(Slightly corrected version of review published in *Notre Dame Philosophical Reviews*, https://ndpr.nd.edu/news/the-scientific-imagination-philosophical-and-psychological-perspectives/)

Arnon Levy and Peter Godfrey-Smith (editors)
The Scientific Imagination: Philosophical and Psychological Perspectives
Oxford: Oxford University Press, 2020. Pp. xiv + 344.

The imagination is not a random idea generator. Nor is it confined to dumb association of ideas. It is often intelligent, well-informed, and sharply focused. This collection of thirteen new essays reflects growing agreement that imagination plays a key role in science, not just in thinking up hypotheses, but in testing them. Although thought experiments and idealized models provide salient examples, they are far from the only ones. But precisely articulating the nature of the scientific imagination and its products is no easy challenge.

The sixteen authors include many of those who have already made significant contributions to meeting the challenge. The volume furthers the debate and will encourage others to get involved. It is greatly to be welcomed; any critical remarks below should be read in that context.

Three essays by cognitive psychologists form the most cohesive part of the book: Tania Lombrozo writes on '"Learning by thinking" in science and in everyday life' (it happens); Deena Skolnick Weisberg wonders 'Is imagination constrained enough for science?' (yes); Igor Bascandziev and Paul Harris ask 'Can children benefit from thought experiments?' (yes). Together, they provide a wide variety of encouraging evidence from experiments with children and adults for the cognitive role of imagination in making correct—and sometimes incorrect—predictions about simple causal set-ups. In short, imagination is present in science from the start.

Conditional thinking is rarely far from such cases. Lombrozo considers an experiment where 'participants were invited to imagine a narrow cylindrical cup and a wide cylindrical cup of equal heights, each filled with water to the same height' and were asked 'would they begin to pour water when tilted to the same angle, or at different angles?'—a question about what would happen *if* the cups were tilted. Skolnick Weisberg reports on the strongly reality-oriented choices children make between alternative continuations of a story; they are doing something like answering the question '*If* these events happened, what would happen next?' She also describes Ruth Byrne's work on the role of counterfactual conditionals in regret, relevant to learning from one's mistakes: '*If* only I hadn't …'. Bascandziev and Harris discuss an experiment where a ball was to be dropped down a tube and children were asked 'to predict where the ball would fall'; whether '*when* dropped' or '*if* dropped' is the understood qualification may make little difference to how the answer is found. In another study, 'three- to five-year-old children were told a story in which A caused B' and then asked 'what would have happened if A had not occurred'.

Even without story-telling, conditionals are closely connected to imagination. The most basic way to assess 'If P then Q' is by assessing the hypothesis that Q on the supposition that P, which typically involves an imaginative exercise of drawing out the supposition's consequences, causal as well as logical (this is a version of the well-known Ramsey test). Conditionals are ubiquitous in science. For example, to test a hypothesis, one

must work out what would be the case *if* it were true. That is one reason to expect imagination to be widespread in science.

Conditionals make a much simpler, more general, and psychologically more basic entry-point than do thought experiments or models to the cognitive role of the imagination. Conditionals have far fewer moving parts to control. Admittedly, their semantics is a notoriously tricky issue, but one cannot finesse it by starting with thought experiments or models instead, for the reasoning associated with them is itself loaded with conditionals.

Surprisingly, conditionals have little more than a walk-on part in this volume. The most extended treatment is in Peter Godfrey-Smith's essay 'Models, fictions, and conditionals'. He explains the value of scientific models in terms of their ability to generate counterfactual conditionals. The only other serious discussion of conditionals here is in Fiora Salis and Roman Frigg's wide-ranging discussion 'Capturing the scientific imagination', which includes an argument that counterfactual conditionals are unsuitable for capturing the kind of imaginative activity at work in thought experiments and scientific models. They prefer an account in terms of make-believe. Salis and Frigg's main objection to conditionals is that if C is irrelevant to a thought experiment or scientific model (with conditions described by M), C should not count as true in the latter, whereas the counterfactual conditional $M \,\square\!\!\rightarrow C$ might easily come out true on the possible world semantics of Robert Stalnaker or David Lewis. That objection is unconvincing. Relevance is a standard Gricean conversational implicature; in any conversational context, many things are true but irrelevant. Those who want more than a conversational implicature can always build the relevance requirement into the meaning of the word 'in' trivially, by stipulating that 'C is true in M' is true if and only if (i) $M \,\square\!\!\rightarrow C$ is true and (ii) C is relevant to M. But it is hardly worth the trouble, since the irrelevant truth is scientifically harmless. Salis and Frigg also complain that there is 'no general agreement' on the epistemology of counterfactual conditionals, but since there is no general agreement on their preferred make-believe account either, they are ill-placed to put much weight on that consideration.

One of several significant features differentiating conditionals from both make-believe and fiction is this. When we assess a conditional in the basic way, we first assess the consequent on the supposition of the antecedent. But there is a further stage, when we discharge that supposition and assess the whole conditional from outside it. In formal systems of natural deduction, that corresponds to the rule of conditional proof (or →-introduction), when having derived C from A and background suppositions, we can derive the indicative conditional $A \rightarrow C$ from those background assumption alone (a more restricted version of the rule is needed for the counterfactual conditional $\square\!\!\rightarrow$). Less formally, in mathematics, that is the standard way of proving conditionals. Without the possibility of discharging suppositions and speaking from outside them, scientific statements cannot be held to account against physical reality, since even contradictions are sometimes correctly asserted on suppositions, for example in a proof by *reductio ad absurdum*. Discharging a supposition is a normal move within science; it does not take one to a radically different kind of discourse. Such moves are often made many times within a single mathematical proof. By contrast, no such move is normally available *within* a game of make-believe or a fiction. One can move from making an assertion 'P' inside a game of make-believe to making the assertion 'In the game, P' outside the game, or from putting 'P' into a novel to putting 'In the novel, P' into a work of literary criticism, but in each case it is a move into a radically different kind of discourse. In this

respect, scientific discourse about a model or thought experiment seems closer to the use of conditionals than to make-believe or fiction.

Amie Thomasson in 'If models were fictions, then what would they be?' and Martin Thomson-Jones in 'Realism about missing systems' argue in detail for the assimilation of scientific models to fiction. Much of their attention goes to ontological problems about models, which they treat similarly to ontological problems about fiction. By contrast, 'if' seems to raise no special ontological problems of its own, which is one attraction of conditional-based accounts of scientific models. That is not to deny that models have some similarity to fiction. Rather, the point is that since we have firmer theoretical control of conditionals than of fiction, it is methodologically questionable to try to understand models in terms of fiction or make-believe before at least trying to understand them in terms of conditionals. The truth-conditions of a conditional differ clearly and overtly from those of its consequent, even when the latter is uttered on the supposition of the antecedent—that is how compositional semantics works. On an account in terms of fiction or make-believe, what corresponds to the consequent of a conditional is the same sentence uttered in story-telling mode or as a move in a game of make-believe, where it is much less clear what to say about its truth-conditions.

Godfrey-Smith slightly muddies the waters here by tentatively endorsing John Mackie's non-factualist theory of counterfactual conditionals, which denies them truth-conditions and makes no sense of their occurrence in more complex sentences. For science, counterfactual conditionals sometimes need to be embedded under quantifiers and truth-functional operators. Thus 'Every sample which would have dissolved if put in water turned red' is intelligible enough, but Mackie's account cannot handle it without cheating. By contrast, standard treatments of counterfactual conditionals like Stalnaker's and Lewis's in a framework of truth-conditional intensional semantics have no comparable difficulty with embedding; they are fully compositional. Accounts of thought experiments and models in terms of counterfactual conditionals have much greater descriptive adequacy and explanatory power when combined with such a truth-conditional compositional semantic theory of those conditionals.

Semantic theories of conditionals have been tested against a much wider range of data than semantic theories of fiction or make-believe, and have a richer and better-understood logical structure. Thus if an account of thought experiments or models in terms of conditionals is false, it should be much easier to refute, and its refutation should yield some useful structural constraints on a more adequate account. That is another reason to start with accounts in terms of conditionals. But, on the evidence of this volume, nothing like such a refutation is currently on offer.

One might think that thought experiments and scientific models obviously *do* raise ontological problems, so a simple conditional account must be missing something. To take one of Thomson-Jones's examples, what are we to make of the sentence 'Simple pendula move sinusoidally'? As literally defined, a simple pendulum is an object meeting various constraints, which entail that it is concrete, but in fact no concrete object meets those constraints, so there literally are no simple pendula. Does this mean that the sentence is about nothing? On his fictionalist view, it is about abstract artefacts which do not literally satisfy the definition, just as fictional characters created by a story-teller do not literally satisfy the descriptions in the story, according to which they came from normal biological procreation.

But why is anything so elaborate needed? At face-value, 'Simple pendula move sinusoidally' is just a modalized generalization, roughly equivalent to 'If there were any simple pendula, they would move sinusoidally'. It is quite different in verb aspect and meaning from 'Simple pendula are moving sinusoidally', whose actual truth really does require actual motion. Hearers unfamiliar with the scientific context might take 'Simple pendula move sinusoidally' to carry the presupposition that there are some simple pendula, but the truth of such a presupposition is in any case not necessary for the sentence to express a true proposition. The compositional semantics of 'If there were any simple pendula, they would move sinusoidally' determines appropriate truth-conditions, through the interaction of the literal meanings of its constituents; why should the compositional semantics of 'Simple pendula move sinusoidally' be so different? Neither the literal meaningfulness nor the literal truth of 'Simple pendula move sinusoidally' requires 'simple pendula' to have a non-empty extension; an intension will suffice.

A literal interpretation is also more promising for making sense of the science, for what physicists use in calculating the motion of a simple pendulum is the literal definition. A non-literal interpretation would have to do much more work to reconstruct what is going on in a real-life scientific text, if it involves transitions between literal and non-literal readings of the same symbols. The validity of those transitions would need to be properly checked.

Of course, story-telling is a thing humans do. Scientists may indeed sometimes talk of simple pendula as fictional characters, and something like Thomasson's or Thomson-Jones's account of fictional characters may then apply. But that seems to be a secondary phenomenon, inessential to the main scientific action.

A metaphysician may still have serious worries about the ontology of possible worlds and the domains of their inhabitants in a standard framework for intensional semantics. But such generic concerns over quantified modal language should be addressed at that level of generality. To treat them as specific to thought experiments or scientific models is to risk special pleading and *ad hoc* moves in special cases of what is really a much wider problem.

Attention to logical structure may also help us understand the apparently fictional aspect of objectual imagination, discussed by Frigg and Salis, as when one imagines a simple pendulum, although there is no simple pendulum for one to imagine. In proofs, the normal way to exploit an existential generalization $\exists x\, F(x)$ is first to suppose an instance $F(a)$, where a is a new 'arbitrary name' (not in $F(x)$), then derive from it (and auxiliary suppositions not containing a) a consequence C (not containing a); then C follows just from $\exists x\, F(x)$ (and the auxiliary suppositions). Imagination may work similarly, though less formally. Imagining a simple pendulum may go beyond just imagining that *there is* a simple pendulum. Read $F(x)$ as 'x is a simple pendulum'. In supposing the arbitrary instance 'a is a simple pendulum' and imaginatively exploiting it, one does something of the same form as supposing of a particular object that *it* is a simple pendulum, and imaginatively exploiting the supposition, although there is no particular object of which one has supposed that *it* is a simple pendulum. None of this requires an ontology of abstract non-literal simple pendula. If one likes that ontology for independent reasons, one might apply it here, but no such objects need figure in *what* the scientist supposes or imagines, any more than they do in the content of mathematical proofs from existential generalizations.

These concerns about the over-interpretation of scientific discourse on thought experiments and models are reminiscent of Stacie Friend's conclusion in 'The fictional character of scientific models' that the ontological disputes make little difference to more

urgent *epistemological* questions about scientific models as fictions, though the present standpoint is less friendly than hers to accounts of models as fictions. It is not that ontology is in general irrelevant to epistemology, for ontology is relevant to *what* we need to know, but in this case the ontological disputes may concern a secondary level of discourse and cognition.

One moral is that we should not give up on compositional semantics too quickly before resorting to non-literal interpretation. This is also relevant to Arnon Levy's considerations in his chapter 'Metaphor and scientific explanations'. Of course, comparisons and analogies are used continually in science, but they can be made in literal terms and differ psychologically and cognitively as well as semantically from metaphors. In this regard, Elisabeth Camp makes a battery of helpful distinctions in 'Imaginative frames for scientific inquiry: metaphors, telling facts, and just-so stories'.

Overall, the book contains rather little on the logic and semantics of the language of thought experiments and models (including mathematical language and notation). One reason why language matters for the scientific imagination is that it is in effect the main interface between the individual processes of imagining and the collective process of science. However, it is not the only interface, for pictures and diagrams also play such a role, as well as providing aids to the imagination: 'a picture is worth a thousand words'. In 'Imagining mechanisms with diagrams', Benjamin Sheredos and William Bechtel provide a fine-grained and well-documented case study of the key role of diagrams in reasoning about the mechanism responsible for circadian rhythms (a sort of body-clock) in cyanobacteria.

The other case study in the volume, at a higher level of generality, is Michael Weisberg's 'Abstraction and representational capacity in computational structures', which responds to arguments that models need to be counterfactual fictions rather than purely mathematical structures in order adequately to represent causal structure. He explains how the causal structure of a target system can be represented by the procedural structure of a computer program, and how the interpretation of the computational structure can be adjusted to achieve the required level of generality, in ways which do not work for fictions.

Conditional accounts of scientific models may not be appropriate for capturing the fine structure of diagrams and computational processes. To that extent, we might treat a conditional account as a *model of modelling*, abstracting away from various nuances and complications to understand something central about what models do. Analogously, Camp discusses *metaphors for metaphor*.

Strikingly, no chapter applies a model-building approach to model-building itself. Closest comes Stephen Yablo's 'Models and reality'. He recycles the title of a notorious paper by Hilary Putnam to cast models in a much more positive semantic role than Putnam did. Yablo's aim is to show how a model's simplifications can be harmless because a false statement can nevertheless be true as far as its intended subject matter goes. Following David Lewis, he treats a subject matter as a partition of possible worlds: worlds are exactly the same as far as the subject matter m goes (m-wise indiscernible) just in case they fall into the same cell of the partition m. Yablo then defines a sentence S to be true in a world α about m if and only if S is true *simpliciter* in all the "best" worlds that are m-wise indiscernible from α. Here one of two m-indiscernible worlds is better in the relevant sense than another 'just if it better illuminates what is going on m-wise—for example, by containing fewer distorting influences or irrelevant complications'. Thus, in the actual world, S may be false *simpliciter* yet still true

about *m* because some worlds are both *m*-wise indiscernible from the actual world and better than it *m*-wise, and *S* is true in all those worlds.

Unfortunately, Lewis's conception of subject matters may not be well-suited to understanding the relation between models and reality, since the complications a model abstracts from are often *aspects* of its intended subject matter, not irrelevant to it. For example, we may want to model a continuous process of diffusion with a computational procedure over a discrete grid of cells, using a discrete time order. Two worlds which differ slightly in the underlying fine-grained continuous process may be exactly the same with respect to the coarse-grained discrete representation. If we take the fine-grained partition as the intended subject matter, Yablo's definition will count many sentences generated by the model as not true *about its subject matter*. For instance, the model may represent the system as having reached equilibrium when in reality it is still slightly changing. Thus his strategy fails to achieve what it was supposed to achieve. If instead we take the coarse-grained partition as the intended subject matter, we move the goalposts by no longer treating the continuous diffusion as the target system, and it is still unclear that Yablo's definition will count all sentences generated by the model as true *about its subject matter*. For instance, in the long run an accumulation of small errors may make the multiply iterated discrete computation diverge even by coarse-grained standards from the continuous diffusion. Again Yablo's strategy fails to achieve what it was supposed to achieve.

Can we treat Yablo's account of how models work as a good but simplified model of how they work? It seems to violate his own standard for that status. For the intended meta-subject matter is the relation between fine-grained reality and coarse-grained models. Since fine-grained differences in reality are relevant to that meta-subject matter, we are back with the same sort of problem as in the previous paragraph. The proposed aim, exact truth about the intended subject matter, is unrealistic for many good models.

Godfrey-Smith makes a more promising suggestion in his conditional account. He considers a model which delivers the counterfactual conditional 'If *A* were the case, *C* would be the case', where *A* is obviously not the case. He suggests that we may then be able to move to a new conditional 'If *A* were approximately the case, *C* would be approximately the case', where *A* is indeed approximately the case (I have altered his wording in inessential ways). As he notes, the move is not deductively valid, but in suitable circumstances it may nevertheless constitute a good non-deductive inference. Given the additional true premise '*A* is approximately the case', we may then infer the true conclusion '*C* is approximately the case'. As Godfrey-Smith is well aware, this is all highly programmatic. We want to know how the 'approximately' operator works, and what makes circumstances 'suitable' for the non-deductive inference, though we cannot expect fully formal answers to either question. Still, Godfrey-Smith's suggestion promises to take us closer to how models actually work in science, and not further away.

Much of this review has concentrated on points of disagreement, in the usual way. However, it must be strongly emphasized that anyone interested in the scientific imagination will have a lot to learn from this volume.