Leuenberger, S. (2009) *Review of Stephen Yablo, Thoughts: Papers on Mind, Meaning, and Modality (Philosophical Papers, Volume 1).* Notre Dame Philosophical Reviews. ISSN 1538-1617

http://eprints.gla.ac.uk/69878/

Deposited on: 24 September 2012

Stephen Yablo, *Thoughts. Papers on Mind, Meaning, and Modality (Philosophical papers*, vol. 1), Oxford University Press, 2008, 327pp, $ 36 (pbk.), ISBN 978-0199266470.

Reviewed by Stephan Leuenberger

This book collects together eleven rich and rewarding articles by Stephen Yablo. All but one have been published previously, and some of them are already classics. Six essays are concerned with modal epistemology, and four with mental causation. In this review, I will focus on the former group, and offer only a brief summary of the latter. I will end by discussing whether there are irresolvable tensions among the papers.

"How in the World" (ch. 7) is the only paper to fall outside the two main clusters of topics. It provides an interesting development of the idea that talk of possible worlds is metaphorical. Yablo observes that metaphors, despite not being literally true, often help us convey factual information. Metaphorical talk of possible worlds allows us to communicate about modal facts. While Yablo's view is fictionalist about possible worlds, it is not thereby fictionalist about modality, as he insists: it does not entail that modal facts are somehow constituted by a many-worlds fiction.

The papers on modal epistemology are "The Real Distinction between Mind and Body" (ch. 1), "Is Conceivability a   Guide to Possibility?" (ch. 2), "Textbook Kripkeanism and the Open Texture of Concepts" (ch. 3), "Coulda, Woulda, Shoulda" (ch. 4) , "No Fool's Cold: Notes on Illusions of Possibility" (ch. 5) and "Beyond Rigidification: The Importance of Being *Really* Actual" (ch. 6). The last-mentioned paper, whose title echoes both Scott Soames and Oscar Wilde, is the only one that appears for the first time in the volume. It partly overlaps with "No Fools's Cold", and partly extends the arguments of that paper.

The main question with which these essays are concerned is how we can have epistemic access to modal facts, given that necessity is distinct from apriority, and possibility is distinct from conceivability. One of Yablo's positive aims is to rehabilitate conceivability as a source of evidence for possibility in the wake of Kripkean examples of the necessary a posteriori. His main negative aim is to bring out ways in which the two-dimensionalist's account of the relation between conceivability and possibility, as develped in most detail by David Chalmers, is unsatisfactory.

We seem to be able to conceive that Hesperus is in the sky but Phophorus is not; that some homogeneous substance, without any atomic or molecular structure, is hot; that there is a necessary God; and that there is no God. Conceiving presents its object as possible, according to Yablo and many others. But, it is widely held, there are compelling arguments against the possibility of the above propositions. In the case of Hesperus and Phosphorus, the argument appeals to an astronomical discovery and the necessity of identity, and in the case of heat, to a physical discovery and certain ideas about what sort of thing heat is. The necessity and the non-existence of God cannot both be possible, if certain widely accepted principles of propositional modal logic are sound. Our faculty of conceiving seems to present impossibilities as possible.

One reaction would be to deny that conceivability is evidence for possibility at all. But Yablo insists that this is not an option, unless we are seriously revisionary about our epistemic practices. Conceivability is evidence for possibility, even though it is evidence that can be overruled. Another reaction would be to adopt a *profiling* strategy: to identify types of propositions with respect to which conceivability evidence should be treated with suspicion. One could do this by cataloguing arguments that have overruled conceivability evidence in particular cases, and analysing what features make a proposition susceptible to such arguments. In the above cases, these features might be identified as the following: being an identity claim involving rigid designators; bearing on the constitution of natural kinds; and featuring a modal operator. If the catalogue is complete and the analysis sound, the claim that non-profiled conceivable propositions are possible is safe from refutation. But the profiling strategy, unless supplemented, does not provide a satisfactory response to worries about the evidential value of conceivability. In practice, we are not sure whether we have identified all kinds of overruling arguments. Even if we were, it is not clear what reason we would have to be confident that non-profiled conceivable propositions are possible. For the strategy does not by itself reveal a difference in the conceivability evidence between the profiled and the non-profiled cases.

Yablo's own strategy, broadly, is to show that the propositions shown to be impossible are not conceivable, contrary to appearances, such that they do not disconfirm the claim that what is conceivable is possible. While trying to implement this strategy, he discusses in detail what is involved in conceiving a proposition: what the respective contributions of imagination and of background beliefs are, what features of the actual world are held fixed in conceiving counterfactual situations, and the like. Yablo is ready to ascribe introspective error: sometimes we think that we conceive something, but don't. But it is distinctive about his approach that he is very careful and scrupulous about ascribing such error. He does it only on a case-by-case basis, eschewing sweeping generalizations. Moreover, he imposes a very stringent constraint (what he calls the "psycho-analytic standard") on when the modal epistemologist may ascribe introspective error: when the conceiver herself is able to come to the realization that she has committed that error.

To the extent that Yablo's positive project succeeds, we can be confident that "carefully handled, conceivability evidence can be trusted, for when impossible $E$ seems possible, that will generally be because of distorting factors that we can discover and control for" (p. 169).[1] To what extent does his project succeed? He does not pretend to have identified distorting factors in all cases where apparent conceivability evidence is overruled. For example, he does not claim to offer an explanation of our error in the case of heat, nor in the cases of the necessity or the non-existence of God. The value of his discussion is not so much in the conclusion he reaches, but in how it increases our understanding of "how we can discover and control for" errors in our conceiving.

Yablo has been one of the most influential commentators on semantic

---

[1] Of course, we would not have defeated a skeptic who denies that there is any reason to take conceivability as evidence for possibility. But defeating such a skeptic is not Yablo's project.

two-dimensionalism. Recent struggles between proponents and opponents of that view have occasionally been perplexing to thrill-seeking onlookers. Two-dimensionalists did not even flinch under seemingly heavy blows. Either they had swiftly moved out of reach, or they had not been where their attackers thought they were.[2] In contrast to other critics, Yablo does not take a pugilist approach. Rather than aiming at a refutation, he tries to undermine the motivation for it by questioning its explanatory potential. Consequently, he is more concerned with its broad strategy than with the details of a particular implementation.

Ignoring the fine print, Yablo takes the two-dimensionalist idea to be that if a proposition P is conceivable but impossible, then the "shifty" cousin P* of P is possible. We can think of P* as the proposition expressed by the sentence that results from deleting all occurrences of `actually' in a sentence canonically expressing P. In "Textbook Kripkeanism", Yablo tries to reconstruct a "transcendental" argument for that claim, one that proceeds from reflection on the conditions of the possibility of thought and language. Since he sees no prospect in such an argument, he asks in subsequent papers whether the two-dimensionalist claim can be supported by an inference to the best explanation.

In "No Fool's Cold" and "Beyond Rigidification", he presents intriguing cases of modal error that standard two-dimensionalist resources cannot explain. We may call these cases of the "a posteriori necessary a priori", because they involve a truth that, for all we know a priori, is an example of the contingent a priori, but turns out to be necessary as well as a priori. If I wrongly believe that there are ivory-billed woodpeckers, it will seem to me that it could have been the case that there are fewer of them than there actually are. If I am ignorant about how gold is constituted, it may seem to me that it could have been the case that it is constituted differently from how it actually is. These are modal errors, according to Yablo, because what seems possible is not. He notes that the two-dimensionalist cannot explain those errors. For deleting `actually' in `There are fewer woodpeckers than there actually are.' or `Gold is constituted differently from how it is actually constituted.' results in a contradictory sentence. Whatever error we are committing in this case, it is not one of elementary logic.

None of this refutes two-dimensionalism as officially presented, say by David Chalmers, and Yablo does not claim otherwise. No pre-theoretical, ordinary notion of conceivability or apparent possibility appears in those presentations, but only a technical surrogate that applies to (mental or linguistic) tokens, which we might call conceivability*. Chalmers claims that if a sentence token is conceivable*, then the primary intension associated with it is possible. That claim is not threatened by

---

[2] Two-dimensionalism is presented by David Chalmers in *The Conscious Mind. In Search for a Fundamental Theory*, Oxford University Press, 1996, in "Does Conceivability Entail Possibility", in Tamar Gendler Szabo and John Hawthorne (ed.), *Conceivability and Possibility*, Oxford University Press 2002, "The Two-Dimensional Argument against Materialism," in Brian McLaughlin (ed.), *The Oxford Handbook of the Philosophy of Mind*, Oxford University Press, forthcoming, among other publications. Two-dimensionalism is by criticized by many philosophers in numerous articles, and at book-length in Scott Soames, *Reference and Description. The Case against Two-Dimensionalism,* Princeton University Press, 2007.

Yablo's examples. But the examples show that some propositions may be conceivable without there being corresponding conceivable* sentence tokens. Two-dimensionalism could be supplemented to give a different account of those other illusions. But even if it were, this may not result in a unified account of modal error.

What is at stake for those of us who are not primarily interested in an explanation of modal error, but merely in a recipe for inferring true possibility claims? As just noted, we have not been given any counterexamples to the claim that what is conceivable* is possible. But the absence of counterexamples confirms the claim only if conceiving* is a natural mental kind --- if it does not result from imposing an ad hoc constraint on conceivability designed to immunize it from certain objections.[3] Otherwise, the two-dimensional strategy faces a similar epistemic problem as the "profiling" strategy mentioned earlier. It is an interesting question whether conceiving* is suitably natural, but it is not a question I can take up here.

Besides conceivability and modal error, the other major topic in this volume is mental causation. According to so-called "exclusion arguments", certain classes of properties or events must be epiphenomenal, since other properties or events are already doing the causal work that we might otherwise attribute to them. Yablo's response to such arguments has been highly influential. In "Mental Causation" (ch. 8), he discusses exclusion arguments against multiply realizable mental properties and events; in "Singling out Properties" (ch. 9), against secondary qualities; and in "Wide Causation" (ch. 10) and "Causal Relevance: Mental, Moral, and Epistemic" (ch. 11) against intentional states that do not supevene on brain properties.

The exclusion argument relies on causal incompatibilism, according to which one event cannot have two distinct causes. Typically, philosophers who try to analyse causation in terms of counterfactuals are causal compatibilists, while those who construe causation as production, as "oomphy", are incompatiblists. Yablo develops a counterfactual theory of causation, yet he is not a compatibilists. He acknowledges causal competition, but thinks that the outcome of the competition is different from what is supposed by proponents of exclusion arguments.

In Yablo's main example, a pigeon called "Sophie" is conditioned to peck at red things. When a scarlet triangle is presented, she pecks. Yablo denies that the property of being red is "excluded" from causal relevance by its determinate. Departing from compatibilism, he also claims that the property of being scarlet is not causally relevant, since it includes too much unneeded detail. It fails to be "proportional" to the effect, as he sometimes puts it.

Yablo's theory changes between papers, and so does his terminology. I will partly use my own terms when outlining the main elements of his theory. The basic thought is that for a given effect, there is a competition between candidate causes about which one gets to be the cause. (Alternatively, the theory could be presented as talking about

---

[3]  The two-dimensionalist might think that any quasi-inductive confirmation is dispensable, since she has a transcendental argument for the claim that what is conceivable* is possible. But those of us who reject such an argument, for the reasons Yablo gives in "Textbook Kripkeanism" or for other reasons, the quasi-inductive reasons have to suffice.

properties rather than events.) A candidate cause is an actually occurring event on which the effect counterfactually depends:

(1) $C$ is a candidate cause for $E$ iff $E$ counterfactually depends on $C$.[4]

A candidate cause is a cause if it wins the competition over its rival candidate causes. To win, it needs to exclude them all, in a technical sense to be specified.

(2) $C$ is a cause of $E$ iff $C$ excludes all other candidate causes for $E$.

In his account of exclusion, Yablo uses the notion of "screening off", which he defines as follows:

Definition: *C screens off C' from E* iff $E$ would still have occurred iff $C$ had occurred without $C'$.

A first shot at a theory of exclusion would be the following (this is in effect Yablo's "superproportionality"constraint on p. 297, which he does not endorse):

(3) $C$ excludes $C'$ with respect to $E$ iff either $C'$ is screened off from $E$ by $C$, or fails to screen off $C$ from $E$.

It follows from (2) and (3) that one event can have at most one cause. But together with reasonable further assumptions, they also imply that many events have no causes. Their combination is far too stringent.

Yablo's refined account of exclusion invokes the relation of determinables to determinates, which he extends from properties to events:

(3') $C$ excludes $C'$ with respect to $E$ iff either $C'$ is a determinate of $C$, and is s screened off from $E$ by $C$, or $C'$ is determinable of $C$, and fails to screen off $C$ from $E$.

This is intended to capture the idea that nothing counts as a cause if either it includes unneeded detail - as it does when it is screened off by one of its determinables – or if it does not include all the detail that is needed – as it does when it fails to screen off one of its determinates. This theory needs to be supplemented by an account of the relation between determinables and determinates. Unfortunately, Yablo only provides rudiments of such an account.[5] I will briefly return to this issue below, when discussing the question whether Yablo's views are consistent with each other.

A regrettable feature of the volume is the lack of an introduction. In the Preface, Yablo attributes this to his failure to identify a unified worldview into which all the

---

[4] This runs into well-known problem with preemption, which plague any counterfactual account. Yablo discusses these in "De Facto Dependence", *The Journal of Philosophy 99* (2002), pp. 130-148.

[5] In "Causal Relevance", Yablo invokes relative naturalness rather than the relation of determination.

essays fit, or even to show that they are consistent with each other. But reconciling the papers, or integrating them into a system, are not the only purposes an introduction might serve. It could be useful as a guide to the reader. If the eleven collected articles had eleven different topics, this might not be needed. But since many papers are on related topics, some of the later papers overlap with, refine, or modify the earlier ones. Some make a fresh start entirely. But it often takes time and effort on the part of the reader to figure out which one of those relations obtains between different papers.

But increased reader friendliness would not have been the only gain to be had from an introduction. I would have been interested to learn why exactly Yablo thinks the papers are not consistent with each other. He writes that before giving up, he "was going to explain why, if conceivability is such a good guide to possibility, the conceivability of zombies doesn't refute supervenience; and how, if disembodied pain is possible, pain can be a determinable of its physical underpinnings; and more of the same general sort." It seems to me that the first pair of views can be reconciled happily, and that the second pair is at least not obviously in tension.

While Yablo takes conceivability evidence seriously, he does not think that it refutes physicalism. This gives rise to the first apparent tension he mentions. A zombie, on the usual understanding, is a physical duplicate of you or me that lacks phenomenal consciousness. On some ways of understanding `supervenience', the possibility of zombies is incompatible with the supervenience of the phenomenal on the physical. But if supervenience is understood in the way recommended by David Lewis, then the two are compatible, as pointed out by John Hawthorne.[6] For Lewis, the phenomenal supervenes on the physical - in the sense required for what he calls "minimal materialism" - if among the worlds in the "inner sphere" around the actual world, no two physical duplicate worlds differ phenomenally. The inner sphere is made of those worlds that do not display any fundamental features that are absent from our world. As long as zombies are not in the inner sphere, they do not threaten the supervenience claim. But if physicalism is true, a fundamental alien property, a "blocker", is required to prevent the emergence of consciousness from the physical base. Zombies that are accompanied by blockers will be in the outer sphere. Using this definition of supervenience is not merely a technical trick to save the letter at the expense of the spirit of the claim that the phenomenal supervenes on the physical. The possibility of such blockers should be perfectly acceptable to the physicalist.[7]

However, even a blocker-friendly supervenience claim, such as Lewis's, is potentially vulnerable to conceivability arguments. If not only zombies, but *unblocked zombies* – zombies that inhabit a world without non-physical fundamental features - are conceivable, then there is a conflict between the supervenience of the phenomenal on the physical and the claim that what is conceivable is possible. It is an interesting question what Yablo's account predicts about the conceivability of unblocked

---

[6] John Hawthorne, "Blocking Definitions of Materialism", *Philosophical Studies 110* (2002), pp. 103-113. Lewis's definition is given in "New Work for a Theory of Universals, *Australasian Journal of Philosophy 61* (1983), pp. 343-377.

[7] This claim is defended in Stephan Leuenberger, "*Ceteris Absentibus* Physicalism", in Dean Zimmerman (ed.), *Oxford Studies in Metaphysics*, vol. 4, Oxford University Press, 2008.

zombies. It seems to me that their existence might come out as neither conceivable nor inconceivable, and thus fall in the same category as the negation of Goldbach's conjecture (which Yablo discusses on p. 68). On Yablo's account, a proposition is conceivable for me if I take it to be true in some world that I can imagine, and inconceivable if I take it to be false in every world that I can imagine. Propositions that are neither conceivable nor inconceivable are "undecidable": I do not take them to be true in any world that I can imagine, and in some of those worlds I do not take them to be false either. We can certainly imagine a world in which a disproof of Goldbach's conjecture is announced. But, Yablo suggests, we do not (or should not) believe about that world that it is a world where the conjecture is false, nor that it is a world where the conjecture is true. For none of our imaginings is true only in a world where a *correct* disproof of Goldbach's conjecture is produced. One could argue that the same applies, mutatis mutandis, to the case of zombies: just as our imagining is silent about whether the proof is correct, i.e. without error, our imagining is silent about whether the zombies are unblocked.

The second potential incompatibility arises from Yablo's defence of the possibility of disembodied spirits and his claim that pain has physical determinates. Presumably, a disembodied spirit can have pain, but none of its physical determinates. But it is not clear how we can generate a contradiction from that. I will consider three options.

First, the conflict might arise from Yablo's gloss of the relation of determination: $P$ is a determinate of (the determinable) $Q$  if to be $P$ is to be $Q$, not simpliciter, but in a specific way (p. 226). If $P$ is what physically underpins my pain, then to be $P$ is to be in pain, not simpliciter, but in a specific way. But, one might think, a disembodied spirit has pain simpliciter if it has pain at all (assuming, for the sake of the argument, that pain is phenomenally determinate). It would follow that the spirit has a determinable without having any of its determinate, which is presumably not possible. The question, however, is why having pain in the unembodied way does not count as having it in a specific way.

Second, the conflict might arise from the thesis, suggested by various authors, that the determinates of one determinable need to be able to be arranged along some dimension of property space. Plausibly, there is no such dimension along which we find both disembodied pain and the physical realizers of pain. However, Yablo does not seem to be committed to the dimensional account of determinables.

Third, the conflict might arise from the theoretical role that the distinction between determinables and determinates is supposed to play in mental causation. If disembodied pain is a determinate of pain, then so is embodied pain, one might think. Suppose I am in pain, and flinch. It is not the case that if I had been in pain but not in embodied pain, I would still have have flinched. Disembodied beings don't flinch, after all. Hence the determinable pain fails to screen off its determinate embodied pain. The pain is not enough for the effect, in Yablo's technical sense. It follows that the pain is not causally relevant for my flinching, which is the result that the theory was designed to avoid.

Again, there might be ways to respond to this problem. It is not obvious that if disembodied pain is possible, then embodied pain is a determinate of pain. Perhaps a realizer (an ultimately determinate physical property of the brain, say) of a

phenomenally ultimate determinate is itself a determinate of pain, but a disjunction of realizers of different phenomenally ultimate determinates is not. I will have to leave it open here whether any such view could be motivated.

There is much more in these papers than I have the space and competence to discuss. There is a wealth of interesting linguistic data, especially in "Shoulda, Coulda, Woulda" and "How in the World?" There are subtle points about the interpretation of Descartes and Kripke. In many places, potentially important philosophically observations are made en passant, and often with the tongue in cheek. For example, Yablo raises a problem for an analysis of counterfactuals in terms of causally isolated worlds, as advocated by David Lewis, in the course of justifying not writing an introduction: "Your counterpart in the nearest possible world where the boring introduction is completed would thank me on your behalf, if she could get a message through."

These papers may not always convince the reader of their main thesis. But they will not leave her without new philosophical ideas, and a richer appreciation of the problems they discuss.[8]

---