in entities that have some causal efficacy, if you need any convincing of that. It is original, imaginative, ingenious, illuminating, and compelling; what more could you want?! And I have noted that counterexamples may also function positively to illuminate a philosophical concept or issue.

Recall, moreover, that in order for a philosophical product to be creative, it is necessary for it to be valuable. These 'negative' heuristics may provide a valuable service in evaluating an existing product—if it survives the threats that they potentially pose to it, this counts in favour of it. For example, if an analysis takes extreme cases in its stride, more power to it. If on the other hand it succumbs to such cases, then this may prompt philosophical progress—the analysis might be modified to handle the cases, yielding a superior final product. A similar point can be made in favour of the reflexivity heuristic. Think of how fertile Russell's paradox has been as a touchstone to set theory—a creative contribution if ever there was one, albeit 'negative'.

And if the analysis falls by the wayside as a result, that may still lead to philosophical progress in a way that resonates with a widely endorsed model of creative thinking in psychology. According to the GENEPLORE model, creative thinking consists of an interaction between two kinds of processes—those involved in GENERATING ideas and those involved in EXPLORING ideas to assess their potential value, significance or utility. (See Ward and Kolomyts 2010 pp. 94-5.)[3] I will soon present some "positive" heuristics that are tailor-made for the generative phase of the philosopher's creative process. Meanwhile the "negative" heuristics may be used in the exploratory or evaluative phase, as a way of assessing the positive ideas that one has already generated. To be sure, my focus in this paper is primarily on creative *products* rather than on the processes leading to them. But if the heuristics may play a role that psychologists recognise in the processes as well, all the better. And the heuristics need not be used purely negatively; sometimes they may fertilize the imagination, the seed of creativity, leading to more 'positive' philosophizing, whatever that might mean.

In any case, let me turn now to some more 'positive' heuristics, to the extent that it makes sense to speak this way. I hope that it will be clear how they may foster creative philosophy.

7. Begetting new arguments out of old

Arguments are often easily transformed from one domain to another.

7.1 Space ⟷ time ⟷ modality

Arguments involving space can often be rewritten to create parallel arguments involving time; arguments involving time can often be rewritten to create parallel arguments involving modality; and we can reverse these directions.

7.1.1 Parfit (1984) has an argument for the irrationality of discounting the future that turns on the absurdity of a similar spatial discounting.

7.1.2 Lewis (1993) provides a nice example of space → time → modality transformations of arguments. He discusses Unger's (1980) "problem of the many": an object such as a cloud has various questionable spatial parts—do we count a water droplet at the outskirts of the cloud as part of it or not? But then many aggregates of droplets of various sizes are equally good candidates to be the cloud. So how many clouds are there? It looks like the answer is: many. Or maybe the answer is: none (each candidate is ruled out by competition by the others). But we don't get the correct answer: one.

In his discussion of this argument, Lewis shows how to convert it into one involving questionable temporal parts (those parts near the beginning or the end of an object's life), and one involving questionable other-worldly parts.

7.2 Modal notions

Various modal notions are closely related: laws of nature, causation, counterfactuals, chance, necessity, and dispositions. For this reason, we often see parallel treatments of some of them. For example, Hume's theory of causation as constant conjunction parallels the regularity theory of laws, according to which the law that all F's are G's is simply the regularity that all F's are G's; this in turn parallels the frequentist theory of chance, according to which the chance of an event type on a trial is simply the relative frequency with which that type occurs in trials of that kind. And arguments against one sort of theory of one modal notion can often be transposed to become arguments against a parallel theory of another one.

Example. According to Hume's theory of causation, whether my hand going in *this* flame caused the pain in *my* hand depends on whether there is a constant conjunction between *other*

hands going in *other* flames and subsequent pains in *those* hands. But offhand (sorry!), it seems that causation is a far more localized matter: the sole protagonists are *this* flame and the pain in *my* hand, and whether or not this was a case of causation should be intrinsic to them. Similarly, according to frequentism about chance, whether *this* radium atom has a chance of roughly ½ of decaying in 1500 years depends on whether roughly ½ of all the *other* radium atoms in the universe decay in 1500 years. But offhand, it seems that chance is a far more localized matter: the sole protagonist is *this* radium atom, and its chance of decay should be intrinsic to it.

7.3 Rationality ←→ morality

Many problems and theses in theories of rationality can be rewritten as problems and theses for various metaethical theories.

7.3.1 The St. Petersburg paradox for decision theory can be rewritten as a problem for a natural version of consequentialism. Imagine a world in which there are denumerably many people (or other loci of value). Order them and label them by the natural numbers: 1, 2, 3, … Consider this action: To person 1, give $2 with probability ½; to person 2, give $4 with probability ¼; to person 3, give $8 with probability 1/8; … If we calculate the total utility in the world by summing the expected utilities given to each person, we get $1 + 1 + 1 + … = \infty$. Then according to a version of consequentialism that regards the moral status of an action as the total expected utility that it accrues, then this action should be judged morally superior to every action that gives with certainty a finite quantity to a finite number of people, no matter how large the quantity and the number are. This seems crazy.

7.3.2 Pollock's (1983) example of the Ever-better wine causes trouble for the principle of rationality that one should maximize expected utility. 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 seems to be too soon; yet never opening it is the worst option of all. Now transform this into a moral problem: the longer you wait to administer the Ever-better drug to a patient, the greater it will benefit her. When should you administer it? We can specify the case so that any time seems to be too soon; yet never administering it is the worst option of all.

8. Some ways to argue that *X* is possible

Why care about what is possible? Because what is possible is relevant to various things that philosophers care about—claims of necessity, supervenience, entailment, validity, equivalence, and conceptual analyses, according to all of which certain things or combinations of things are *impossible*. Moreover, philosophers often deploy thought experiments in support of their positions; when they do so, they typically assume that the scenarios that they have us imagine are in some sense possible.

There are various senses of 'possible' (logical, metaphysical, nomological, epistemic, doxastic, deontic, …), and the methods for showing that something is possible in the relevant sense vary somewhat accordingly. The first method below for showing that something is possible works for most of them; the later ones are usually targeted at metaphysical possibility, but some of them generalize. Even where they may not be decisive, they may nonetheless provide good guides to what is possible.

Suppose you want to show that something, call it *X*, is possible. (*X* may be a proposition, or a state of affairs, or a being, or what have you.) There are two components to the claim '*X* is possible': '*X*', and 'is possible'. Accordingly, there are two main ways of arguing for the claim:

1. Begin with some *other suitable property* of *X*, and infer that *X* is possible.

2. Begin with *something else*, *Y*, that is possible and suitably related to *X*, and infer that *X* is possible too. (Or begin with two or more things, *Y*, *Z* ...)

8.1 Begin with some other suitable property of X, and infer that X is possible

8.1.1 Actuality

Actuality implies possibility for most of the modalities;[4] so if you have a good grip on what actuality is like, you have a head start on supporting various possibility claims. And actuality gets surprisingly weird—think of quantum mechanics, and special and general relativity.

8.1.2 Positive probability

The next argument form is:

1. *X* has positive probability.

2. If *X* has positive probability, then *X* is possible.

Hence,

3. *X* is possible.

Different senses of probability will lead to different senses of possibility. If the probability is *objective chance*, then we may infer *physical possibility* (and hence metaphysical possibility). If the probability is *subjective* (of some rational agent), then we may infer *epistemic* and *doxastic* possibility (for that agent).

Example. Quantum tunneling of medium-sized dry goods is physically possible—for example, it is physically possible for you to quantum tunnel to the North Star. After all, according to quantum mechanics, this has positive probability (tiny, admittedly!), and it is thus

---

[4] Not doxastic or deontic.

possible. And knowing this fact, you should assign positive subjective probability to it (equally tiny, to be sure!), and it thus should be epistemically and doxastically possible for you.

8.1.3 Conceivability

The next argument form is:

1. *X* is conceivable.

2. If *X* is conceivable, then *X* is possible.[5]

Hence,

3. *X* is possible.

Example: Chalmers (1996) argues this way for the possibility of zombies (physical duplicates of someone with a mental life, who lack mental lives).

8.1.4 Arbitrariness

Lewis (1986a, 103) observes that "when something seems arbitrary, we are apt to think that it might well have been different". If not-*X* is apparently arbitrary, that is a reason to think that *X* is possible.

Example. It is apparently arbitrary that the gravitational constant has exactly the value that it does—so that is a reason to think that it could have been otherwise, against necessitarians about laws such as Shoemaker (1980).

---

[5] A clause may need to be added that excludes defeaters. There is considerable controversy over this form of argument, but this is not the place to enter the fray!

Now that we have under our belts some ways of arguing that various things are possible, we can use them as starting points to argue that various *other* suitably related things are possible. This brings us to our second main way of arguing that *X* is possible.

8.2 Begin with something else, Y, that is possible and suitably related to X, and infer that X is possible too. (Or begin with two or more things, Y, Z, …)

8.2.1 Begin with almost-X

1. Argue that almost-X is possible.

2. Argue that the small difference between almost-X and X does not make a difference to what's possible: either both are possible, or neither is.

Examples:

8.2.1.1 Bigelow and Pargetter (1987) argue, contra Lewis (1986a), that it is possible to have "island universes" within a given world that are not spatio-temporally related. They contend that it is possible to have a world with near-islands as parts, joined only by a worm-hole. Then they argue that removing this worm-hole would make these parts entirely spatio-temporally isolated from another; but the small difference that this makes does not make a difference to what's possible.

8.2.1.2 Adams (1979) argues, contra the identity of indiscernibles, that it is possible for there to be two indiscernible objects that are not identical. Clearly there can be two *almost* indiscernible spheres that are not identical. Then, just change one of them a little bit (e.g. remove a tiny impurity on one of them) so that they are indiscernible. That change does not make a difference to what's possible.

8.2.1.3 Contra behaviorism, it seems to be possible for someone to have a mental life, without any behavioral manifestation of it. Clearly someone who is nearly completely paralyzed can have a mental life—think of Stephen Hawking. Now imagine this person becoming completely paralyzed. Notice that method 8.1.1 supports the first step: near-total paralysis with mental life is possible, because it's actual.

Related:

8.2.2 *Extrapolation*

Begin with a clear case of possibility, and work by small steps to the case at issue. Someone who doesn't like this kind of reasoning might pejoratively call it 'slippery slope' reasoning; but sometimes it is effective. The method is a little different from the previous method, where there was just a single step.

8.2.2.1 Frank Jackson (in conversation) suggested the following argument that it is possible for something functionally like Blockhead—Ned Block's (1981) imaginary computer that can carry on a conversation by having all its responses pre-programmed—to be intelligent. There is no impugning your intelligence if you look up square roots in a look-up table. Now, slowly add more things that you look up; still your intelligence is not impugned. Eventually you get to something functionally like Blockhead.

8.2.2.2 Baldwin (1996) argues that an entirely empty world is possible. Consider a possible world with just finitely many objects. Now suppose each object disappears in turn. Eventually we reach the empty world.

The last two methods of showing that something is possible are fruitful (as I hope the examples have made clear), but they must be handled with care. Beware of 'showing too much' to be possible, by sequences of small steps beginning with something that is possible: *Y*, almost-

Y, almost-almost-*Y*, almost-almost-almost-*Y*, … If the relevant space is path-connected (in a suitable sense), then we can get from anywhere in the space to anywhere else by a sequence of small steps—but we may cross the line from possibility to impossibility along the way. It is obvious that this method has its limits. In fact, *limits* in mathematics provide good examples of its limits. Think of a function approaching but never reaching an asymptote, the way that $f(x) = 1/x$ approaches but never reaches 0. It is possible for the function to *almost reach* its limit (by any standard for 'almost'), but not possible for it to *reach* its limit.

Related:

8.2.3 Interpolation

Show that *W* and *Y* are each possible, and that *X* falls between *W* and *Y* on some relevant axis, or with respect to some gradable property.

Example: Hume's missing shade of blue is possible. Take two actual—and thus possible—shades of blue, one on either side of the missing shade, and interpolate.

8.2.4 Combinatorialism

Take any things that are separately possible, and put them together in any arrangement permitted by shape and size: according to combinatorialism, the result is possible. So by these lights, it is possible for a pub to sit on the top of Mt Everest.[6]

8.2.5 Physical symmetry principles

Physics has various symmetry principles according to which the laws of nature are preserved under various transformations. Start with a physically possible scenario, apply one of these transformations, and *voilà*, you have another one. For example, take a scenario

---

[6] Or at least a counterpart of Mt Everest, as Lewis (1986a) would have it.

that is physically possible, time-reverse it, and you get another scenario that is physically possible.

Example. John Norton (2003) imagines a mass sitting on top of a dome. The obvious solution to Newton's equations sees the mass continuing to sit there. But Norton argues that there are also infinitely many other 'unexpected' solutions, according to which the mass spontaneously begins moving and slides in some direction or other down to the base of the dome. To see that this is possible, consider the time reversal of one of these trajectories: the mass starts at the base of the dome, and it is projected towards the apex with just the right initial velocity so that it makes it to the apex, but no further. This trajectory is physically possible; hence, so is its 'unexpected' time-reversal. Norton concludes that classical mechanics is indeterministic.

More generally, the so-called *CPT symmetry* of physics implies that a 'mirror-image' of a physically possible world is physically possible—one with all objects having their positions reflected by an imaginary plane (parity inversion), all momenta reversed (time inversion), and all matter replaced by antimatter (charge inversion). Rich material for thought experiments!

9. 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.

9.1 Consider the problem of coming up with a single truth function that is expressively complete in the sense that we can express all truth-functional compounds in terms of it. Sheffer (1913) solved the problem: 'not and', otherwise known as the 'Sheffer stroke', and its dual 'nor', both have this property. Coming up with the problem was ingenious; but solving it need not have

been. After all, there are only 16 truth functions to check. So by running systematically through them, it should not take one long to come up with the answer.

9.2 There are various ingenious counterexamples to various theories of causation. According to Lewis's original (1973) theory, c is a cause of e just in case c and e both occur, and there is a chain of counterfactual dependences (of the form 'if c_i had not occurred then e_i would not have occurred') linking c to e. In cases of so-called "late preemption", there are two potential causes of an effect, one preempting the other, and the causal process leading from the preempted cause to the effect is cut off after the successful cause has brought about the effect.

Late preemption cases are ingenious counterexamples to Lewis's theory. But one could find them by a process of trial and error. 'Neuron diagrams' or 'causal graphs' are two ways of representing causal structures, with nodes representing events or variables, and arrows between them representing relations of causal promotion or inhibition. One could systematically work through various such causal diagrams, starting with simple ones and gradually building up their complexity, hunting for counterexamples. Eventually one would stumble upon late preemption cases—they require some complexity, but not that much. Similarly, the structure of many of the well-known counterexamples to other theories of causation could be found by trial and error. To be sure, having found the structure, one would then want to find a plausible real-world case that fits it. But that is not so difficult; the hard work is finding the structure that does the job in the first place. A trial-and-error approach removes the need for ingenuity there. More generally, you may find a model or structure that serves your purposes by trial and error; having found it, you may then look for a case that fits it.

And if you have some programming skills, you can write a program to do the hard work of finding the structure for you. More generally, where the search for a counterexample can be reduced to an algorithm, a computer can find it for you. (Not that you need to tell anyone that

that's how you arrived at it!) In this way the process of coming up with a creative philosophical product can be partly automated. Examples of using computer tools to discover philosophically interesting technical results can be found in some recent papers by Branden Fitelson. His (2008) shows how one can put to good philosophical use a user-friendly decision procedure for the probability calculus, and his (MS) shows how first-order theorem provers and model finders can yield an interesting generalization of Gibbard's theorem that, given certain seemingly plausible assumptions, the indicative conditional collapses to the material conditional.

Sometimes a *hybrid* strategy is the way to go: partly ingenuity, partly trial and error. Sometimes by ingenuity you suspect that a counterexample lurks in a certain relatively small sub-class of cases; then you can just work through them by trial and error (perhaps with the help of a computer).

Sometimes the trial and error heuristic can effectively work in tandem with one or more of the other heuristics. For example, you might use the 'check extreme cases' heuristic to narrow the search to a few 'corners' of the search space, and then go through each of them by trial and error.

The trial-and-error heuristic also serves as a meta-heuristic when you are otherwise at a loss: *systematically run through the heuristics until you find one that helps you*! It also works internally within some of the multi-faceted heuristics. For example, in the previous section I provided a long list of heuristics for arguing that *X* is possible, through which one could work systematically. You might be pleasantly surprised at what you come up with.

10. Future projects: dissertations and books waiting to be written

When you are looking for a big project to work on, take some big philosophical idea or program, and apply it to a new case. The scheme is to apply philosophical system *X* to specific problem *Y*, for suitable *X* and *Y*.

This is the closest I can come to producing a heuristic for producing ground-breaking philosophy. Here the thought is that rather than merely responding to someone else's agenda, you can do some agenda-setting of your own. And if ground-breaking philosophy when it succeeds doesn't count as creative, I don't know what does. However, even if the results are not quite so dramatic, still the heuristic encourages one to look beyond entrenched ways of thinking about an issue. System *X* is illuminated if a hitherto unrecognized application of it is revealed; progress may be made on recalcitrant problem *Y* if it is approached from a fresh perspective.

10.1 Kantian ethics has been fruitfully applied to a number of moral domains; find one to which it hasn't. Ethical issues concerning new technologies may be promising places to look—for example, privacy issues raised by the new social networks, like Facebook and Twitter. If this turns out to be passé—I'm always behind the curve when it comes to all things technological—I'm sure you can do better yourself.

10.2 Apply the Rawlsian theory of justice to intellectual property—again, some recent innovations may be good places to start.

10.3 The semantic view of theories—as opposed to the 'received' syntactic view—has been fruitfully applied to biology (Lloyd 1988) and to physics (Frigg 2006). As far as I know, it has not been applied to various other scientific and social-scientific theories—geology, astronomy, economics, and what have you.

10.4 The so-called 'Canberra plan' is an approach to conceptual analysis according to which we identify a folk-theoretical role for a concept to play, and then look to the world for the best

player of that role. Various concepts have been Canberra-planned—e.g. mental states (Lewis 1972, Jackson 1998), moral concepts (Smith 1992), and causation (Menzies 1996). I haven't heard of anyone Canberra-planning aesthetic concepts. I must tell my Canberra-based colleagues to get to work on that!

10.5 Various fictionalist accounts of problematic notions have been given—of numbers (Field 1980), unobservables in scientific theories (van Fraassen 1980), possible worlds (Rosen 1990), and morality (Joyce 2005). As far as I know, nobody has yet given fictionalist accounts of various notions in the philosophy of language, such as reference.

10.6 Closest to my heart, Bayesian confirmation theory has illuminated the confirmation of scientific theories. (See Howson and Urbach 2006.) I believe it has yet to be applied to the confirmation of historical theories.

And so it goes. Some of these avenues might be dead-ends—I'm not sure I'd want to take on Canberra-planning aesthetics!—but some might be fertile. And you may well be able to come up with better examples. To be sure, it takes good judgment to work out which of these projects are promising. Not just any pairing of a philosophical system with a specific problem will be fruitful. This heuristic, like any heuristic, can be misused, but then so can calculus and general relativity—that's hardly *their* fault. Once again, using heuristics only takes you so far; it leaves plenty of work for more sophisticated skills. Philosophical *noûs* will not be left jobless just yet.

****

We are about to change gears. Before we do so, here's a trivia question for you. Which two countries in the world have the letter 'x' in their name? I invite you to spend a couple of minutes thinking about this question if need be (take longer if you like). If you can't come up with the answer, so much the better. Soon I'll explain the point of this exercise.

11. Philosophical creativity and the heuristics

A number of philosophical heuristics are now before you. Let's step back for a moment and consider their relationship to philosophical creativity.

Most of the philosophers to whom I've spoken about this project of mine applaud it. But a handful of my interlocutors think that it should be discouraged, because they think that I will stifle the creativity of philosophy students. It's as if they have visions of my churning out legions of cookie-cutter students, tongues hanging out, parroting my heuristics at each other: "I refute your position with heuristic #17(ii)"; "No, no, no, I defend it with heuristic #328(c)"...[7] While I am flattered that my project should prompt such Orwellian fantasies, I have far less grandiose ambitions for it.

It would be strange if we had a similar concern about chess heuristics: "Yes, castling early helps—*but don't tell anyone, because it will stifle their creativity*!" Or should we be concerned that Pólya gave generations of mathematics students heuristics for solving mathematics problems, because it *stifled their creativity*? Much as I find it hard to take these imaginary concerns too seriously, I don't take the corresponding actual concerns about my project too seriously.

For starters, even after you have mastered a heuristic, or a hundred of them, there is still so much left over to challenge your philosophical acumen. There is plenty of room for creativity to come in there. Also, as I have noted, there are creative ways of using the heuristics that I offer. ('Castle early' is not like this—there are few ways in which you can execute that advice.) And the talented chess player would discover many of the heuristics for herself eventually in any

---

[7] I'm reminded of the old joke about the annual comedians' convention, at which the same comedians gather year after year. They all know the same jokes so well that they give the jokes numbers; when they get up on stage they merely call out numbers, to the gales of laughter of their audience.

case—she would notice soon enough the benefits of castling early, and so on. So it is with the philosophical heuristics. But above all, it is not as if the heuristics must *compete* for your limited cognitive resources. Surely you can *both* attend to some of the heuristics on a given occasion, *and* freely exercise your native cunning. Or if you cannot, then by all means attend to the heuristics for just a few minutes, and then set them aside so that your native cunning can work unfettered. Let me assure you: *it's not that hard*. Indeed, I will soon argue that the heuristics may promote rather than inhibit creativity.

Another objection that I have heard is that students should have good philosophical intuitions and instincts on their own. To which I reply: *What is this—survival of the philosophically fittest?* If you have such students, or are such a student, more power to you. But what about students who could use some more guidance? More to the point, as I have contended, even the best of us can benefit from internalizing these heuristics. And internalizing them only helps to instill good intuitions and instincts. Once internalized, they can become second nature. This objection makes a mockery of the idea of teaching philosophy. ('Go on: PHILOSOPHIZE!') Also, it isn't at all clear that a student who comes to philosophy with good intuitions and instincts hasn't learned them.[8]

The objection to my project that I take most seriously runs as follows. Sometimes one can be too self-conscious about what one is doing; it can be better just to *do* it, rather than to *think* about doing it. Once when I was playing pool with Andy Egan, he asked me whether I hold my breath or not while taking a shot. I had never thought about it before, and for the rest of the game I was self-conscious about my breathing, which threw me off my game—which of course was exactly his plan! And John Searle tells a story of when he was a ski racer, and he had an Austrian coach to whom he turned for advice after doing a run. The coach's advice was simple: *"Schneller!"*

---

[8] Thanks here to Katrina Hutchison.

("*Faster!*"). The coach's point was that Searle should not overthink what he was doing. Instead of being preoccupied with his weight distribution or hand position, he should just *think fast*, and let his body do the rest. Similarly, perhaps one should not overthink the doing of philosophy either, but rather get down to *doing* it.

I take this objection seriously, but I think it can be easily answered. For starters, in both the pool and skiing cases, presumably there was at least *some* instruction: there certainly was in the case of my pool-playing, and I assume there was in Searle's case too. So the objection should not be to all teaching of heuristics, but rather to too much of it. (Eating broccoli is good for you, but you should not each too much of it.) Similarly, I advocate using the heuristics, not *overusing* them. Then one should ask to what extent the learning of heuristics interferes with rather than enhances the expression of some skill. It surely does so more in sports than in intellectual enterprises. I can see how concentrating on a heuristic could temporarily slow down more automatic mental processes, but the philosopher hardly needs the split-second reaction times of the ski racer! And as one acquires expertise in an area, explicit rules and reasoning get replaced by automatic processes. When this happens, it becomes easier to perform other tasks concurrently. If using the heuristics works this way, there is still value in recognizing and teaching them, just as there is value in explicit ski instruction. But the expert philosopher may rarely explicitly use them (and her performance might be impeded by using them), just as the expert skier rarely explicitly thinks about how to ski (and her performance might be impeded by such thinking).[9]

But I am being too defensive. Far from being an impediment to philosophical creativity, the heuristics can actually enhance it. For starters, there is some experimental evidence that conscious processes play an essential role in creativity more generally, and even that self-

---

[9] Thanks here to Angela Mendelovici and Katrina Hutchison.

consciously striving *for* creativity is conducive to increasing it—see Baumeister et al., (this volume). Moreover, creativity can be enhanced by constraints[10]—poetry writing, for example, is sometimes improved when it is constrained by a given rhyming pattern or metre. The heuristics are not constraints, of course (it's not as if you *must* apply any of them), but they may *focus* and *guide* the mind much as constraints do.

More importantly, many of the heuristics show how hard problems can be broken down into easier sub-problems, so that *less* cognitive horsepower is required to solve them than the original problems required. The heuristics can work like hints, leading you closer to the solutions you seek. An analogy: reducing the distance between two pieces of metal makes it easier for a spark to jump across them. A closer analogy: Go back to the trivia question about countries with the letter 'x' in their name. How did you do? If you were unable to come up with the answer, try the following. You can be confident that no country's name *starts* with the letter 'x', but in any case, let me assure you that this is the case. So any 'x' must be preceded by another letter, presumably a vowel. (That's not a guarantee—just a good heuristic.) Now, slowly vocalize the sounds that could be so generated, while scanning your mental country database to check whether any names have these sounds in them:

"… ax …", "… ex …", "… ix …", "… ox …", "… ux …".

Did a spark suddenly jump for you? If not, relax, clear your mind as much as you can, and try again. If still not, don't worry—I'll reveal the answer shortly. You should still appreciate the point of the analogy. Loading up these sounds in your mind brings you slightly closer to your

---

[10] Thanks here to Uriah Kriegel.

targets, thus making it easier for you to find them.[11] If you have time, you could bring your mind still closer by running successively through all the consonants that might prefix these sounds:

"… bax …", "cax …", "…dax…", "… fax…", …

"… bex …", "cex …", "…dex…", "… fex…", …

Similarly, if you reduce the 'distance' of a philosophical target, it is easier for an idea to spark across to it, easier for you to make the mental leap that will complete it. *The heuristics help get you partway to your targets*. By making your tasks easier, then far from interfering with your creativity, the heuristics free you up to exercise it all the more. Far from competing with your creativity, they can *cooperate* with it.

Some of the heuristics have more of a 'brute force' nature, which may seem more antithetical to creativity. And yet even that may not in fact be the case. One good chess heuristic is 'check every check'—by trial and error running through a relatively small number of potential moves that directly attack your opponent's king, you may hit upon a powerful move. It is a good heuristic twice over: it is easy to use, and checks greatly restrict your opponent's immediate options. Indeed, my 'trial and error' heuristic is rather like this chess heuristic. Now suppose a chess player finds what looks like a brilliant queen sacrifice by systematically checking every check. Does that fact diminish the brilliance of the move? Similarly, suppose that you find what looks like a brilliant counterexample to a philosophical view by the exercise of some heuristic. Does that fact diminish the brilliance of the counterexample?

You may reply that brilliance is one thing, creativity another, and that while we may attribute brilliance to a product, creativity is essentially a matter of the process that led to it. A sterner

---

[11] David Chalmers used this strategy to answer the question successfully at a trivia night we recently attended.

critic of my project might go so far as to say that the very application of a heuristic diminishes to that extent the creativity of the product produced by it.

Let me begin my response with a thought experiment. Consider two philosophers who independently hit upon the same idea, one that we judge to be creative. Then we learn that one of them came to the idea directly by pure philosophical *noûs*, while the other self-consciously applied some heuristic. Should we *withdraw* our judgment that the latter philosopher's idea was creative, while maintaining our judgment that the former philosopher's was? That seems absurd. In that case, we ought to suspend pretty much *all* our judgments of creativity of philosophical ideas. If certain kinds of workings rather than others are essential to an idea's being creative, then we are rarely in a position to make judgments of creativity. Yet we frequently make such judgments. It's as if a version of the Turing test applies to philosophical creativity: we judge philosophical *outputs*, not the thought processes that produced them. If it *looks* creative, we count it *as* creative. How could it be otherwise? We seldom have access to the processes in the case of others, and we often have rather imperfect access to them even in our own case. Again, I am regarding the creativity of philosophical products as a matter of their originality and value, however they were produced. This accords with the definition of "creativity" widely used in psychology—in particular Amabile's Consensual Assessment Technique, according to which the best way to measure creativity is by using experts' assessment of the relevant products. (See Amabile 1996, especially pp. 33-5.)

Let me summarise. Starting with the philosopher, let's ask: is she less likely to produce a creative product if she uses heuristics such as I have identified? No—on the contrary, the heuristics make creative breakthroughs easier to achieve by making hard problems easier, and by reducing the distance to philosophical targets. Once internalized they free the mind all the more to exercise the natural light of reason. Starting with the product, let's ask: is it less creative if

someone produced it by following heuristics rather than by a leap of insight? No—insofar as a philosophical product is creative to the extent that it is novel and valuable, it does not matter how it was produced. And if a heuristic that previously was used consciously is internalized and now operates implicitly, then the upshot of its operation may well count as a leap of insight in any case.

But suppose I concede for the sake of the argument that philosophical creativity requires certain mental processes rather than others, *and* that employing my heuristics thwarts those certain processes (two big concessions!). Then I am left wondering why we should care about creativity, so understood. Give me a less creative (so understood) but superior product over a more creative but inferior product, any day. If the heuristics stifle creativity (so understood) but promote first-rate products, more power to them, I say. Whoever thought that creativity was the only thing, or even the main thing, that we value? We should value philosophical understanding and philosophical progress, however they are achieved.

By the way, the answer to the trivia question is *Mexico and Luxembourg*.

****

This is just a start. As I have said, I have hundreds more heuristics, of varying degrees of usefulness. I invite you to come up with some of your own; this may prove to be a somewhat creative enterprise in itself! If you care to share them with me, I would be glad to hear about them. Or perhaps you will find a way to use some of my heuristics against me. There—I even managed to finish with a gesture at one of them: self-reference.

Go forth and be creative![12]

REFERENCES

Adams, Robert Merrihew (1979), 'Primitive Thisness and Primitive Identity', *Journal of Philosophy*, 76: 5–26.

Amabile, Teresa M. (1996), *Creativity in Context* (Boulder, CO: Westview Press).

Armstrong, D.M. (1989), *Universals: An Opinionated Introduction* (Boulder, CO: Westview Press).

Baldwin, Thomas (1996), 'There Might Be Nothing', *Analysis*, 56: 231-38.

Baumeister, Roy, Brandon J. Schmeichel, and C. Nathan DeWall (this volume), "Creativity and Consciousness: Evidence from Psychology Experiments".

Bigelow, John and Robert Pargetter (1987), 'Beyond the Blank Stare', *Theoria*, 53: 97-114.

Block, Ned (1981), 'Psychologism and Behaviorism', *The Philosophical Review*, 90(1): 5-43.

Boden, Margaret (2004), *The Creative Mind: Myths and Mechanisms*. 2nd ed. (London: Routledge).

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

Church, Alonzo (1940), 'On the Concept of a Random Sequence', *Bulletin of the American Mathematical Society*, 46: 130-135.

Descartes, René (1637), *Discourse on Method*; Laurence J. Lafleur (trans.) (1960), *Discourse on Method and Meditations* (New York: The Liberal Arts Press).

Field, Hartry (1980), *Science Without Numbers* (Princeton: Princeton University Press).

Fitelson, Branden (2008), "A Decision Procedure for Probability Calculus with Applications", *Review of Symbolic Logic*, 1: 111-125.

Fitelson, Branden (MS): "Gibbard's Collapse Theorem for the Indicative Conditional: An Axiomatic Approach".

Frigg, Roman (2006), 'Scientific Representation and the Semantic View of Theories', *Theoria*, 21(1): 49-65.

Gaut, Berys (2010), 'Philosophy of Creativity', *Philosophy Compass*, 5(1): 1-13.

Howson, Colin and Peter Urbach (2006), *Scientific Reasoning: A Bayesian Approach*, 3rd Ed. (Open Court).

Jackson, Frank (1998), *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Clarendon Press).

Jammer, Max (1974), *The Philosophy of Quantum Mechanics: The Interpretations of Quantum Mechanics in Historical Perspective* (New York: Wiley-Interscience).

Joyce, Richard (2005), 'Moral Fictionalism', in Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics* (Oxford: Clarendon Press): 287-313.

Kim, Jaegwon (1982), 'Psychophysical Supervenience', *Philosophical Studies* 41: 51-70.

Lewis, David (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, 50: 249–58.

Lewis, David (1973), 'Causation', *Journal of Philosophy*, 70: 556–67; also in Lewis (1986a).

Lewis, David (1986a), *On the Plurality of Worlds* (Oxford: Blackwell).

Lewis, David (1986b), *Philosophical Papers*, vol. II (Oxford: Oxford University Press).

Lewis, David (1993), 'Many, but Almost One' in Bacon, John (ed.), *Ontology, Causality and Mind: Essays in Honour of D M Armstrong* (New York: Cambridge University Press).

Lewis, David (1999), *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press).

Lloyd, Elisabeth (1988), *The Structure and Confirmation of Evolutionary Theory* (Princeton: Princeton University Press).

Menzies, Peter (1996), "Probabilistic Causation and the Pre-emption Problem", *Mind*, 104: 85–117.

Norton, John (2003), "Causation as Folk Science", *Philosophers' Imprint* 3, No. 4 (November), www.philosophersimprint.org/003004/.

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

Pollock, J. (1983), 'How Do You Maximize Expectation Value?', *Nous*, 17: 409–421.

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

Quine, W.V.O. (1951), 'Two Dogmas of Empiricism', *The Philosophical Review*, 60: 20-43.

Rosen, Gideon (1990), 'Modal Fictionalism', *Mind*, 99: 327-354.

Sheffer, H. M. (1913), 'A Set of Five Independent Postulates for Boolean Algebras, with Application to Logical Constants', *Transactions of the American Mathematical* Society, 14: 481-488.

Shoemaker, Sydney (1980), 'Causality and Properties', in Peter van Inwagen (ed.), *Time and Cause* (Dordrecht: Reidel).

Smith, Michael (1992), *The Moral Problem* (Oxford: Oxford University Press).

Smullyan, Raymond M. (2000), *Forever Undecided: A Puzzle Guide to Gödel*, Oxford: Oxford University Press.

Sorensen, Roy (1996), 'Modal Bloopers: Why Believable Impossibilities are Necessary', *American Philosophical Quarterly* 33(1): 247-261.

Sorensen, Roy (2000), 'Faking Munchausen's Syndrome', *Analysis* 60(2): 202-209.

Stove, David (1991), *The Plato Cult and Other Philosophical Follies* (Oxford: Blackwell).

Thomasson, Amie (1999), *Fiction and Metaphysics* (Cambridge: Cambridge University Press).

Unger, Peter (1980), 'The Problem of the Many', *Midwest Studies in Philosophy*, 5: 411-468.

van Fraassen, Bas (1980), *The Scientific Image* (Oxford: Oxford University Press).

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

Ward, Thomas B. and Yuliya Kolomyts (2010), 'Cognition and Creativity', in James C. Kaufman and Robert J. Sternberg (eds.), *The Cambridge Handbook of Creativity* (Cambridge: Cambridge University Press).