ORIGINAL RESEARCH

Beyond the limits of imagination: abductive inferences from imagined phenomena

Michael Traynor[1]

Abstract

The present paper proposes a route to modal claims that allows us to infer to certain possibilities even if they are sensorily unimaginable and beyond the evidential capacity of stipulative imagining. After a brief introduction, Sect. 2 discusses imaginative resistance to help carve a niche for the kinds of inferences about which this essay is chiefly concerned. Section 3 provides three classic examples, along with a discussion of their similarities and differences. Section 4 recasts the notion of potential explanation in Lipton's (Inference to the best explanation, Routledge, Abingdon, 2004) in order to accommodate inferences to possibility claims; Sect. 5 then attempts to characterise a principle underlying such inferences. Section 6 concludes by discussing how the proposal relates to other modal epistemologies, with emphasis on the potential of such inferences to produce genuinely new ideas.

Keywords Imaginative resistance · Thought experiments · Perception · Time without change · Turing test · Absolute space · IBE · Abduction

> If we had to travel to other possible worlds to learn what goes on there, we could know nothing about uninstantiated possibilities or necessities; for possible worlds are not foreign countries or planets. Rather, they are alternative complete universes and so are physically inaccessible. Nevertheless, there are standard patterns of inference entitling us to infer facts about ways things could be—about other possible worlds. (Sorensen, *Thought Experiments* (1992, p. 77))

✉ Michael Traynor
   michaelthomastraynor@gmail.com

[1] Fife, UK

 ⌀ Springer

1 Introduction

It is common in philosophy to appeal to imagination for evidence concerning what is possible. Hart, (1988) and Yablo, (1993) have suggested a compelling analogy to make sense of this: Imagination is to the possible as perception is to the actual. There are limits to the justificatory force of imagination, however. While, for many, the imagination has an exalted status as a window to the possible, an appeal to the imaginability of a certain event or type of entity falls flat if one's interlocutors find themselves unable to conjure up scenarios or worlds as they are asked. Moreover, should they succeed in doing so, they may yet doubt the strength of the supposed connection: imagination may be a guide to modal space, but, like perception—perhaps even to a greater extent—it is fallible.

Disputes about what is imaginable thus seem particularly prone to dialectical stalemate. Since we cannot peer into others' minds and inspect the contents, the best we can do is call a draw, or, perhaps worse, bemoan our rivals' limited horizon—or their attachment to claims beyond the bounds of sense, as the case may be. We might go a little further, and offer an error theory, expounding on the source of their false claims, but for the purpose of convincing them to change their minds, this puts the cart before the horse, since such an explanation assumes, hence does not prove, that the interlocutor is wrong about what they can imagine.

This paper explores connections between imagination and perception, with regard to the limits of each, and particularly with regard to the capacity of the former to resolve philosophical disputes where imaginative failure is a factor. We do not limit ourselves to perception when investigating the actual world: we can also construct theories, and believe some of these theories when they do a good job of accounting for what we perceive. Extending the perception-imagination analogy, then, we can also construct theories about imagined phenomena. Might theorising about imagined phenomena provide justification for belief in possibilities which themselves lie beyond the reach of imagination? I argue that this is so. That is not to say that imagination and perception are perfectly analogous—I grant that imagination may well be more fallible, and, as I say, particularly apt to lead to disagreement, certainly in matters of metaphysics. It is perhaps in the nature of metaphysics, if not philosophy more generally, to deal with the putatively unobservable; so, to the extent that imagination is tied to perception, and that metaphysics deals in possibilities, metaphysics has a particular interest in means of justification that transcend imagination.

To this end, I aim to make space for a type of inference that utilises imagination in order to go beyond it. I adopt the widespread assumption that imagination *can* furnish modal knowledge; that is, I am not addressing the global modal sceptic; rather, the aim is to legitimate a type of inference that may be utilised in particular instances when imagination runs aground. With this in mind, the next section (Sect. 2) discusses in general terms the kind of dialectical context in which such inferences may be appealed to.

Section 3 provides three classic examples of such inferences, along with a discussion of their similarities and differences. Section 4 recasts the notion of potential explanation in Lipton's, (2004) in order to accommodate abductive inferences to possibility claims; Sect. 5 then attempts to characterise a principle underlying such inferences.

Section 6 concludes by considering the potential for such inferences to produce genuinely new ideas. Comparing Hanrahan's, (2007) explanationism will bring out the need to acknowledge that such inferences are not equipped to deal with the modal sceptic who denies the capacity of imagination to support even the most mundane of modal claims; but if imagination is granted the *potential* to justify modal claims, I contend that appeal to the best explanation of an imagined scenario has the potential to justify the possibility of the unimaginable.

2 Failures of imagination

This paper takes it for granted that, in general, imagination is a guide to possibility. However, the usefulness of imagination to this end is reduced to very little as soon as any party to the debate in question claims to be unable to imagine the relevant scenario or world. Such imaginative failure may stem from the nature of the entities in question, the nature of imagination, or the nature of the interlocutor. This section explores the limits of imagination qua justifier of possibility claims, in order to pave the way for a novel way of circumventing these limits, such that certain possibility claims may be justified in spite of imaginative failure. The aim here is to give a sense of how an appeal to imagination can cease to be dialectically effective, without being discounted as a guide to possibility in general terms—that is, both parties to the dispute regard imagination as a guide, but a reason has been given as to why imagination fails in a *given* case; in such a context, an inference drawn from imagined phenomena can legitimately reach beyond the justificatory limits of sensory or stipulative imagination.

In her survey article on imagination, Gendler characterises *imaginative resistance* as follows:

> *Imaginative resistance* occurs when a subject finds it difficult or problematic to engage in some sort of prompted imaginative activity. Suppose, for example, that you were confronted with a variation of *Macbeth* where "the facts of [Duncan's] murder remain as they are in fact presented in the play, but it is prescribed in this alternate fiction that this was unfortunate only for having interfered with Macbeth's sleep" (Moran, 1994). If you found it difficult to imagine this, even though the author had done everything authors usually do to make such a story fictionally true, then you would be experiencing imaginative resistance. (Gendler, 2013, Sect. 5.2)

Gendler goes on to distinguish two ways of interpreting a philosopher's claim to be unable to imagine something, by way of two kinds of theory: "Can't" theories, and "Won't" theories. *Won't* theories involve the would-be imaginer being *unwilling* to imagine something, for instance, "because doing so might lead them to look at the (actual) world in a way that they prefer to avoid" (Gendler, 2013, Sect. 5.2). While it is certainly possible for failure to imagine some phenomena to be a matter of the will, I focus on *can't* theory for two reasons: Firstly, a *won't* theory brings with it a perhaps uncharitable accusation of disingenuousness. And secondly, where a *won't* theory applies, it is perhaps appropriate for modal epistemology to give way to a discussion of the would-be imaginer's psychology.

Can't theories can then be separated into two broad categories, the simple and the sophisticated.

> Simple *can't* theories often embrace some sort of *impossibility hypothesis*, suggesting that propositions that evoke imaginative resistance are impossible in the context of the stories where they appear, and that this explains why readers fail to imagine them as true in the fiction. … Brian Weatherson (2004) offers a more sophisticated version of a *can't* theory, suggesting that resistance puzzles arise in the face of a certain *type* of impossibility: they arise in cases where the lower-level facts of the story and the higher-level claims of the author exhibit a particular kind of incoherence. (Gendler, 2013, Sect. 5.2)

I submit that simple *can't* theories are relatively toothless. There are two broad ways in which an impossibility hypothesis might be justified by imaginative failure. The first is where it is claimed that unimaginability entails impossibility. In the absence of further substantive claims, this is implausible given that there are contingent limitations on our imaginative capacities: failure to imagine a chiliagon does not imply that corresponding objects are impossible (Gallois, 1974). The second is by way of an inference to the best explanation, where it is claimed that the impossibility hypothesis *best explains* imaginative failure. But in order for an inference to the best explanation to be dialectically effective, the data to be explained must be accepted by those we are trying to convince. Thus, for example, if Turing were to claim that thinking machines are possible because he can imagine them, and his interlocutors claim that he must be mistaken about the content of his imagining because such are impossible, we have reached dialectical deadlock. We should also note that, even if all parties find themselves unable to imagine something, an impossibility hypothesis would need to beat out alternative explanations for this imaginative failure, such as fall under the category of sophisticated *can't* theories.

Sophisticated *can't* theories perhaps have more bite, providing a principled reason for imaginative failure. A sophisticated *can't* theory along the lines suggested by Weatherson, (2004) might proceed from a particular theory about what it is to imagine something. According to Martin, (2002) for example (and presaged in Peacocke, (1985) and Gallois, (1974)), "to imagine sensorily a φ is to imagine experiencing a φ" (Martin, 2002, p. 404). This takes the link between (sensory) imagination and perception quite seriously, and has implications for the capacity of imagination to evidence possibilities which would lie beyond any possible experience. Consider, for example, the debate concerning whether time could pass in the absence of change, which we will consider further below. In his article 'Time without Change', which furnishes one of our paradigm cases to be considered below, Shoemaker notes that "it is plausible to suppose that as long as one is aware of the passage of time some change must be occurring, namely, at a minimum, a change in one's own cognitive state" (Shoemaker, 1969, p. 367), and hence that "it is logically impossible … for someone to be aware of the existence of a changeless interval during that interval itself" (Shoemaker, 1969, p. 368). Hence, where φ is 'a period of time without any change', we are left with a certain *incoherence*, between the suggestion that one's imagery is of time passing without change, and a higher-level claim about what it is to imagine something. (Incidentally, such incoherence would rival, qua explanation of imaginative failure, an impossibility

hypothesis of the kind mentioned in the previous paragraph. While this may seem to help the case for the possibility of time without change, it nonetheless, if successful, blocks any simple argument from imaginability to possibility, by providing a principled reason why it cannot be imagined.) As a consequence of this account of sensory imagination, if we are to imagine a world in which time passes without change, it must be stipulation—as opposed to any sensory features of the imagining—that furnishes the crucial properties.

However, as Kung, (2010) argues, the probative force of imagination is doubtful where all the work is being done by supposition. Kung distinguishes two aspects of imaginings: the *assigned* content and the basic *qualitative* content, which correspond roughly to stipulated content and imagistic content, respectively. On Kung's account, there is a sense in which we can imagine impossibilities, but generally only where the assigned content does all the work. This does not impugn the modal justificatory force of imagination in general, because on Kung's account, it is the basic qualitative content of imagining that gives it its probative force. This is because the basic qualitative content of an imaginative experience—just as that of a perceptual experience—presents a way that space *could* be filled around the perceiver (Kung, 2010, p. 637).[1] As such, failure to put together the qualitative content required to provide an imagining with probative force can (in principle) be explained by the inconsistency of the arrangement we are attempting to construct in our mind's eye. In this way, the explanatory potential of impossibility hypotheses is itself explained by way of a link between imagination and experience. But there remains a sense in which we can imagine impossibilities: for Kung, it is the stipulative aspect of imagining—the assigned content—that all-too-easily permits the imagination of impossibilities, for "stipulations and labels are virtually unconstrained, and what minimal constraints there are have no modal epistemological value" (Kung, 2010, p. 634).[2]

As Peacocke, (1985, p. 19) points out, whether one is imagining a suitcase, or a suitcase with a cat hidden behind it, may be determined not by the mental image itself, but by the stipulations attached to it, namely, that there is a cat hidden from view. This is of course an uncontroversial case, the imaginability or possibility of which nobody could sensibly deny, but in other cases the coherence of an image with a stipulation about it may be at the very centre of a philosophical debate. In many metaphysical disputes, it is clear that imagery alone will not suffice to establish the coherence of the possibilities in question, and in the context of a debate where such possibilities are

---

[1] This echoes Yablo's gloss on the imagination-perception analogy: "Just as someone who perceives that p enjoys the appearance that p is true, whoever finds p conceivable enjoys something worth describing as the appearance that it is possible" (Yablo, 1993, p. 5).

[2] "[T]he principal constraint on assignments is certainty. I said that so long as we find P believable, epistemically possible in the strongest sense that it is true for all we know for certain, or possibly true for all we know for certain, we will be able to imagine P via stipulation or label. Let P be some proposition whose possibility we are trying to establish via imagining. The mere fact that we find P (or possibly P) believable, and hence are capable of making the assignments required to make P true in the imagined situation, is not good evidence for P's possibility. Believability just is lack of certainty. (Let us use 'non-certainty' to denote lack of certainty; it avoids the unwanted connotations of 'uncertain'.) It would be very odd if our non-certainty counted as evidence of P's possibility" (Kung, 2010, p. 634). The principle I advance in Sect. 5 takes stipulation's lack of probative force into account when allocating its role, namely as a gap-filler for *uncontroversial* details, the possibility of which will be granted by all parties in the debate at hand.

doubted, stipulation, too, is bound to fail to convince. This paper aims to provide a way of moving forward in such cases when imagination runs aground—as in cases where all parties to a dispute allow that imagination can warrant belief in some possibilities, but where there is disagreement on the imaginability of a *given* scenario, or on whether an imagined world can be said to be one in which a *particular* proposition is verified. Given that we cannot determinately imagine whole worlds, there will always be aspects of imagined worlds the possibility of which is not settled by the act of imagining alone; moreover, dialectical constraints are often in play: where there is a particular disagreement about what *can* be imagined, some other method—besides asking one's interlocutor to imagine the 'possibility' up for debate—is needed. The next section considers three well-known thought experiments, relating to the possibility of thinking machines, of changeless passage of time, and of absolute space. These will provide the basis for our later attempt to characterise a principle according to which possibility claims lying beyond the evidential reach of imagination can be justified through an appraisal of merely possible phenomena.

3 Inferring to possibility, from imagined phenomena

Inference to the best explanation (IBE) is a procedure whereby a shortlist of potential, mutually exclusive explanations of a given set of data is drawn up, and these are compared in terms of ideological and ontological simplicity, explanatory power, coherence with fundamental or well-established principles, internal coherence, etc., until it is determined whether there is one among these theories that can be said to be the best of the bunch. If so, that theory is inferred to be true. Such a method is of particular interest to scientific realists because it offers the potential to justify belief in entities that cannot be directly observed; indeed, such entities might be said to be the raison d'etre of IBE, for where *observable* entities are posited, the most secure way to determine whether reality abides is to seek to observe them.

It might reasonably be said that it is in the nature of metaphysics to consider that which lies beyond the senses, and in many cases—again, depending on the strength of the imagination-perception connection—that which cannot be conjured up by imagery. If we grant the legitimacy of IBE as a route to belief in the unobservable, by way of consideration of what *can* be observed, might we not do the same for possibilities beyond simple confirmation by imagination, proceeding by way of that which *can* be imagined?

There is a class of thought experiments which seek to do just that, which we will consider by way of three exemplary cases.

Newton's spheres: Imagine a world in which two spheres are joined by a cord. The material of the spheres is such that neither attracts nor repels the other. The cord is taut, and yet the spheres are not moving towards each other under the force of the cord. What explains this set up, Newton proposed, is that the spheres are rotating—but with respect to what? Not with respect to each other, for each retains its position relative to the other throughout. And yet no other concrete objects have been posited as part of the scenario; thus, to explain the phenomena so described, Newton proposes, we need to posit absolute space. (Newton, 1966, p. 12).

Shoemaker's freezes: Imagine a universe divided into three populated regions, A, B and C. The inhabitants find that each region undergoes periodic freezes, each freeze lasting one year. They can come to know this because the local freezes are staggered: by observations of clocks in unfrozen regions in the times before and after the freezes, the inhabitants find that region A freezes for one year during every third year; region B freezes for one year every fourth year; and region C freezes for one year every fifth year. Having noticed this regularity of local freezes, they realise that if each individual region continues to follow its observed pattern of freezes, then every sixtieth year will be one in which all three regions are frozen (for all and only multiples of 60 are multiples of *all three* numbers, 3, 4 and 5). Since the three regions exhaust this imagined universe, this would amount to a *total* freeze: every sixtieth year would be one in which time passes without change. (Shoemaker, 1969, pp. 370–371).

Turing's test: Turing, (1950) famously proposed a test for the presence of mind, thought or intelligence in machines.

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 percent chance of making the right identification after five minutes of questioning. … I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. (Turing, 1950, p. 442)

Putting aside the correctness of Turing's predictions, and the demandingness of the test he proposes, he is effectively claiming that, given that a computer could convince us—the average interrogator—that it is human more often than not, we should conclude that it is possible for machines to think. While Turing is perhaps more concerned about what *will* be the case, than he is about what *could* be the case, and about the specific question about machine intelligence, there seems to be a more general principle underlying his test, such that the best explanation of some imagined phenomena is itself possible[3,4].

Each of these cases has its nuances. Turing describes an experiment that could actually be performed—in principle, that is, for the computers existing at the time Turing wrote these words were yet some way from the capacity he spoke of—whereas both Newton's and Shoemaker's cases are of decidedly other-worldly phenomena. Turing's case also essentially involves an observer, the one playing the imitation game; for Shoemaker's case, we might omit the observers, while Newton's world is crucially desolate: in order to compel us to infer to the existence of absolute space, there needs to be an absence of things with respect to which the spheres can be said to be rotating.

---

[3] It appears that many have read Turing as claiming that winning the imitation game is a sufficient condition for the presence of mind, a claim which is difficult to maintain in the face of thought experiments such as Searle's, (1981) "Chinese Room". In this connection, it is interesting to note the probabilistic criteria that Turing offers, of winning 70% of the time, suggesting a probabilistic *guide* to the presence of a mind. As Oppy and Dowe, (2020, Sect. 4.4) note, this probabilistic aspect of Turing's discussion is often overlooked.

[4] Turing begins his paper "I propose to consider the question, "Can machines think?"" (Turing, 1950, p. 433). While it is true that Turing talks of 'replacing' this question with another—namely, that of whether a machine could win the imitation game—his proposed replacement is intended as a route to answering the first modal question, which he returns to, and after all refers to as "the main question" (heading of Sect. 6).

However, there is a common theme that needs exploring: the idea that an appeal to merely possible phenomena—and what might reasonably be concluded from it—can justify a claim about a possible world beyond what is conjured up by stipulative or sensory imagination; can justify, in other words, belief in a possibility.

That this is what is going on is perhaps not so clear in Newton's case. Newton seems to move from the claim that absolute space must exist in order for us to make sense of what happens *in the two-sphere world*, to the claim that it must exist in the *actual* world. However, this seems unwarranted. Since the inference to the existence of absolute space at the two-sphere world depends on features of that world which do not pertain to the actual world, more needs to be said if the two-sphere world is to provide us with any grounds for thinking that absolute space *actually* exists.[5]

However, what does seem reasonable is that, if indeed Newton has shown that absolute space is part of the best explanation of the goings on at the two-sphere world, then we should grant that it is *possible* for space to be absolute. This is an instance of a kind of argument that can be particularly useful in contexts where the dispute is about some possibility claim, and where an appeal to imagination, for one reason or another, is ineffective. In Newton's case, this may well be why thought experiments like the two-sphere one are appealed to: it is not at all clear that absolute space can be imagined, so appeal to imagination is ineffective against someone who doubts its possibility. This is in line with Brown's, (2004) characterisation of the thought experiment.

> Newton's spheres do not so much give us a new result but, rather, give us a remarkable phenomenon, something that needs to be explained. The thought experiment establishes a phenomenon; the explanation comes later. And the best explanation, according to Newton, is the existence of absolute space. This way of looking at it is confirmed, it seems to me, by the way in which Berkeley and Mach reacted to the thought experiment. They didn't deny that rotation with respect to absolute space is the best explanation for the tension in the chord. Instead, they denied there would be any tension in the first place, if the spheres are not moving together. Or the two spheres would move together, if there were any tension. That is, they didn't bother to challenge the explanation of the phenomena that Newton posited; they challenged the alleged phenomenon itself. (Brown, 2004, pp. 29–30)

To these remarks about Berkeley and Mach, we should also add that, had their strategy been to challenge the explanation of the phenomena that Newton posited—say, by providing an alternative explanation—this would amount to tacit acceptance of

---

[5] What more needs to be said? Perhaps the bridging assumption is that the nature of space—relative or absolute—is a matter of nomological or metaphysical necessity. In addition, an anonymous reviewer proposes: "If it is about nomological possibility, then the possibility claim plausibly implies the actuality claim. After all, if the imagined situation can only occur in absolute space and it can occur in the actual world, then the actual world has absolute space." The assumption here may be that matters of metaphysics, such as whether space is absolute or relative, cannot affect whether some phenomenon is nomologically possible; hence if the two-sphere scenario must involve absolute space, and is nomologically possible, then so must every nomologically possible world (including, of course, the actual). However, if such further reasoning is given to establish what is actual, it nonetheless proceeds from a claim about what is *possible*, based on an inference to a possibility (that of absolute space), from a consideration of merely possible phenomena.

the general method of inferring possibility from an appraisal of merely possible phenomena, the principle that the best explanation of some possible phenomenon is itself possible.

In our other two examples, the concern is likewise, and perhaps more explicitly, to establish a possibility claim: Shoemaker's that time could pass without change, Turing's that machines could think. In each case, this is done via an ampliative inference from an appraisal of imagined evidence.[6]

Is such a method of argumentation legitimate? The question invites us to search for a principle that abstracts from the particulars of each case, to see whether this *type* of inference can withstand scrutiny. This will be useful for assessing the conclusions drawn on the basis of the above cases, but the discussion will also, I hope, be of broader significance. What I am interested in is not merely the particular controversies of each case—whether time could pass without change, whether machines could think, whether absolute space is possible—but rather the method by which the philosophers concerned attempted to weigh in on them.

One might think that such inferences must be counted as legitimate: once it has been granted that IBE is a route to truth about actual phenomena, it is hard to see why we ought not to draw abductive inferences about merely possible phenomena. After all, in a sense, the imagination has an advantage over perception; as Sorensen notes:

> Regular experiments need to be replicated, reviewed, and revisited. The mental folding need not be repeated. Imagination gets the job done on the first try. [...] There is no need to check for leaks, contamination, or procedural errors. The thought experimenter harnesses the clean, renewable power of stipulation. Imagination achieves the experimentalist's ideal of control. (Sorensen, 2016: p. 422)

Given the power to stipulate anything which one's interlocutors are willing to grant as possible without argument, we must be allowed to consider what an experiment *would* show, were it carried out, given a particular set of results, and a consideration of what would best explain a given set of results can be carried out without having to actually undertake the experiment. But there is more work to do. The next section recasts IBE in order to accommodate inferences to possibility claims; Sect. 5 then attempts to characterise a principle underlying such inferences. Section 6 concludes by considering the potential of such inferences to produce genuinely new ideas.

4 IBE for the modal realm

In his classic discussion of inference to the best explanation, Lipton, (2004) provides a characterisation of this mode of inference that may seem to be at odds with the idea that one can infer to a possibility claim on the basis that it would explain the phenomena of some possible world. Operating on an assumption of inferential and explanatory realism for much of the book, he says that "for something to be an actual explanation, it must be (at least approximately) true" (Lipton, 2004, p. 57). Hence,

---

[6] Further examples of such arguments are discussed in Oppy, (2004), and Parsons, (2013) offers one explicitly modelled on Shoemaker's, (1969).

since competing explanations are typically mutually exclusive, "Inference to the Best Explanation cannot be understood as inference to the best of the *actual* explanations"; instead, Lipton suggests, we ought to distinguish *actual* from *potential* explanation (Lipton, 2004, p. 57). Now a natural reading of 'potential' is one according to which it entails possibility. As Lipton explains, "the potential explanations of some phenomena are those that do explain them in a possible world where our observations hold" (Lipton, 2004, p. 59). However, if we are to allow abductive inferences to possibility claims, choosing from a list of mutually exclusive candidate explanations, we need some other way of cashing out '*potential* explanation', for it would beg the question to assume the possibility of all candidate explanations before the inference is drawn. For a realist, a claim can't be explanatory unless it is true; but IBE involves selecting from candidate—i.e. potential—explanations. So how do we generate a list of candidate explanations before determining what is a *possible* explanation?

Lipton's understanding of IBE, as a selection from a list of candidates that each explain our observations at some possible world, is equally problematic for Biggs's, (2011) abductive route to necessity claims, albeit for a slightly different reason. Biggs argues that abductive principles can favour certain explanans over others where the explanans in question are either modal claims, or else claims that entail modal claims (beyond those less metaphysically interesting possibility claims that are entailed by claims about what is actual). In the example Biggs gives, the explanandum is the co-occurrence of water and H_2O, and three potential explanations are offered:

(I) H_2O is identical with water.
(M) H_2O merely metaphysically necessitates water—where 'merely' indicates that the necessitation holds in only one direction.
(N) H_2O merely nomologically necessitates water—where 'merely' indicates that the necessitation is not metaphysical. (Biggs, 2011, p. 297).

As Biggs notes, these three potential explanations are mutually exclusive. If H_2O is identical with water, as according to (I), then water and H_2O metaphysically necessitate each other (assuming, that is, that property identities are metaphysically necessary). Thus (I) rules out (M) because (I) has it that the necessitation goes in both directions, while (M) has it that the necessitation holds in only one direction; and (I) rules out (N) because (I) implies that the necessitation is metaphysical, not merely nomological. Similarly, (M) and (N) rule each other out because they contradict each other on the strength of the necessitation—metaphysical according to the former, and nomological according to the latter.

That the competing candidate explanations are mutually exclusive is typical of an abductive assessment (Lipton, 2004, p. 57). Less typical, however, is the strength with which Biggs's potential explanations exclude each other, for if any of (I), (M), or (N) is true, then the other two candidates are each ruled out *as possibilities*.

Thus, if IBE is to justify modal claims—whether of possibility or necessity—it is necessary to adopt an epistemic sense of 'potential explanation', where a proposition is a potential explanation of xyz where, if—perhaps, *for all we know*, *per impossibile*—it were true, it would explain xyz. On reflection, however, this is perhaps how it should be. That an impossibility *might* count as a potential/candidate explanation can be seen from cases where the impossibility is not yet known to be as such; where it is not yet known

whether two things are identical or distinct, and assuming the necessity of identity, their identity or distinctness might be central to candidate explanations of some facts about them: for example, the non-identity of Batman and Bruce Wayne might explain why they appear to have different voices. With this in mind, the requirement that all candidate explanations must be metaphysically possible is too stringent.

5 Specifying the principle

It is difficult to see why we should not extend the reach of inference to the best explanation in this way—why it should be legitimate with respect to actualia, but not possibilia. Indeed, if such inferences are rejected in principle with respect to the merely possible, we need some principled reason for this, or else the legitimacy of inference to the best explanation of actual evidence is threatened. If the best explanation of some actual phenomenon is thereby considered actual, then the best explanation of some possible phenomenon should therefore be considered possible. However, it remains to formulate a principle characterising such a modal epistemology.

Yablo, (1993) provides a helpful starting point, where he considers a sense of 'conceivability', drawn from Putnam's, (1975) critique of conceivability as a guide to possibility, according to which "[t]o conceive a proposition … is to imagine acquiring evidence that justifies you in believing it" (Yablo, 1993, p. 22). Putting aside doubts about whether it can properly be called a sense of 'conceiving', this sounds rather similar to what is going on in our three thought experiments, so perhaps this can help us formulate a principle underlying them.

If I can imagine a scenario in which I am rationally persuaded of p, then p is possible.

But of course, this will not do, and Yablo notes the fundamental flaw: we can imagine being rationally persuaded of almost *anything*—including necessary falsehoods. Thus this principle immediately falls to counterexamples. One such is familiar from the work of Kripke and Putnam: if we grant that water is necessarily identical to H2O, one can yet easily imagine being rationally persuaded that some watery stuff XYZ ($\neq$ H_2O) is water. Another kind might involve a possible agent relying on the testimony of others: one can imagine the entire community of mathematicians testifying that Goldbach's conjecture has been proven; one can also imagine a world in which all mathematicians testify that a proof has shown Goldbach's conjecture to be false. Assuming it is plausible to believe the testimony of experts (we can assume that they are generally reliable in both cases), our first formulation tells us both that Goldbach's conjecture is possibly true, and possibly false.

Reformulating such a principle, then, we need to avoid approving inferences to possibility claims where there is a lack of relevant knowledge; we need to avoid imagining ourselves into a world where we know less than we actually do—such imaginings are unlikely to licence any new possibility claims. We might also note the importance of agreement on the possibility of the imagined phenomena. Taking these considerations

into account, we should substitute reference to the *imagined* for reference to *possibilities*, and rather than *scenarios* or *phenomena*, refer to *worlds*. It is worlds, not scenarios, that make true possibility claims; moreover, referring to worlds will allow us to avoid drawing inferences from underspecified situations. Kung's (2010) aforementioned reservations about the probative force of stipulative imagining do not prevent us from marshalling stipulation to fill out uncontroversial details (where the possibility of the stipulated aspects is not doubted by either party to the debate), which is just as well: we cannot hope for the qualitative content of our imaginings alone to render the world imagined as determinate. As Yablo, (1993, p. 28) notes, we do not (and perhaps cannot) determinately imagine all aspects of a scenario or world; but we do imagine these as being determinate. This is crucial, in the first instance because ampliative inferences about possible scenarios would be superfluous were we able to simply imagine the possibility in question; and, in the second, because a determinate claim can only be made true by a world that is determinate, at least in the relevant respect.[7]

How much should be specified? Not everything about the world: this is impossible in practical terms, but aside from this, inferences to possibility claims based on a world about which everything is already stipulated would be dialectically useless. At the very least, we need to avoid stipulating the truth of the proposition whose possibility is the subject of debate. How do we arrive at the right degree of specification, such that the world is not underspecified to the point where something relevant to our inference is neglected, but that we do not overspecify to the point where those who doubt the possibility of the proposition call us out for begging the question?

The key lies in our three cases: the propositions that they concerned themselves with were each plausibly beyond direct confirmation by experience. The imitation game purports to test for thought, but would be otiose if this were the kind of thing one could simply locate by observation; similarly, the passage of time could not be observed in the absence of change—at the very least, noticing the passage of time plausibly must involve change in one's mental state; and as for the existence of absolute space, this is of course something that cannot be confirmed by any of the senses.

My second attempt is:

If there is a possible world w at which the observable facts are such that it is objectively reasonable to conclude that p is true at w, then we should conclude that p is possibly true.

---

[7] "To imagine an object as determinate is to imagine it as possessing the higher-order property … of possessing a determinate property for each of its determinables. There is a world of difference, then, between imagining an object as determinate—as possessing determinates for each of its determinables—and determinately imagining it—specifying in each case what the underlying determinate is" (Yablo, 1993, p. 28).

By looking to observable facts[8] on the left side of the conditional, we are explicitly accommodating those who claim that only the observable is imaginable, and ensuring that anything imagined here has a basis in *qualitative* imagining, to accommodate Kung's, (2010) claim mentioned above, that stipulative imagining lacks probative force.[9] Referring to a possible world, rather than an imagined one, draws on the importance of agreement from both parties to the debate (even if imagination is the source of this agreement). Claims making up the world referred to on the left of the conditional must be agreed by both parties in the debate in question: the principle has no bite unless such agreement is reached (as we saw above, Berkeley and Mach's refusal to grant Newton's description of the case stopped his argument in its tracks). And by the phrase 'objectively reasonable' I intend to cover inferences to the best explanation, but only those that are drawn rationally, where all relevant facts have been considered and the inference is not to the truth of "the best of a bad lot" (van Fraassen, 1989, p. 143; Lipton, 2004, pp. 151–163). This parallels the actual practice of drawing an inference to the best explanation: this can only be rational where one thinks that they have considered all relevant evidence. We have also now omitted reference to a possible observer, meaning our principle covers cases like Newton's two-sphere case, which crucially lacked any objects other than the two spheres (another example might be cases of putative distinct indiscernibles, which must lack a symmetry-breaking observer).[10]

The problem here is that this tells us we ought to conclude that p is true even in cases where we have not considered any such world. That is, the mere fact that *w* is a possible world would require us to draw an inference about it, whether we have given any thought to *w* and p or not! This can be remedied by amending the start of the principle as follows:

---

[8] What counts as observable facts here can be thought of as turning on what the qualitative contents of our imaginings are. Whether these ought to include relatively high-level properties, such as being a cat, or exclusively colour-at-location depends on one's theory of perception. As we saw in Sect. 2, following Kung, (2010), I distinguish *qualitative* from *assigned* contents, and take being a cat to fall under the latter category. However, I admit that alternative theories of perception could make the principle less useful—for instance, if one includes modal properties among the qualitative contents of percepts and imaginings, then perhaps we need not go beyond experience or imaginings of machines to determine whether they could think. The kinds of possibilities for which such inferences are useful are those for which confirmation by experience alone fails; if the qualitative content of an imagining alone would settle the issue, then an ampliative inference would be obsolete, as in the case of actual IBE and perception. While it is beyond the scope of this paper to provide a fleshed out theory of perception, I take it that the passage of time without change, the presence of absolute space and thinking machine are not merely 'observable facts', i.e. things which could be determined by looking. What is crucial for my purposes is that, when the principle is brought to bear in a disagreement, both parties agree on the possibility of the case as it is described; stipulation can play a significant role in fleshing out the case, but assuming that stipulation alone cannot determine what is possible, any such stipulations should in principle be traceable to the stripped down, colour-at-location qualitative content. Thanks to an anonymous reviewer for raising this issue.

[9] Recall that, for Kung, qualitative imagining presents a way that space could be filled around the perceiver (Kung, 2010, p. 637).

[10] The goings-on in the two-sphere case can nonetheless be classified as "observable facts"—in the sense that they involve concrete objects that would leave an impression, were there an observer there to see them; it may also be permitted that a disembodied observer is present in the case, depending on the philosophical beliefs of the interlocutor.

> If *we consider that* there is a possible world *w* at which the observable facts are such that it is objectively reasonable to conclude that p is true at *w*, then we should conclude that p is possibly true.

However, our principle is now rather trivial: it effectively says that, if we think that we should conclude that *p* is true at *w*, then—since *w* is a possible world—we should conclude that *p* is possibly true.

I therefore suggest that we characterise the three thought experiments with the following principle:

> If there is a possible world *w* at which the observable facts are such that it is objectively reasonable to conclude that p is true at *w*, then p is possibly true.

Here 'we should conclude that' has been removed from the consequent, therefore removing the need for restricting the principle to those worlds under our consideration. At what cost? It may be said that, given that even idealised IBE is fallible, the *epistemic* fact that it is objectively reasonable to conclude that p is true at a world should not entail the *metaphysical* fact that p is true at that world.[11] In fact, this does not contradict the principle as stated; I will return to this in the following paragraphs, but first, a brief discussion of the supposed barrier to implication, from the epistemic to the metaphysical. The phrase 'objectively reasonable' is intended to encode for rationally drawn inferences to the best explanation (whereby hypotheses are compared by way of principles of theory selection such as Occam's razor) which are drawn from an appraisal of all relevant facts, my assumption being that it is objectively reasonable to infer to the truth of such an explanation. If the claim that a certain hypothesis is the best among all other viable candidates is itself an epistemic claim, then the viability of IBE more generally is threatened, for the latter involves not merely inferring to a claim about what it is reasonable to believe; rather, it involves making the leap to claims about what is the case, often in matters of ontology, and hence metaphysics. Of course, it is a legitimate position to deny the validity of inferences to the best explanation; my claim here is that, if we allow for inferences to the best explanation based on an appraisal of actual goings on, then, provided certain conditions are met, we should treat merely *possible* goings on similarly. I intend the stated principle to be a normative one, but the normative aspect comes from my saying that we ought to apply the principle, which is premised on the norm of drawing IBEs based on appraisals of actual phenomena. IBE is indeed fallible, and so there is no *guarantee* that the above principle will turn up true possibility claims, just as there is no guarantee that IBE will turn up truths about the actual world; nonetheless, one should apply the principle provided one engages in the general practice of drawing abductive inferences.

The reader may have noticed that the right side of the principle affirms a possibility without affirming it of the specific world imagined. It may have been expected that such inferences would be about the specific world imagined, given what has been said above about imagining our scenarios as part of complete worlds, and as being determinate (in spite of not being determinately imagined). Thus, with our principle so stated, in the case of the Turing test, the inference is to the possibility of thinking

---

[11] Thanks to an anonymous reviewer for raising this.

machines—not to the claim that there is a thinking machine specific scenario/world imagined. There are two reasons for characterising the principle in this way.

First, any such inference to a claim about a specific world could be said to be far too unlikely to be warranted. As our cases involve ampliative inferences, the evidence from which these inferences are drawn is consistent with the falsity of the conclusions drawn. In the case of Shoemaker's scenario, for example, it is entirely consistent with the description of the case that there are in fact no periods of total freeze (which is to say, that the regularity of the freezes in each region is reliably scuppered every sixtieth year). To say that the best explanation of Shoemaker's scenario is that there is a year-long total freeze—a period of time without change—every sixtieth year may not allow us to say, of any particular world conforming to Shoemaker's description, that it is *likely* that there is a total freeze. When reasoning about particular possible worlds, we are effectively in barn-façade territory.[12]

Secondly, as noted, IBE is fallible—the best explanation *need not* obtain—and so the fact that the best explanation of some phenomena at *w* is *p* should not entail that p is true *at w*. To see why, consider that affirming, of any world conforming to Shoemaker's description, that it is a world in which there are periods of changeless time (and likewise for the other two cases and their intended conclusions) would prove too much. There is of course an infinite number of such worlds, varying in irrelevant respects (such as the number of items in each region); while all such worlds may differ from each other in any number of respects, each has its membership in that class by conforming to Shoemaker's stipulations. But, if we should say of *each* of these worlds that time passes without change in that world, then, surely, we should say that they are *all* worlds in which time passes without change. This conflicts with the fact that the observable evidence in Shoemaker's thought experiment is *consistent* with there being no periods of changeless time.[13] What I am proposing for possible worlds is therefore weaker than IBE—strictly speaking, not to infer to the *truth* of the best explanation at a particular world under consideration, but to the *possibility* of the best explanation—its truth at *some* world conforming to the stipulations of the case.[14]

It is interesting to note that, given that such inferences are ampliative, the rationality of concluding that p is possibly true does not depend on it being *true* at the world we are considering. It does however need to be true at *some* world to count as modal knowledge. This suggests that we can gain modal knowledge from a consideration of worlds that make the proposition in question false: the justification of the proposition and the truth of the proposition need not be contributed by the same worlds. Again, with regard to the Turing test, this is how it should be: characterising the principle as we have done allows us to claim that the presence of a thinking being is the best

---

[12] Having said this, Biggs and Wilson, (2017) have convincingly argued that the epistemic value of abductive principles does not depend on whether theories satisfying these principles are more likely to be true.

[13] Faraci, (2013) raises this problem, as a modal analogue of the lottery paradox, albeit with regard to a different debate—concerning the objectivity of moral claims. Faraci proposes dropping the agglomeration rule, such that asserting a, b, c, etc. should not force us to assert their conjunction. By inferring to *possibility*—truth at *some* world—rather than truth *at a particular world*, I mean to avoid committing to this move.

[14] Nonetheless, IBE is a realist mode of inference—it moves from epistemic considerations to truth claims; as such, the principle I have stated warrants inferences not merely to *justified belief* about modal space, but to what that space includes.

explanation for any case in which a machine passes the test, while leaving open the *possibility* that a machine might win the imitation game—'fooling' the human—while nonetheless *not* being a thinking being. This fits with an interpretation of the Turing test thought experiment that doesn't claim, implausibly, that winning the imitation game *entails* that one is a thinking being.

6 Genuinely new ideas

In their introductory chapter to their *Conceivability and Possibility* volume, Gendler and Hawthorne raise the following intriguing question: "can imagination/conception enable us to gain access to genuinely new ideas? Or do they merely *re*-present, albeit perhaps in a modified form, what we are already acquainted with through other means, such as perception?" (Gendler & Hawthorne, 2002, p. 9). We have met the idea that stipulation can be combined with sensory imagination, but as we saw, it is highly questionable whether possibility claims arrived at in this way have any argumentative force.

What about recombination more generally? Specifically, consider the principle of recombination, which states "that anything can coexist with anything else, at least if they occupy distinct spatiotemporal positions. Likewise, anything can fail to coexist with anything else" (Lewis, 1986, p. 88). Behind this lies another plausible principle, the Humean dictum, that there are no necessary connections between distinct existences. However, often modal disputes are primarily about whether certain things really are distinct, where the possibility of separation cannot be presupposed without question-begging (we see this, for example, in the debates over whether time = change, or whether identity *just is* indiscernibility). In such cases, where the identity or distinctness of some things is what is at issue, the principle of recombination offers little help; for example, recombination cannot tell us whether time could exist without change until we find out whether time and change are the same or distinct. In some cases, we might appeal to imagination or conceivability to help us decide whether some things are distinct, and then let the principle of recombination do its work. As Schaffer has noted, "[r]ecombination and conceivability are interrelated. Part of the justification for recombination is the conceivability of what results" (Schaffer, 2005, p. 12). But one will only grant the conceivability of what results if one grants the conceivability of the elements to be recombined. Thus there is a certain limit to the capacity of recombination to resolve disputes about what is possible. It seems fair to say that recombination counts as re-presenting, in a modified form, what we are already acquainted with.

As I suggested above, my aim is to forward a modal epistemology that has the capacity to extend our modal horizons, perhaps beyond the limits of what can be upheld by appeal imagination (whether sensory or stipulative) alone. In this connection, we now turn to a comparison of what is perhaps the closest precedent, Hanrahan's, (2007) modal explanationism.

Hanrahan's modal explanationism has it that we can justify claims about possible worlds by thinking about possible evidence, evidence conceived of through the combination of sensorial imagining and storytelling. In particular, the kind of evidence

relevant to Hanrahan's project consists in non-factive mental states; just as one can have a bear represented to one's mind without there actually being a bear there, when one conceives of a possible perception *as of* there being a bear in one's backyard, we have not necessarily conceived of a possible world in which there *is* a bear in one's backyard. Possible percepts can be misleading, just as actual percepts can. This is crucial, if the aim is to justify the claim that imagination is a guide to possibility, for it would beg the question to assume that whenever we imagine a possible percept, the world imagined is one in which that percept is veridical (Hanrahan, 2007, p .136, p. 142).

Modal claims are then justified by comparing the merits of competing potential explanations of possible percepts. Hanrahan asks us to consider her epistemic twin, Rebecca, who has the same mental life as Hanrahan up to the point where Hanrahan intentionally conjures up an image of a bear in her backyard; Rebecca, on the other hand, has an image of a bear in her backyard come to her unbidden. Hanrahan herself has recently heard that there was a bear sighting on someone's porch, and she notes that her neighbours keep many bird feeders stocked up with feed, and that this can be a draw for bears.

Hanrahan argues that her epistemic twin is justified in taking her images as of a bear to be veridical, and that this in turn justifies Hanrahan in believing that (B) *A bear is in Hanrahan's backyard*, is possibly true. It is argued that the mental state of Hanrahan's epistemic twin, and the account that the latter gives of that state, satisfy three requirements, sufficient to justify the epistemic twin in taking her mental representations as of a bear to be veridical:

1. The account of that state as a veridical perception must possess all of the epistemic virtues to a greater degree than any of the other accounts we could have given of that state and it must also increase or at least preserve the epistemic virtues of our best explanation over the long run.
2. The account of that state as a veridical perception must, of course, include or imply that the state is a product of the normal workings of our senses which accurately reflects the way the world is.
3. The state must be sensorial in nature and it must usually (*though not always*) be highly forceful, vivacious, and determinate. (Hanrahan, 2007, p. 130)

Hanrahan justifies the satisfaction of requirements 1 and 2 as follows:

Now because Rebecca possesses nearly all the beliefs I possess, all the information I have recently gathered about bears is available to her. Thus, I know that the story she can tell about these images that best preserves the epistemic virtues of her best explanation is one that deems them to be veridical perceptions. In this story, this bear had in the past found a mound of birdseed to dine on in Rebecca's neighbor's backyard. And this bear (being a creature of habit) returned to these bird feeders (via Rebecca's yard) to see if there was more that very morning as Rebecca was looking out her back window. Thus, requirements 1 and 2 above have been satisfied. (Hanrahan, 2007, pp. 137–138)

As for requirement 3, its satisfaction depends on how vivid Hanrahan's own imagery was, for, being epistemic twins, Hanrahan's and Rebecca's mental states are the same

on this score. If it is granted that Hanrahan's sensorial imagination of a bear in her backyard is highly forceful, vivacious and determinate, then it must also be granted that her (possible) epistemic twin's mental representation of a bear also possesses these qualities. But given that some people's imaginations can manifest these qualities to a high degree, we should grant the former point, and hence that 3 is satisfied by her epistemic twin's mental state (Hanrahan, 2007, pp. 138–139).

The next step in the argument is to show that, because Hanrahan's epistemic twin is justified in taking her mental state to be veridical, and thus in believing that there is a bear in her backyard, Hanrahan is also actually justified in the claim that there could be a bear in her backyard. Hanrahan does this by noting, first, that Rebecca is justified in taking her image to be a veridical one, and, but for the fact that Rebecca's image came to her unbidden, they are epistemic twins; as such, it seems only reasonable that Hanarahan would draw the same conclusion as Rebecca were the former in the latter's shoes. Secondly, Hanrahan notes that we have been given no reason to think that Rebecca was intoxicated, or her vision obscured or impaired, or that there is a holographic representation in her garden; and since Rebecca's imagery comes to her unbidden, it makes more sense to conclude that her mental state is veridical, than it does to take it to be a product of her imagination, as Hanrahan's was (Hanrahan, 2007, pp. 140–141).

A significant point of analogy between Hanrahan's view and the present paper is that both are concerned to circumvent what Hanrahan terms 'the problem of adjudication', where competing theorists fail to agree on whether something is imaginable or conceivable—as happens in cases of imaginative failure, discussed above. A prominent case is the claim that philosophical zombies are conceivable, and therefore possible—that there is a possible world just like ours except that our counterparts are without consciousness (Chalmers, 1996, p. 96). One may deny the intuition, or, alternatively but to similar effect, one may have an intuition of something else that contradicts it. In an example of the latter, Frankish (2007) mounts an argument against the conceivability of zombies, from an intuition that conflicts with it: the conceivability of "anti-zombies". Anti-zombies are beings very much like zombies would be (if they were possible), except that the purely physical facts about them necessitate their consciousness. But there is no room in modal space for both zombies and anti-zombies, for, if the purely physical features of anti-zombies necessitate their consciousness, they therefore make any would-be zombies conscious too. Frankish suggests that the result is a stand-off, and hence "that considerations of conceivability have little role to play in debates about the nature of consciousness" (Frankish, 2007, p. 666).

Hanrahan's claim is that adjudication is *possible* on modal explanationism, and not possible on modal intuitionism. While in *some* cases disagreement may persist, on which is the most fruitful, coherent, and simple theory, there will at least be some cases where the abductive success of a theory is enough to overcome some original misgivings about the concepts appealed to in that theory, or misgivings based on the extent to which that theory coheres with previously held beliefs. That is, it is at least possible for opposing parties to evaluate each other's theory based on abductive

criteria; it is not clear that the same can be said for apparently opposing modal intuitions (Hanrahan, 2007, p. 145).[15]

Another point of analogy between Hanrahan's and my own route to possibility claims is that we both assume some possibilities in order to infer others. With this in mind, Hanrahan considers the following objection:

> I want to justify my belief that B is possibly true. In order to provide myself with this justification, I have made an assumption that I have tried to defend. But in this defence, I have employed some of my beliefs about what is possible. For example, my above argument depends on the claim that the images we produce via our imagination could have come to us via an alternative pathway. So, I am seemingly trapped in a circle. (Hanrahan, 2007, p. 141)

A similar objection might be put to me: in any argument from possible phenomena to a possibility claim, the possibility of at least some of that phenomena must be assumed. For example, in Shoemaker's case of evidence for the possibility of time without change, we must assume that his world of partial freezes (apart from whether it is a world where every sixtieth year is one of total freeze) is itself possible.

Hanrahan responds that the circle is not a vicious one.

> When we seek to establish that a proposition is possibly true, we need not start from scratch, throwing out all of our modal beliefs. We only need to avoid assuming the possibility of the proposition whose modal status is in question. This I have done. (Hanrahan, 2007, pp. 141–142)

The trouble is that Hanrahan goes on to tentatively endorse a "limited skepticism about possibility", based on the limitations of her own modal epistemology, namely that "[b]y design, it is radically ill-equipped to deal with possible worlds that are radically different from this [i.e. the actual] world" (Hanrahan, 2007, p. 143). If the suggestion is that we should be sceptical of modal claims that cannot be justified by Hanrahan's method, then it remains problematic that Hanrahan herself appeals to such modal claims—in particular, the claim that the images we produce via our imagination could have come to us via an alternative pathway (which forms part of the scenario to be explained). One might think that the latter claim is justified by everyday experience; but then why not say the same of other mundane modal claims like the one that Hanrahan justifies using her method (that it is possible for there to be a bear in her backyard)? As Hanrahan notes, the kinds of possibilities that Hanrahan's epistemology aims to justify are mundane, the possible inferences she refers to based on "variations in our immediate surroundings" (Hanrahan, 2007, p. 143). Thus the modal claims she takes for granted are on a par, as far as the need for justification

---

[15] It should be noted here that Chalmers acknowledges the difficulty of adjudicating conceivability arguments, "particularly where strange ideas such as these [i.e. zombies] are concerned" (1996, 99), and goes on to offer alternative arguments against the logical supervenience of consciousness on the physical. It should also be said that, while Chalmers finds the logical possibility of zombies "obvious" (1996, 96), he goes beyond the mere affirmation of intuition, describing a system functionally isomorphic to that of a consciousness-producing brain to provide indirect support for the possibility of physical replicas lacking in consciousness (97). I would therefore not go so far as to say that adjudication is impossible for the modal intuitionist: it is possible, provided they are willing to extend their argument in such ways—indeed, explanationist considerations can be counted among these.

is concerned, with those she seeks to justify. If the sceptic is suspicious of a claim like "The image of a bear could have come to me by an alternative pathway", then any conclusions drawn therewith will be equally suspect; but if the interlocutor is not so sceptical, then explanationism has no work to do—we could just as well infer the possibility of a bear in the backyard from imagining it directly, rather than going to the trouble of inferring to the best explanation of some imagined phenomena.

That is why I set my principle up for use in dialectical contexts of limited imaginative failure, rather than those in which the imagination is hardly admitted as a source of *any* modal beliefs. I do not present my route to possibility claims as a *general* alternative, to supplant all others,[16] nor as an antidote to a general modal scepticism that would challenge even the most mundane of modal claims; but in disputes where imaginative resistance is a factor, appeal to possible evidence can make some headway where simple appeals to imagination or modal intuition of the possibility in question are inappropriate. If I make use of imagination to justify the possibility of some evidential situation, I do not assume what is to be shown, for I do not aim to justify the claim that imagination is a guide to possibility; thus I can, without circularity or redundancy, appeal to imagination or modal intuition to make modal claims regarding possible phenomena, in order to consider what conclusions it would be reasonable to infer. Again, in terms of the kind of dialectic I have been concerned with, this makes sense, for those who take imaginative resistance to be a barrier to justification of possibility have effectively granted imagination as a guide; that is, the idea that imaginative resistance is a barrier to the justification of a given modal claim presupposes that the imagination *would have* provided some justification, were it successful.

To summarise, the context of general modal scepticism works both for and against Hanrahan's project: for, because it motivates explanationism—such scepticism means that imagination is not enough, and even the most mundane of modal claims (namely, 'There could be a bear in my backyard') requires justification that goes beyond the assumption that what is imagined is generally possible; but against, because in any argument from explanation, the explanandum must be granted—yet any context in which a modal claim as mundane as 'There could be a bear in my backyard' is treated with scepticism will be one in which 'The image of a bear could have come to me unbidden' is equally suspect. By contrast, the kinds of inferences which I am advocating are inherently ambitious, tending towards claims that may not be reachable by other means. I take it for granted that perception can be a guide to actuality, and imagination a guide to possibility; after all, IBE needs raw data to draw from (though there is plenty of room for disagreement about the extent of imagination's reach—one may be more inclined to grant the possibility of Turing's case than Shoemaker's, the latter requiring physical goings on quite alien to what we find in actuality). In any case where a hypothesis could be confirmed by observation, it would make sense to strive towards such observations; it is with respect to that which cannot be otherwise verified that IBE is often most appropriate. Similarly, where a possibility claim can be justified by imagination, it is to imagination that we should turn; it is where imagination falters that I find the perception-imagination analogy to be most intriguing: if IBE, operating

---

[16] Thus I can be said to be, in Wirling's, (2020) terms, a non-uniformist about modal epistemology.

with actual phenomena, can transcend the limits of perception, then perhaps, operating with merely possible phenomena, it can similarly transcend the limits of imagination.

7 Conclusion

Kripke famously claimed that "[g]enerally, things aren't 'found out' about a counterfactual situation, they are stipulated" (Kripke, 1981, p. 49); "[p]ossible worlds' are stipulated, not discovered by powerful telescopes" (Kripke, 1981, p. 45). It has also been suggested (blocking the inference of necessity from a priority, as in the knowledge argument for property dualism) that one cannot legitimately draw metaphysical conclusions from epistemic premises (Kripke, 1981, p. 36). Concomitant to its potential to expand our conceptual horizons, a modal epistemology that warrants ampliative inferences from possible phenomena to possibility claims would represent a departure from these assertions. New concepts are naturally faced with scepticism, but may gain acceptance as they figure in theoretically virtuous explanations. We can see Turing's test, for example, as attempting to break free from conceptual arguments, by showing that an alternative concept of machines (as things that could think) is involved in the best explanation of certain possible data. This idea that theory success is a source of concept advancement is captured by Sider in his *Writing the Book of the World*:

> A good theory isn't merely likely to be *true*. Its ideology is also likely to carve at the joints. For the conceptual decisions made in adopting that theory – and not just the theory's ontology – were vindicated; those conceptual decisions also took part in a theoretical success, and also inherit a borrowed luster. So we can add to the Quinean advice: regard the ideology of your best theory as carving at the joints. (Sider, 2011, p. 13)

In this way, explanation can play a crucial role, for sensory imagination and stipulation would lack probative force in a dialectic where not just the imaginability of a world verifying a proposition is in doubt, but also where the legitimacy of a certain concept is a point of contention. To give a striking example from the history of philosophy and science, Locke was dismissive of the notion of action at a distance in earlier editions of his *Essay*, but revised his attitude in light of the success of Newton's theory of universal gravitation. Interestingly, he nonetheless maintained that action at a distance was beyond his comprehension, but granted its possibility in spite of this. In a letter to Stillingfleet, he wrote:

It is true, I say, "that bodies operate by impulse, and nothing else". And so I thought when I writ it, and can yet conceive no other way of their operation. But I am since convinced by the judicious Mr Newton's incomparable book, that it is too bold a presumption to limit God's power, in this point, by my narrow conceptions. The gravitation of matter towards matter, is not only a demonstration that God can, if he please, put into bodies, powers and ways of operation, above what can be derived by our idea of body, or can be explained by what we know of matter, but also an unquestionable and everywhere visible instance, that he has done so. And therefore

in the next passage of my book I shall take care to have that passage rectified (Locke, 1823, Volume 4, pp. 467–8)

Scientists need to consider what their experiments would show, if certain conditions were met and certain underdetermining results were obtained. It is likely that in some cases there will be imaginative failure regarding the phenomena being tested for. Locke's claim that action at a distance cannot occur because it is inconceivable, revised upon consideration of the success of Newton's theory of universal gravitation, is echoed in the development of quantum mechanics, where theory success perhaps outstrips our powers of conception. In some cases, there may be contingent limitations on the capacity to design or execute experiments that have nothing to do with the truth of the proposition per se (just as our inability to conjure up imagery corresponding to a chiliagon has nothing to do with the possibility of thousand sided objects); in other cases, there may be *necessary* limitations on the capacity to execute experiments that may have nothing to do with the *possibility* of the proposition per se (as we noted in the cases of Shoemaker's freezes and Newton's spheres). Inference to the best explanation of imagined phenomena can enable us to determine what is possible, in the face of imaginative resistance. Just as a scientific realist will draw inferences to claims about unobservables, similar patterns of reasoning can allow us to reach beyond sensory imagination, and beyond the dialectical limits of stipulation, to justify possibility claims about the unimaginable.

This is all to carve out a niche for inferring to the possibility of the best explanation of imagined phenomena. Specifically, this is appropriate in a discussion where imagination is assumed to provide evidence for possibility in general, but where there is a principled reason why the target phenomenon either cannot be imagined, or can, but only in such a way that provides no evidence for possibility (as in the case where stipulation does all the work); in other words, it is useful where the imaginability of the phenomena under scrutiny is part of what is at issue, such that appeal to its imaginability would be dialectically inappropriate. These points are related, for in the context of a debate about a possibility where imaginative resistance is a sticking point, there is likely to be tacit acceptance of the idea that imagination is a guide to possibility: were it imaginable, this would have favoured the case for its possibility. Returning to the perception-imagination analogy, we might also note that unobservability is analogous to unimaginability. Thus, just as scientific realists allow inferences to reach beyond what can be directly confirmed by perception—that is, to the unobservable—by drawing ampliative inferences based on imagined phenomena, we can reach beyond what can be shown directly by imagination, beyond the sensorily imaginable. Thought experiments like our three cases do not depend on the imaginability of what they aim to demonstrate, and therefore offer a way of moving the debate forwards, beyond the impasse of disagreement about what is imaginable.

References

Biggs, S. (2011). Abduction and modality. *Philosophy and Phenomenological Research, 83*(2), 283–326.

Biggs, S., & Wilson, J. (2017). The a priority of abduction. *Philosophical Studies, 174*(3), 735–758.

Brown, J. R. (2004). Why thought experiments transcend experience. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science* (pp. 23–43). Malden, Mass: Blackwell.

Chalmers, D. (1996). *The conscious mind*. Oxford University Press.

Faraci, D. (2013). Brown on mackie: Echoes of the lottery paradox. *Philosophia, 41*, 751–755.

Frankish, K. (2007). The anti-zombie argument. *Philosophical Quarterly, 57*, 650–666.

Gallois, A. (1974). Berkeley's master argument. *The Philosophical Review, 83*(1), 55–69.

Gendler, T. (2013). Imagination. *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=imagination

Gendler, T. S., & Hawthorne, J. (2002). Introduction. In T. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 1–70). Oxford: Clarendon Press.

Hanrahan, R. (2007). Imagination and possibility. *Philosophical Forum, 38*(2), 125–146.

Hart, W. D. (1988). *Engines of the soul*. Cambridge University Press.

Kripke, S. (1981). *Naming and necessity*. Blackwell.

Kung, P. (2010). Imagining as a guide to possibility. *Philosophy and Phenomenological Research, 81*(3), 621–663.

Lewis, D. (1986). *On the plurality of worlds*. Blackwell.

Lipton, P. (2004). *Inference to the best explanation*. Routledge.

Locke, J. (1823). *The Works of John Locke, a new edition, corrected, ten volumes.* London.

Martin, M. G. F. (2002). The transparency of experience. *Mind & Language, 17*(4), 376–425.

Moran, R. (1994). The Expression of feeling in imagination. *The Philosophical Review, 103*(1), 75–106.

Newton, I. (1966). *(Principia), Mathematical Principles of Natural Philosophy*, trans. F. Cajori. Berkeley, CA: University of California Press (original work published 1687).

Oppy, G. (2004). Can we Describe Possible Circumstances in which we would Have *Most* Reason to Believe that Time is Two-Dimensional?. *Ratio (new series)* XVII 1 March 2004: 67–83.

Oppy, G., & Dowe, D. (2020). The Turing Test. *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/turing-test/>.

Parsons, J. (2013). The Earth and the Aleph. Manuscript. Available at https://philpapers.org/archive/PARTEA-3.pdf . Accessed 28/12/20.

Peacocke, C. (1985). Imagination, experience and possibility: A Berkelian view defended. In J. Foster & H. Robinson (Eds.), *Essays on Berkeley* (pp. 19–35). Oxford University Press.

Putnam, H. (1975). *Mind, Language and Reality, Philosophical Papers ii*. Cambridge University Press.

Schaffer, J. (2005). Quiddistic knowledge. *Philosophical Studies, 123*(1–2), 1–32.

Searle, J. (1981). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–457.

Shoemaker, S. (1969). Time without change. *Journal of Philosophy, 66*(12), 363–381.

Sider, T. (2011). *Writing the book of the world*. Oxford University Press.

Sorensen, R. (1992). *Thought experiments*. Oxford University Press.

Sorensen, R. (2016). Thought Experiments and Imagination. In *The Routledge Handbook of Philosophy of Imagination*. Routledge

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford University Press.

Weatherson, B. (2004). Morality, fiction, and possibility. *Philosophers' Imprint, 4*(3), 1–27.

Wirling, Y. S. (2020). Non-uniformism about the epistemology of modality: Strong and weak. *Analytic Philosophy, 61*(2), 152–173.

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research, 53*(1), 1–42.