Thinking Off Your Feet
Reply to My Critics

Michael Strevens

Analysis, 82, 343–353, 2022

Many thanks to my critics for their thoughtful comments.

1. Vindicating Analysis

All three critics ask, in varying ways, how far my "vindication" of analysis really goes.

Andow frames the project of vindication in rather specific terms: "What seems to need vindicating is a model of philosophy in which armchair philosophy is dominant and in which only around 10% is experimental rather than around 30%". He observes that TOYF does not come close to accomplishing this task, saying that the book might in fact be used to "build a case in favour of the more-experimental-philosophy and less-philosophical-analysis agenda". In a similar vein, Love wonders what TOYF's attempted vindication means for philosophical pursuits besides analysis, such as explication or conceptual engineering.

The answer might be discerned between the covers of TOYF if it is treated not as a sequence of claims about, but as an example of, philosophical inquiry. Its author appears to be a thorough-going methodological pluralist in philosophy, and one for whom analysis is by no means the dominant or most favored approach. As someone with personal knowledge of his proclivities, I can confirm that speculation.

At the same time, TOYF contains one significant piece of analysis, that of the reference or extension of a concept. It spans only a single chapter, but a crucial one. If TOYF is to succeed in its aims, its claims about reference had better be on the right track. In this respect the project undertaken in TOYF resembles a great deal of empirically informed philosophy: though largely a posteriori in character, it turns at the same time on certain critical armchair considerations. Further, as Andow correctly observes, much experimental philosophy in the "positive" vein is predicated on similar, and similarly puzzling, assumptions as philosophy of the armchair variety, most notably that case judgments made on an obscure or opaque basis can serve as reliable evidence for and against philosophical hypotheses.

These remarks, I hope, help to frame the attitude that I bring to the project of vindication. Without having to agree on quotas or to build a methodological consensus, a broad range of philosophers ought to be able to agree that a lot turns on the validity or otherwise of the classic armchair method. The aim of my book is to explore a strategy for understanding how it might work.

The term "vindication" is perhaps best known to philosophers from Hans Reichenbach's claim to have vindicated inductive reasoning. TOYF does not suggest that we can hope for more assurance from a vindication of armchair philosophy than we could expect from a vindication of scientific inquiry. At best, we will get a reasonable expectation, if certain skeptical scenarios are not realized, that we will eventually get to the truth. When Machery writes that "It is hard to see how [TOYF's] long-term perspective will bring comfort to the armchair philosopher", then, I think he is asking for too much. Newton did not live to see the perfection and then the overthrow of his theory of gravity, and likewise, the pioneers of the analysis of causation predeceased the development of sophisticated regularity theories, the apparatus of interventionism, and so on. If there is still more work to do, I think that is a calamity for neither scientific nor philosophical inquiry.

2. Case Certainty

Case certainty, as I wrote in the précis, is my name for the intimation that a case judgment is sure to be correct, or at least, that it is highly likely to be correct and not subject to further revision.

"To what degree should case certainty be such a significant, bedrock feature of philosophical inquiry?", asks Love. That runs together two quite separate questions. The first is a psychological question: to what extent does the impression of case certainty pervade armchair analysis, and why? That is a matter of psychology. The second is a normative question: to what extent is the impression correct, and consequently, to what extent should we allow it to figure into our philosophical deliberation? Simply to take the first question seriously, as I do, is not to advocate the supremacy of the intuitions. Yet Love goes on to characterize my concern with case certainty as "dogmatism".

I understand why an experimentally-minded philosopher might take the paradigm armchair philosopher's confidence in their intuitions as a form of dogmatism, and seek to dismiss it out of hand. My response is different: I want to understand it. I approach this task without an agenda. It might well be that case certainty is little more than an expression of institutional complacency. But to suppose this to be the case without any further investigation is in itself, I think, rather dogmatic. In any case, the psychology of philosophy had better have something to say about it, if it aims to paint a complete picture of the armchair enterprise.

Let me emphasize that case certainty is, for my project, more a bane than a boon, as it poses a gnarly problem for my inductivist psychology. If our judgments about cases are delivered by a corrigible theory, why would we take them to be definitive? I have nothing to gain by taking case certainty seriously; on the contrary, at critical junctures of TOYF, it constitutes one of my greatest challenges. The project would unfold far more smoothly if it turned out to be a sociological artifact.

Indeed, the situation is, as I say in a passage quoted but then effectively ig-

nored by Love, puzzling no matter what your attitude to armchair philosophy. For those who want to take epistemic succor from case certainty, the variability in case judgments that experimental philosophers have found among the wider population is (or should be) unsettling, even undermining. But that same variability makes the attitude of certainty itself into something of a mystery, the more so when it persists even once the variability is sincerely acknowledged. Yes, it might all be false consciousness. But then again, it might be an opportunity to learn something truly interesting about the psychology of philosophical analysis.

3. Explanatory Convergence

TOYF's argument that our philosophical concepts pick out substantial categories of objective interest, rather than a hodgepodge of things dictated by the thinker's ephemeral or idiosyncratic interests, turns on the idea that the concepts in question are theories that have an inductive tendency to converge on explanatory structures. Our concepts, that is to say, consist of hypotheses about the explanatory connections between certain things (such as causal connections in the case of material things); as further information arrives, or as reflection proceeds, they will tend to reflect that structure more accurately. What we end up with when inquiry is done will therefore tend to be a map of some piece of "local explanatory reality".

Love is skeptical, writing that "Strevens presumes an ever-encroaching scientific image that converges on objective, true fundamental structures" and "Strevens thinks concepts are governed by the almighty hand of Science in its imperial march to explain everything". I find these remarks puzzling.

Suppose we have, as a part of our concept of swan, a causal hypothesis of the form "Something about swans causes them to grow white feathers". We then travel to Australasia, where we discover birds that we classify as swans on the basis of our other swan beliefs, but which have black feathers. We say to ourselves: that whiteness hypothesis needs to be revised. We replace it with

(perhaps) "Something about swans causes them to grow either white or black feathers, depending on the species". Or suppose that, uncertain about swans' natural diet, we carefully observe the eating habits of several different species. We find that they consume various kinds of underwater plants. So we add to our "theory" of swans a new hypothesis: swans are herbivorous.[1]

I suppose you might describe such workaday inductive inferences as building "a scientific image" that "converges on objective, true fundamental structures". It seems rather more strained to say that the hypotheses making up the swan concept are "governed by the almighty hand of Science in its imperial march to explain everything", which suggests a kind of metaphysical grandiosity teetering on the brink of absurdity, when in fact what is happening is some modest and unremarkable learning from experience. It is no more than this that I have in mind when I say that our concepts converge on local explanatory reality.

Perhaps Love is moved by the antirealist movement in the philosophy of science. Home base for scientific antirealism is fundamental physics, where it is indeed possible to reasonably wonder whether we will ever get to the bottom of gravity, matter, fields, and spacetime. Even antirealists, though, concede that there's not much doubt about the existence of bacteria, viruses, DNA, H_2O molecules, and so on, along with the causal mechanisms by which they interact with the world around them (Stanford 2006). The convergence arguments in TOYF are largely pitched at this level, a level at which surely only a trenchant skeptic would deny the attainability of causal knowledge.

In any case, throughout TOYF I am quite happy to rest with the thesis that philosophical analysis is no less reliable or parochial than empirical inquiry; if empirical inquiry has limits that I have not anticipated, then I will happily accept that philosophical analysis must be limited in the same way.

1. In my view, such generic hypotheses amount to causal claims: something about swans causes them to feed (almost) exclusively on plant matter. But it is enough for my purposes here that they are explanatory in one way or another, or just that they reflect a certain objective fact about swans.

I might add that conceptual convergence is not solely within the purview of science, capitalized or otherwise. Our moral concepts, I speculate, consist of theories that seek to delineate a certain explanatory structure in the moral order. The explanatory links might very well not be causal, but rather something non-natural that exerts normative rather than nomological force. (The question whether we should believe in such things is quite separate; TOYF is not in the business of laying down an inventory of all of reality.) Although certain kinds of empirical facts may be relevant to a non-natural moral theory—think of the considerations in favor of "expanding the circle" to include, say, certain animals as morally considerable beings—the process is hardly one of scientific inquiry. Yet in its own way, it holds out the promise that our concepts will gravitate toward features of explanatory interest, rather than imposing on the terrain a mistaken or arbitrary taxonomy.

4. Conceptual Fragmentation

Machery and I agree that it is important, in order to evaluate the prospects of philosophical analysis, to understand the psychological sources of the variability in the case judgments that drive analysis. There can be little doubt that such variability exists. Asked whether Gettierized belief is knowledge, for example, or whether Twin Earth's X_yZ is water, respondents are anything but univocal in their answers. The crucial question is whether this diversity is a sign of a pervasive inaccuracy that would undermine analysis. (In what follows, I focus on concerns about reliability and put aside the problem of parochialism.)

Machery's preferred explanation of case judgment variability is a kind of fragmentation in philosophical concepts that he compares to polysemy in natural language. In the mental equivalent of polysemy, a single mental representation—let's say for convenience a category term in the language of thought—picks out any one of several categories, depending on the inferential context. To use Machery's example, in one context the mental equivalent of

the word "book" might pick out the physical object, pages and all, and in another it might pick out the content in a more abstract sense.

Polysemy in natural language is not a great challenge to communication; we are typically able to figure out what is meant by any particular use of the word "book", and we don't regard Machery's example sentences—"*Thinking Off Your Feet* weighs 150g" and "Taking *Thinking Off Your Feet* with me for vacation didn't make my carry-on heavier because I downloaded it"—as constituting a contradiction. Likewise, if there is polysemy in the mind—if there are mental terms with several related but distinct meanings—we presumably manage our thought processes with the same level of competence. I do not infer, from my belief that I have TOYF with me on holiday, that my carry-on contains something that I can use to prop up my laptop. More generally, I do not equivocate, allowing polysemous terms to switch meanings in the middle of an inferential chain. And indeed, if equivocation were at all common, our lives would surely be a cognitive shambles. Human reasoning has many failings, but this sort of error does not seem to be among them.

Or at least that is true when the meanings in question are quite distinct, as in the example of "book". Equivocation might not be so disastrous, or even at all obvious, if it were between almost identical meanings, largely overlapping in the things they pick out. Machery, as I understand him, has this sort of possibility in mind, proposing the following picture of thought. Mental category terms, he suspects, typically have many closely related meanings, picking out a cluster of largely overlapping categories. An item that falls under one of these categories, then, will tend to fall under them all. Consequently, equivocation in practical reasoning will not substantially undermine the reliability of our inferences, at least in extensional contexts. Confusion will be confined to a few edge cases.

Edge cases, however, are what philosophical analysis is all about. It is in that apparently safest of all items of furniture, then—the armchair—that our equivocation will endanger us most. Whether some strange substance in an

equally strange scenario counts as water will very much depend on whether we are thinking about $water_1$ or $water_2$, and in Machery's picture of the mind, which of the two we are thinking about is continually getting away from us.

The same kind of story can be told without invoking the semantic notion of polysemy. Even if our mental term "water" has a univocal meaning, suggests Machery (or so I read him), its inferential role might be perpetually contested by rival sets of beliefs (either temporary coalitions thrown together by inferential context, or more long-lasting clusters corresponding to what you might think of as distinct "conceptions" of water that co-exist in the mind). The effect is very much the same as in the story told in terms of polysemy and equivocation. If the competing sets of beliefs were sufficiently different, our minds would be thrown into utter chaos. But if the beliefs agree on almost, but not quite everything—diverging, say, on the question whether X_yZ counts as water—then for the most part we could go about our thinking lives quite unaware that this unending struggle for cognitive supremacy is taking place inside our skulls.

Can this explain the variability found in philosophical case judgments? Without some supplementary hypothesis, I think not. Equivocation, or fragmentation more generally, would explain a pattern of variability in which individuals would be unable to settle clearly on a verdict about a given case. Asked whether a stopped-clock belief (in which a correct belief about the time is inferred from a broken clock) is knowledge, they would first use one set of beliefs to answer the question and then another (perhaps corresponding to different, related meanings of the English word "knowledge"). The beliefs would give different verdicts. The judger, then, would think sometimes yes, sometimes no.

That might account for certain "ordering" and "framing" effects uncovered by experimental philosophers. But professional philosophers, at least, do not appear to have unstable judgments about the cases that have figured most prominently in the last century of analysis. (No doubt it is the very stability

of these judgments that accounts in part for their influence.) What's needed, if Machery and other "negative" experimental philosophers are to undermine our faith in analysis, is some psychological story that gives us reason to doubt that even these stable judgments have zeroed in on the truth. Prima facie, the fragmentation account seems simply not to apply.

Machery might respond as follows. There is, in fact, a considerable degree of internal variability in the way these judgments are made even by a single person, as the fragmentation picture implies. To participate in professional armchair philosophy, however, a thinker must make univocal judgments concerning at least some significant cases. So they arbitrarily select one set of beliefs fighting over the control of the term in question, or one of its many competing meanings (in the polysemy version), and consistently favor it when making the judgment. Uniformity is thereby imposed on their responses. But because of the arbitrariness, they are in effect selecting their chosen answer at random (or perhaps to favor their own agenda).

I reply that this enhanced appeal to fragmentation is still not true to the reality of armchair analysis. It is not merely individual philosophers who have achieved stability in their judgments, but the community of philosophers as a whole. The vast majority of philosophers make the same judgments about the most central Gettier cases (and again, this stability is surely in part what makes for a central case). It is natural, then, to augment the fragmentation story with one further idea: philosophers have coordinated their arbitrary choices. They have found a way to single out a particular meaning for "knowledge", or a particular set of categorization-guiding beliefs about knowledge, and to establish that as the institutional norm.

If so, however, then the arbitrariness of the choice seems far less threatening to the prospects of analysis. In the polysemy version of the story, philosophers have selected a single meaning for the term "knowledge", and they have proceeded to judge whether stopped-clock beliefs, fake-barn beliefs, and so on are instances of knowledge in that particular, quite definite sense—then us-

ing those judgments to test various hypotheses about the nature of knowledge in precisely the same sense.

There is no longer any obvious source of error. The disparity between the philosophical near-consensus and the diverse case judgments of ordinary people turns out to be a consequence, not of some factor that persistently throws such judgments off course, but rather of the fact that the philosophers have successfully fixed on a determinate subject matter while ordinary people have not.

(You might still worry, as does Machery, about parochialism—that philosophers' chosen precisification of knowledge is a poor choice of research topic—but that is not the problem I am tackling here. In any case, to pursue this worry you would need a far better sense of the way in which these choices are made than Machery, on behalf of the fragmentation view, has given us.)

The situation is not so clear in the version where there is no polysemy but rather where the philosophical profession has privileged a particular set of beliefs about knowledge as determiners of their collective case judgments. Let me first remark that this is not fragmentation as discussed by David Lewis in the passage quoted by Machery and as investigated by many contemporary epistemologists. Those philosophers' concern is a situation where different sets of beliefs take over in different circumstances. What I am now discussing is a case where a community of inquirers as a whole has overcome fragmentation by, as it were, selecting a particular fragment and handing it complete control over the deployment of the term "knowledge" in perpetuity. Could the chosen beliefs somehow fall badly short when acquitting the responsibility foisted upon them—too sparse, too lop-sided, too wrong?

It surely makes more sense to interpret what philosophers have done, in singling out this set of beliefs as the arbiters of "knowledge", as more or less identical to what the polysemous philosophers have done: they have pinpointed a very particular conception of knowledge and made that the target of their investigation. (Certainly, this would be the verdict rendered by

my own theory of reference.) As in the case of polysemous fragmentation, the existence of variability in case judgments made by "the folk" turns out to be irrelevant to assessing the prospect of this exercise, because the judgments of the general population have a substantially different subject matter (or rather, matters).

I conclude that Machery's fragmentation explanation of variability is in fact far more helpful to the proponents of philosophical analysis than to its enemies. My own inductivist explanation of variability, as we will see next, offers greater hope to the skeptic—hope that, of course, I intend to dash.

5. Philosophers as Experts

According to the inductivist picture, categorization under a concept is typically guided by something like a theory—a cluster of beliefs about the place of the corresponding category in the explanatory order. Consequently, categorization often has the aspect of something like inference to the best explanation: from various observable properties of a specimen, the categorizer concludes (perhaps rather tentatively) that it belongs to a certain category, membership of which would explain some or all of those properties. In concrete terms, using my standard example, when we infer that a certain bird is a swan, we might do so because we suppose that its being a swan is the best available (and is a good enough) explanation of its white feathers, cygnid shape, trumpeting call, and so on.[2]

If categorization works the same way for categories of philosophical interest such as knowledge, then many of our case judgments, and in particular those based on the kinds of specific details laid out in a typical thought experiment, are also explanatory inference of a sort. Thus, the thought processes by which case judgments are made traverse complex explanatory structures

2. We might also conclude that it is a swan because an expert tells us so; categorization can in principle be based on any information relevant to category membership. But I focus here on the variety relevant to philosophical case judgments.

by following inductive links.[3]

That immediately suggests some possible mechanisms for variability in case judgments even under circumstances where judgers share more or less the same relevant beliefs, that is, when they have roughly the same "theory of knowledge". First, as with any complex inference, some inferrers may go wrong: they may fail to notice that a certain condition that must be satisfied for the inference to go through does not in fact obtain. Second, as with any inductive inference (or so most of us would say), there is a certain latitude in interpreting the strength of the evidence. One inferrer may hold that the relevant information is good enough to license an ascription of knowledge; another equally reasonable inferrer may be less sure.

Such possibilities suggest that our categorizations are in principle vulnerable to intellectual vagaries that the enterprise of philosophical analysis seems not to countenance. By contrast with the fragmentation scenarios discussed in the previous section, here the inference genuinely is being either knocked off course (by error) or defocused (by variability in inferential style) to the point that the truth is no longer clearly and accurately illuminated. Further, the potential for error might well be at its greatest in the kinds of complex, finely calibrated edge cases that serve as crucial thought experiments in philosophical analysis—to the point where a cautious philosopher might suppose that judgments about such cases simply ought not to be relied on in serious intellectual inquiry.

Let me focus, for simplicity's sake, on the starker problem of outright error. Both Andow and Machery emphasize this ominous possibility, and rightly so. A vindicator of analysis had better respond to the challenge. My favored strategy, as Machery notes, is very broadly what is called the "expertise defense". Philosophers, I hold, are simply less likely to make inferential errors.

3. My best guess as to the way in which a category such as knowledge is embedded in these structures attributes to it a role that is rather different from swanhood in the typical "theory of swans". That guess is outlined in TOYF, §13.4. In this short reply, I will write at a level of generality that applies to any kind of explanatory embedding.

That might appear less than fully relevant to the question at hand. As Andow notes, philosophers surely are better than most laypeople at "explicit, deliberate inductive reasoning". But philosophical analysis's case judgments do not fit this description. We are typically not aware of the course of our reasoning when we classify a stopped-clock belief as non-knowledge, for example. If the inductivist picture is correct, then we are deliberating, but we evidently have no conscious control over our deliberation, and therefore no avenue to bring our hard-won logical prowess to bear.

Perhaps our expertise is anyway manifested in these reasoning processes, without any effort on our part. Another possibility is that what we contribute consciously when making case judgments is not so much elegant thinking as a certain level of commitment and attention. What distinguishes the philosopher from the layperson, in other words, is the additional motivation they typically bring to a philosophical thought experiment, focusing intently on the relevant aspects of the case, resisting the urge to rush to a conclusion, backing up if something feels a little off.

It may strike you as odd to say that we can control these aspects of a thought process whose contents are largely opaque to us, but (if I may indulge in some armchair psychology) it seems quite correct to me. Here, my own experience with the famous fake barn case inspires me. For many years I regarded the case as a toss-up, an example of an "intuition" that could go either way.[4] At a certain point, however, the case began to matter more to my professional endeavors (though I still had no vested interest in the outcome). I thought harder about the scenario—without knowing exactly what I was thinking. And I came to realize that the belief in the thought experiment is indeed not knowledge. I was not guiding my thinking, but I was creating an environment in which it proceeded with more care and attention than it had

4. I had not grasped that there is an official consensus that true ascriptions of barnhood in fake-barn country are not knowledge, though otherwise I was apparently in good company, as some interesting experimental philosophy has shown that many philosophers fail to arrive at the official conclusion (Horvath and Wiegmann 2016).

been previously allotted.

That, I suggest, is the key to the difference between professional and lay judgments in these cases. It is not talent but effort, not expertise but dedication, that allows a small cadre of professional thinkers to converge on the truth while others around them fumble and confuse the issues. Or at least, that is how it goes under ideal conditions—a certain degree of quiet and a very comfortable armchair.

Andow, considering the possibility that professional philosophers' judgments are "less noisy" than ordinary people's—more or less what I have proposed here—suggests that the solution to the problem of noise is to increase the sample size, by which he means, to add more people's case judgments to the data set. That is not quite the right analogy, I think. Individual thinkers are like measuring instruments: some are more accurate, some more affected by noise. The correct procedure is to use the more accurate instruments, when available, to obtain all the points in the data set.

Machery's response to the kind of move I've made is rather different: he suggests that convergence of opinion among professional philosophers is due to "conformism and the requirement to pay to play". It's worth bearing in mind, however, that precisely because the thought processes involved in making case judgments are outside our control, it is far from clear that even trucklers and mercenaries have the capacity to tailor their judgments to suit their self-interest. And among the rest of us, even those of very modest virtue have surely had the experience of arriving at a case judgment of considerable professional inconvenience. Still, there is of course such a thing as unconscious bias, and the question of whether, or to what degree, peer pressure steers case judgments might usefully be further researched. Let the psychology of philosophy thrive!

References

Horvath, J. and A. Wiegmann. (2016). Intuitive expertise and intuitions about knowledge. *Philosophical Studies* 173:2701–2726.

Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press, Oxford.