MODAL REALISM AND MODAL EPISTEMOLOGY: A HUGE GAP

Otávio Bueno & Scott Shalkowski*

Department of Philosophy, University of South Carolina
* School of Philosophy, University of Leeds

ABSTRACT. According to the modal realist, possible worlds exist, and in terms of them, it’s possible to articulate a systematic approach to the theory of modality (Lewis 1986). Given, however, that we have no access to such worlds, how can we know that they in fact exist? To answer this question, Lewis developed two modal epistemological strategies. First, he explored an analogy with mathematics, noting that we do have mathematical and modal knowledge, both being cases of a priori knowledge. Second, he defended a theoretical utility argument, insisting that the theoretical utility of postulating possible worlds is a good reason to believe in the existence of these objects. In this paper, we examine critically both strategies, and argue that they fail. We then sketch an alternative way of developing a modal epistemology, and argue that, because it’s not committed to possible worlds, it doesn’t face the problems faced by Lewis’ proposal.

1. INTRODUCTION

According to modal realism, there are possible worlds, and in terms of such worlds, a systematic analysis of modality – among other notions – can be developed (see Lewis 1986: 1-69). Given the crucial role that the existence of possible worlds plays in modal realism, the issue arises as to how we can know that there are such worlds. Lewis is, of course, aware of the issue, and he provides two moves to ground his modal epistemology (see 1986: 3-5 and 108-115). In this paper, we critically assess these moves, and argue that they fail. In the end, the modal realist still has to provide a viable modal epistemology. In response, we outline an alternative way of developing the epistemology of modality – one in which possible worlds play no role. In fact, according to this account, we do have modal knowledge, but this knowledge neither presupposes nor can be identified with knowledge of the existence of possible worlds.

A caveat: for reasons of simplicity, we will focus here only on knowledge of the claim that there are possible worlds. We won’t be addressing the issue of how we can know the content of specific modal claims, such as whether there are worlds with talking donkeys, flying pigs, or in which the continuum hypothesis fails. For the purposes of this paper, we will focus, with Lewis, only on the more general issue of our knowledge of whether there are or aren’t possible worlds.

1 Our thanks go to John Divers, Joseph Melia, Tim De Mey, and Gideon Rosen for extremely helpful discussions and suggestions.
2. LEWIS' MODAL EPISTEMOLOGY

To develop an epistemology for modal realism, Lewis provides two arguments: one is based on an analogy with mathematics and on how we come to have mathematical knowledge; the other explores the theoretical utility of postulating possible worlds, and uses this as a reason to believe in the existence of such worlds. We will examine each of these arguments in turn.

2.1 The argument based on the analogy with mathematics

Lewis starts articulating his epistemology for modality with an analogy with mathematical knowledge. This is, at face value, a “common origin” argument; in the sense that both mathematical epistemology and modal epistemology have the same “source”—they are particular cases of a priori knowledge (1986: 110-113). But more than simply having the same source, what grounds the adequacy of mathematical and modal knowledge is, in the end, their a priori nature. Clearly, we have no access to mathematical objects and their relations, and such objects, of course, bear no causal relations with us. Mathematical objects are not located in space and time, and are not the kind of things to which we expect to have any causal access. If we follow the mathematical realist, who postulates the existence of mathematical objects, how can we explain the possibility of mathematical knowledge? In response, it's not surprising that mathematical knowledge is a priori. What else could it be (given that we have no access to the relevant objects)? Furthermore, no one would deny that we do have plenty of mathematical knowledge. As Lewis points out, however we may obtain mathematical knowledge, we have more reason to think that we have such knowledge than to accept any epistemology that entails the nonexistence (or high improbability) of mathematical knowledge (1986: 109).

We don’t think, however, that this move provides an epistemology for mathematics. At best, it gives a principle to rule out mathematical epistemologies. So, the issue is still open as to how we come to have all the mathematical knowledge that we are presupposed to have. It’s simply not enough to point out that any mathematical epistemology unable to accommodate the possibility (or plausibility) of the knowledge of mathematics cannot be accepted. One still needs to articulate an epistemological account of mathematics. A simple mention of a priori knowledge is not enough. This just identifies the kind of knowledge we are examining; it says nothing, for instance, regarding its reliability.²

According to Lewis, the crucial points above also apply to modal epistemology. After all, there is no doubt that we have plenty of modal knowledge, that is, knowledge of what is possible, necessary, and actual.³ Moreover, we have more reason to believe that we are in possession of such modal knowledge than to accept an epistemological account that entails the lack (or the high implausibility) of modal knowledge. And similarly to mathematical knowledge, modal knowledge is a priori. What else could it be? We have no access to possible worlds (see 1986: 109-110).

² Not to mention the fact that mathematical knowledge is typically taken to be the paradigmatic case of a priori knowledge. If this is assumed, then to invoke a priori knowledge as a way of providing an epistemological account of mathematics comes very close to begging the question.

³ Of course, in Lewis’ case, assuming modal realism, knowledge of what is possible, necessary and actual involves having knowledge of the existence of possible worlds. As we will discuss at the end of this paper, without modal realism, modal knowledge need not invoke possible worlds.
We don’t think, however, that it’s enough simply to have a principle to rule out epistemological accounts of modality. One still needs to articulate a proposal that explains how is it possible to have the modal knowledge we think we have. In other words, a modal epistemology is required for modal realism. Given that, in the case of mathematics, the main challenge for mathematical realism emerges at the level of epistemology, the same difficulty is found with modal epistemology. As a result, the modal realist has to provide an epistemological story to accompany modal realism. It doesn’t suffice simply to suggest a criterion to reject unacceptable epistemological accounts, nor is it enough simply to point out that modal knowledge is a priori knowledge – at least an examination of the reliability of the proposed account is required.

2.2 The argument based on theoretical utility

The second, and most general, argument provided by Lewis in support of his modal epistemology brings in the theoretical utility of postulating possible worlds (see Lewis 1986: vii and 3-5; see also Divers 2002: 151). The key idea is that the theoretical utility of the possible worlds hypothesis gives us good reason to believe that such a hypothesis is true. According to Lewis, this is precisely the situation with the modal realism hypothesis: undeniably, it’s theoretically useful (see 1986: 1-69). Thus, the argument goes, we have good reason to believe in the existence of possible worlds. As Lewis stresses in the opening passages of *On the Plurality of Worlds*:

> I begin [...] by reviewing the many ways in which systematic philosophy goes more easily if we may presuppose modal realism in our analyses. *I take this to be a good reason to think that modal realism is true, just as the utility of set theory in mathematics is a good reason to believe that there are sets.* Then I state some tenets of the kind of modal realism I favor. (1986: vii; italics added.)

Note that we find here both the theoretical utility argument and the analogy with mathematics at work. On Lewis’ formulation, the two arguments are closely linked. That’s why we would resist any suggestion that the argument based on the analogy with mathematics may not be so fundamental. In a sense, the analogy with mathematics stands or falls with the theoretical utility argument. However, it’s important to note that the theoretical utility argument is the crucial argument. After all, the analogy with mathematics can itself be constructed as a theoretical utility argument. As Lewis notes, given that it’s theoretically useful to quantify over sets, we have here a good reason to believe that there are such objects – and *mutatis mutandis* for the existence of possible worlds.

More explicitly, in John Divers’ reconstruction, the theoretical utility argument can be formulated as follows:

(P1) If an ontological hypothesis has eminent utility (i.e. sufficient net utility and greater net utility than its rivals) then that gives us good reason to believe that it is true.

(P2) Such reason for believing an ontological hypothesis is warranting (i.e. given that the theory is true, having grounds of eminent utility is sufficient for knowing that the theory is true).

(P3) GR [genuine modal realism] has eminent utility.

Therefore, (C) We (are in position to) know that GR is true. (2002: 151)

---

4 For a discussion, see Divers 2002: 149-150. Divers doesn’t endorse this suggestion.
Although Lewis himself may not have presented the theoretical utility argument as explicitly as in Divers’ formulation, Divers is certainly right in attributing the above argument to the modal realist. After all, as we saw, Lewis does invoke the theoretical utility of set theory in mathematics as a reason to believe in the existence of sets. And, on his view, the theoretical utility of postulating possible worlds supports the truth of modal realism.

In fact, this brings in yet another way of conceptualizing the theoretical utility argument: as an *indispensability* argument (Quine 1960, Putnam 1971 and Colyvan 2001). According to this argument, the entities we ought to be ontologically committed to are just those that are *indispensable* to the best scientific theories we have developed. More explicitly, the *first premise* of the indispensability argument states that we ought to be ontologically committed to all (and only) those entities that are *indispensable* to our best theories of the world. Moreover, let’s suppose that we include, among our best theories of the world, our theories of modality. This is not an unreasonable assumption, given that science, scientific practice and our use of the vernacular include a huge amount of modal discourse. In this case, the *second premise* of the indispensability argument (as applied to modal discourse) insists that possible worlds are indispensable to our best theories. We then conclude that we ought to be ontologically committed to such worlds.

Now, in what sense are possible worlds *indispensable*? They are indispensable in the sense that, without quantifying over such worlds – conceptualized as genuine worlds, in the way Lewis does, and not as set-theoretic representations – we wouldn’t be able to construct a theory about modality that is as explanatory, as unified and with so much content. We could, of course, construct *weaker* theories that do not quantify over genuine worlds, but as Lewis points out, we would lose either in explanatory power or in expressiveness (Lewis 1986: 136-191). In this sense, to construct the more informative, unified and explanatory theory about modality, we *do need* possible worlds. This is the force of the *indispensability* argument. It’s not that we don’t have an option but to believe that there are possible worlds. There certainly are various options. But, according to Lewis, none of them is as attractive, or as systematic, as postulating genuine worlds.

Note that Lewis’ theoretical utility argument is *valid* – whether it’s reconstructed in Divers’ way or taken as an indispensability argument. So, to assess it, we need to determine whether its premises are true. This is one of the goals of what follows. But before doing that, let’s first assess Lewis’ argument based on the analogy with mathematics.

3. PROBLEMS FOR LEWIS’ EPISTEMOLOGY

3.1. Problems for the argument based on the analogy with mathematics

There are various difficulties with the modal realist’s analogy between mathematical and modal knowledge. We’ll discuss them in turn.

(i) If the analogy with mathematics is taken seriously, it may actually provide a reason to doubt that we have any knowledge of modality. One of the main challenges for platonism about mathematics comes from the epistemological front, given that we have no access to mathematical entities – and so it’s difficult to explain the reliability of our mathematical beliefs. The same difficulty emerges for modal realism, of course. After all, despite the fact that, on Lewis’ account, possible worlds are *concrete* objects, rather than abstract ones, we have no access to them. Reasons to be skeptical about a priori knowledge regarding mathematics can be easily “transferred” to the modal case, in the sense that difficulties we may have to *establish* a given mathematical statement may have a counterpart in establishing certain modal claims. For
example, how can we know that a mathematical theory, say ZFC, is consistent? Well, we can’t know that in general; we have, at best, relative consistency proofs. And the consistency of the set theories in which such proofs are carried out is far more controversial than the consistency of ZFC itself, given that such theories need to postulate the existence of inaccessible cardinals and other objects of this sort.

Similarly, how can we know the consistency of modal realism itself? Given the huge number of worlds that are postulated, it’s unclear how we could establish this result in general. And even if we could establish the result (presumably by constructing some sort of “model” for modal realism itself), this would require (a) a substantial amount of mathematical knowledge (in order to construct the appropriate “model”) as well as (b) the supposition that the existence of a model of a theory guarantees its consistency. There are, of course, limitations to the analogy with mathematics — any analogy only goes so far. And perhaps this is, in part, the reason why Lewis introduces the theoretical utility argument.

(iii) This may suggest that the argument based on the analogy with mathematics is just that: an analogy, and not much significance should be attached to it (for a discussion, see Divers 2002: 149-150). We disagree. The argument is important for Lewis’ purpose, given that, on his view, it provides a strong source of support for modal realism (see 1986: vii and 3-5). Moreover, the analogy with mathematics is apt, given that, in response to modal realism, there are those who attempt to reduce talk of possible worlds to talk of abstract entities, such as sets or propositions. (Lewis calls them ersatz modal realists; see 1986: 136-191.) As a result, the issue of how we can have knowledge of such abstract entities becomes particularly pressing for the ersatz modal realist. Given that, as Lewis acknowledges, he is trying to get a standoff (1986: 112), the move makes sense: the (genuine) modal realist is at least not worst off than the ersatz modal realist on the epistemological front — both need to explain the possibility of knowledge of abstract objects. But, alas, without the development of a proper epistemology — i.e. an epistemology that explains, in particular, a priori knowledge — in the end the move may fail to deliver the intended result.

(iii) For Lewis, we need not have any access to possible worlds to have knowledge of them. If access were required, there wouldn’t be any modal knowledge — in the sense of knowledge of possible worlds — given that, clearly, we have no contact with such objects (except, of course, for the actual world!). In this sense, as we saw, knowledge of the existence of possible worlds is similar to knowledge of the existence of mathematical objects — no access to the corresponding objects is required. Moreover, mathematical and modal knowledge are also similar in their necessity: both are knowledge of non-contingent matters, in the sense that mathematical and modal statements, if true, are necessarily true. So, for Lewis, to the extent that we have knowledge of the truth-values of these statements, we have infallible knowledge (1986: 114-115).

Nevertheless, how can we have knowledge of the truth-value of mathematical and modal statements? Clearly, not by directly inspecting the relevant objects (mathematical entities and possible worlds). A different strategy is required. The key idea is that we have mathematical knowledge by drawing (truth-preserving) consequences from (true) mathematical principles. And we have modal knowledge by drawing (truth-preserving) consequences from (true) modal principles, such as the principle of recombination⁵ (see Lewis 1986: p. 115). However, in each case, we could only have knowledge (let alone infallible knowledge) if the mathematical and modal principles in question were true. But how can we know that such principles are true? Again, not by “direct inspection”, given that we have no access to possible worlds or to

⁵ According to the principle of recombination, patching together parts of different possible worlds produces another possible world (see 1986: 87-88).
mathematical objects. Nor can we know the truth of these mathematical and modal principles by drawing consequences from the principles themselves, given that this wouldn’t validly establish the truth of the principles in question. And if we invoke additional modal and mathematical principles from which the original principles presumably follow, how can we know the truth of the former? Clearly, a regress starts here.

To avoid these familiar conundrums, one could claim that in both mathematics and in the theory of modality, there’s no need for one to be an internalist. To have knowledge, in each case, all we need is to have a reliable mechanism of belief formation; we need not know that the mechanism itself is reliable. As long as the mechanism is reliable, we will have knowledge. In particular, we need not know that the modal and mathematical principles that we use to generate our modal and mathematical beliefs are true. If they are true, and we use truth-preserving inference rules, the conclusions we obtain will be true, and we will have knowledge.

Of course, by following this strategy, we will not know everything that there is to be known about each domain (whether mathematical or modal). In fact, we know that we will not know everything that there is to be known about each domain. For as long as the language we use is rich enough to express some arithmetic (which is the case), and as long as the mathematical and modal principles we use are consistent (which, we hope, is also the case), by a Gödelian argument, there will be true sentences that cannot be derived from our mathematical or modal principles. But this is as it should be. To get incompleteness in our knowledge when that knowledge is incomplete isn’t a failure.

But does this mean that as long as Lewis is an externalist about modal knowledge – and doesn’t require that we know the truth of the modal principles that we invoke to derive our modal knowledge (such as the principle of recombination) – he will be safe on the epistemological front? Not really. For Lewis may need not have to show that his mechanism of formation of modal beliefs is reliable; but he still needs to have a reliable mechanism of belief formation. Can we say that drawing (truth-preserving) consequences from true mathematical or modal principles is a reliable mechanism of belief formation about mathematical and modal matters? Yes, if the principles in question are indeed true and the inference rules are in fact truth preserving. But how do we decide which mathematical and modal principles are true? Clearly, the externalist need not know that the principles in question are true, but he or she would have to select the true relevant principles. But it’s unclear how the externalist could articulate a modal or a mathematical epistemology in this way without knowing which of the modal or mathematical principles are true. Just pointing out that all we need are true modal or mathematical principles (e.g. the principle of recombination or the axiom of choice) isn’t enough to tell us which of various modal or mathematical principles are true. And if the principles in question were not true, then clearly the proposed mechanism of belief formation (drawing consequences from the relevant principles) wouldn’t be reliable. It’s difficult to see a way out.

Leaving aside the externalist option, how could the modal realist know the truth of the proposed analysis of modality? That is, it’s not clear how the modal realist could know that:

(*) \( P \) is possible if, and only if, there is a world in which \( P \).

Either (*) is a case of meaning equivalence or it is not. If it is not a case of meaning equivalence, then it’s not clear that we have a conceptual analysis, given that the right- and the left-hand side terms in (*) have different meanings. If the biconditional in (*) is meaning preserving, then it’s difficult to understand why most people (outside perhaps philosophical circles) would find it puzzling that modal talk should commit them to the existence of possible worlds. In fact, for this reason, it’s implausible to think that (*) is a case of meaning equivalence.
Note also that (*) can only be adequate if on the right-hand side, there are exactly the number of worlds that there could possibly be for the analysis to be right. But this presupposes an independent access to the worlds, which we clearly lack.

Perhaps it could be claimed that we have access to possible worlds through the principle of recombination, just as we have access to mathematical objects through the relevant mathematical principles. But this presupposes the truth of the principle of recombination, which is precisely what the modal realist has to account for.

So, how can the modal realist know that the equivalence between modal concepts and possible worlds in (*) holds at all? Can we say that (*) is known to be true through its consequences? That is, if we know that (*) is true, then we know that we can systematize our modal metaphysics, explain the connection between several apparently unrelated notions (such as properties, verisimilitude, content etc.), and thus accommodate the widespread role played by modality in metaphysics. But any argument of this form provides at best an inductive (and formally invalid) reason to accept (*). Clearly, such an argument fails to provide a very strong reason to believe in the truth of (*), and hence to believe in the existence of possible worlds.

3.2 Problems for the theoretical utility argument

Given these difficulties with the argument based on the analogy with mathematics, perhaps the modal realist could move to the second type of argument to support his or her modal epistemology: the theoretical utility argument. But here we also find difficulties.

(i) Let's first consider the formulation of the theoretical utility argument provided by Divers (see section 2.2, above). The first premise of the argument states that the theoretical utility of a hypothesis that postulates certain objects (e.g. possible worlds) gives us reason to believe in the truth of that hypothesis (leading, in this particular case, to the belief in the existence of possible worlds). But is this premise true?

To answer this question, we need to be clear about what it means to say that a given hypothesis is theoretically useful (or has eminent utility). It means that the hypothesis provides the best overall balance between a number of theoretical virtues, such as simplicity, unification, and explanatory power (see, e.g., Lewis 1986: 3-5, and Colyvan 2001). These are definitely desirable features of a hypothesis, and it's reasonable to prefer hypotheses that exemplify these virtues. However, it has been argued that these theoretical virtues are pragmatic in character only; they are not epistemic. That is, these theoretical virtues tell us about what is significant to us, as users of the hypothesis in question, but they need not tell us something about the connection between the hypothesis and the world. To assume, for instance, that the simplicity of a given hypothesis to us provides a reason for its truth is ultimately to assume some form of anthropocentrism. Hence, if a certain hypothesis exemplifies various pragmatic virtues, this certainly gives us reason to accept that hypothesis; but the mere satisfaction of these virtues doesn't give us reason to believe in the hypothesis' truth (see, e.g., van Fraassen 1980 and 1985). After all, it's possible that the theoretical virtues in question are all satisfied, but the hypothesis under consideration turns out not to be true. This was precisely the case, for instance, with Newtonian mechanics, which had an excellent overall balance between simplicity, unification and explanatory power, but which turned out to be false. Philosophical theories (such as modal realism) need not be different in this respect.

---

6 We develop this point further, in the context of a discussion of modal realism, in Bueno and Shalkowski 2000.
More generally, there seems to be a systematic correlation between the *satisfaction of theoretical virtues* and the *falsity* of the theory that satisfies such virtues. This is, in fact, one of the outcomes of the pessimistic meta-induction in the philosophy of science. All past scientific theories, including those that provided an excellent overall balance between theoretical virtues, turned out to be false (see, e.g., Laudan 1981). This indicates that the connection between the presence of theoretical virtues and the belief in the truth of a theory may not even be reliable.

Note that the same case can also be made for *non-empirical* theories, such as mathematics. In fact, there are numerous examples of mathematical theories that satisfied the relevant theoretical virtues but turned out to be inconsistent. Consider, for instance, the original formulation of the calculus, naïve set theory, or Frege’s logicist reconstruction of arithmetic. In all of these cases, we have theories that were simple, unified, and explanatory. But, assuming classical logic, all of these theories were false, given their inconsistency.

This seems to provide enough evidence to challenge the correlation between satisfaction of theoretical virtues and truth. And so, there’s no reason to accept the theoretical utility argument in the context of modal realism (for further details, see Bueno and Shalkowski 2000). In Divers’ formulation, the theoretical utility argument, although clearly valid, is not sound – the first premise is false.

(ii) Let’s now consider the formulation of the theoretical utility argument as an *indispensability* argument (although the point we make here also applies, *mutatis mutandis*, to Divers’ version). The indispensability argument seems to assume a particular criterion of ontological commitment. Being *indispensable* to our best theories is taken to provide a good reason to believe in the theory’s truth, and hence to be committed to the existence of whatever ontology the theory introduces. But why does the indispensability of certain objects in the formulation of a theory provide a good reason for belief in the truth of that theory? Presumably, because the existence of the postulated entities is required for the theory’s truth. This is the point of the argument, which insists that the indispensability of certain entities is a condition for the commitment to the existence of such entities. But is this argument sound?

One way of resisting the indispensability argument is by denying its *first* premise. According to that premise, we ought to be ontologically committed to all and only those entities that are indispensable to our best theories of the world. But even though certain entities might be indispensable to our best theories, this may not give us reason to believe in the existence of the corresponding entities. The indispensability might be simply a feature of our “linguistic needs”, an aid to overcome the limitation of our expressive power, rather than the manifestation of some intrinsic necessity of the existence of certain objects. We may refer to certain objects not because we are committed to their existence, but just because doing so simplifies immensely what we want to say.

To support this point, let’s first note a distinction between *two kinds of commitment* that seem to be conflated in the indispensability argument (see Azzouni 2004). On the one hand, we incur a *quantifier commitment* by quantifying over certain objects, in the logical sense of the term; that is, by taking certain objects in the domain of our quantifiers. On the other hand, we have an *ontological commitment* when the existence of a given object is asserted. It’s important to note, however, that quantifier commitment does not entail ontological commitment, given that, in our own language, we naturally recognize the fact that we may quantify over objects whose existence we have *no reason to believe*. Suppose that our best theories of fictional discourse, as well as our practices with fiction, require the quantification over fictional characters. Perhaps the best way of making sense of the Sherlock Holmes stories is by quantifying over a detective who solved difficult cases in brilliant ways. Clearly, we incur a *quantifier commitment* here. But we would immediately deny the claim that we are *ontologically committed* to Sherlock Holmes! The natural
response would be to acknowledge that even if we quantify over fictional characters, we have no reason to believe that they exist.⁷

But under what conditions would we be entitled to claim that something exists? We need a criterion of existence to answer this question. There are, of course, several possible criteria. For instance, having causal access to certain objects is one way of determining which objects exist (criterion of causal access). Being able to interact with certain objects is another way (interaction criterion). Or if we are able to observe certain objects, we can say that the observed objects exist (criterion of observability). Independence is another existence criterion: if an object exists independently of us — that is, if we didn’t make it up — we clearly have good grounds to claim that the object exists (criterion of independence; see Azzoumi 2004).

Among these possible criteria, some are epistemological (such as the causal access, the interaction, and the observability criteria); others are ontological (such as the independence criterion). The advantage of the epistemological criteria is that it is easier to determine whether the criteria apply or not. Typically, there’s not much controversy as to whether we have causal access to, or have interacted with, certain objects. The disadvantage of the epistemological criteria is that it’s not so obvious why epistemological constraints provide adequate grounds to decide what there is. Why should one restrict existence to observability, or to the things to which we have causal access? Why can’t there be unobservable entities in the world, or entities that we haven’t causally interacted with?

Ontological criteria, in turn, do not face these difficulties, given that they are not epistemologically constrained. However, it’s not so obvious how we can determine whether an ontological criterion applies or not. Consider, for instance, the criterion of independence. Someone could claim that mathematical objects don’t exist because we made them up, and so they are not independent of us. By invoking this criterion, nominalists about mathematics could deny any ontological commitment to mathematical objects. However, those who believe in the existence of mathematical objects (platonists) could easily resist this move, by maintaining that mathematical objects exist independently of us. In fact, certain platonists insist that even if humans had never existed, there would still be mathematical objects. So, using the same (ontological) criterion, we have opposed answers regarding what exists. And it’s unclear how we could decide which of the two applications of the ontological criterion is correct. As a result, it’s not obvious how we could settle disputes about what there is with ontological criteria of existence.

Of course, a significant requirement in debates over criteria of existence is to try to avoid begging questions against the proposals under scrutiny. It’s not clear, however, that this can always be done (see Azzoumi 2004). After all, in adopting a particular criterion of existence, we immediately rule out certain entities that would be taken to exist had a different criterion been accepted. And those who believe in the existence of the offending entities would immediately claim that the question had been begged against them. This explains why ontological disputes are notoriously hard to settle.

Luckily for our purposes here, we need not resolve the issue about which criterion of existence should be adopted in the debate over the existence of possible worlds. Our point here is simply to highlight that Lewis has implicitly invoked one such criterion, namely, the indispensability of

⁷ Quine, of course, would agree, given that, on his view, fictional characters can always be dispensed with. But, unfortunately, such paraphrases are not always available (see, e.g., Melia 1995). And even in the absence of such paraphrases, that is, even when we acknowledge the indispensability of quantifying over certain objects (such as fictional characters), we correctly resist ontological commitment to them.
quantifying over certain objects. But, as the previous discussion makes clear, nothing requires the adoption of this particular criterion. Moreover, if a different criterion had been adopted, we would have obtained a different answer regarding the existence of possible worlds than Lewis'. For instance, had we adopted the causal accessibility criterion, we wouldn't take possible worlds in Lewis' sense to exist – given that we have no causal access to such objects. Similarly, had we adopted the observability criterion, we would have no reason to assume that possible worlds exist either. After all, possible worlds are not observable – certainly not from the actual world! Of course, we are not advocating here any particular criterion of existence; we are only pointing out how the choice of a criterion bears on the issue of the existence of possible worlds, and indicating the significant contingency of adopting the criterion that Lewis ultimately adopted.

Moreover, by distinguishing quantifier commitment and ontological commitment, it's possible to deny the first premise of the indispensability argument. Even if certain entities are indispensable to our best theories of the world and we inevitably quantify over such entities, this is not enough to give us reason to be ontologically committed to such entities. Not only fictional characters clearly illustrate this situation, but also mathematical objects, in the context of the mathematics applied to our best physical theories. Typically, physicists insist that the mathematics is only a useful device, and despite quantifying over mathematical objects, they resist the claim that they are ontologically committed to such objects (see, e.g., Dirac 1958). By carefully distinguishing quantifier commitment and ontological commitment, it's possible to make perfect sense of this situation, and undermine the indispensability argument along the way.

The same point applies to the discussion of possible worlds. Even if we grant that quantifying over possible worlds is indispensable to our best theories of modality, this is not enough to commit us ontologically to the existence of such objects. We have, at best, a quantifier commitment to possible worlds, but this is not sufficient for ontological commitment. An additional criterion of existence needs to be met before we are entitled to the latter commitment. With the theoretical utility argument, Lewis simply assumed one criterion of existence: the indispensability (or the theoretical utility) of postulating possible worlds. But as we pointed out, nothing requires the adoption of this particular criterion. And with different criteria in place, we get different answers regarding the existence of such worlds. So, we don't think there are good reasons to accept the first premise of the indispensability argument. And, in this way, the modal realist still owes us an epistemological story to support his or her view.

4. SKETCHING AN ALTERNATIVE VIEW

Is it possible to develop an epistemology for modal discourse that is not subject to the difficulties faced by the modal realist? In this section, we will sketch such an alternative. Our goal here is simply to indicate, in outline, the key moves. We will develop the details in future works.

The strategy developed here has two components: First, there's a logical component: modal knowledge involves knowledge of what follows from what; in particular, what follows from a

---

8 The same point also applies, of course, for the theoretical utility of postulating possible worlds. In fact, in Diver's formulation of the theoretical utility argument, the utility of postulating possible worlds was ultimately taken as a criterion of existence. In this way, the present discussion also provides additional worries for Divers' (and Lewis') version of the argument.

9 As we point out below, we don't think there are good reasons to accept the second premise of the indispensability argument either. Possible worlds are not indispensable to our best theories of modality (see the next section).
given body of modal beliefs. Second, there’s a modal component: modal knowledge involves an irreducible modal element in that to know what is possible (or necessary), we need to determine what is consistent (or inconsistent) with certain claims. So, as opposed to Lewis, we are not trying to provide an analysis of modal discourse in non-modal terms. The difficulties discussed above seem to be enough to motivate the need for a different approach.

The previous discussion also indicates the troubles of providing a modal epistemology if modal knowledge is thought of as knowledge of the existence of possible worlds. Perhaps the best strategy to accommodate modal knowledge is to resist the postulation of possible worlds — in terms of which modal claims are analyzed — and instead embrace the idea that we already have some primitive notion of modality, such as a primitive notion of consistency. This notion is then used to guide us in the process of assessing the overall consistency of our modal beliefs, and in helping us to determine which modal beliefs (now thought of as beliefs about what is possible and what is necessary) follow from other modal beliefs. We are thoroughly immersed in a modal language: our own language clearly has modal idioms, and it’s probably a hopeless task to try to analyze them in non-modal terms. By assessing the connections between the various modal beliefs we have, we are in position to articulate what we know and what we fail to know regarding modality.

Lewis is certainly right in recognizing the role that imaginative thought experiments play in forming our modal beliefs (see Lewis 1986: 115). Certainly, they provide significant ways in which we come to have the modal beliefs we have. But we are often able to imagine all sorts of things, including impossible things. And the difficulty here is to be able to determine which thought experiments are acceptable and which aren’t. Clearly, simply imagining something is not sufficient to establish the possibility of what we have imagined. For we may think we have imagined X, but, in fact, we have imagined something slightly different. We may have imagined that we have squared the circle, by thinking in general terms of some operations that would lead to this outcome. But, in fact, we haven’t done such a thing. There are proofs to the effect that this outcome can’t be done: a contradiction obtains if we suppose that this outcome is the case (assuming classical logic, this establishes the impossibility of the outcome). Hence, we need some additional constraints to guide our modal judgments.

If we take modal knowledge to be knowledge of the existence of possible worlds, it’s much harder to come up with such constraints and articulate a modal epistemology. But once we realize that possible worlds play no role whatsoever in forming our modal beliefs, and that we are typically guided by the primitive notions of modality that we already have in place (from our own language), modal knowledge becomes much less mysterious. We start from this primitive notion of possibility (consistency), and we determine what else is possible by adding new modal beliefs that are consistent with the modal beliefs we already have. And we determine what is necessary by evaluating what follows necessarily from our modal beliefs. In doing this, we try to form maximally consistent sets of modal beliefs. We may stumble into inconsistency along the way, and we revise our modal beliefs accordingly. As a result of this process, we end up also refining our initial, primitive notion of consistency, and evaluating its adequacy and limitations.

In this sense, modal epistemology and mathematical epistemology are not too far apart. Just as modal epistemology need not be an epistemology of possible worlds, mathematical epistemology need not be an epistemology of mathematical objects. To develop a mathematical

10 For simplicity, we are assuming here that the underlying logic is classical. Of course, this need not be the case. For example, if the underlying logic is paraconsistent, then there will be inconsistent but non-trivial belief sets; that is, sets of beliefs that are inconsistent, but such that not every belief follows from them (see da Costa 1997). In this case, the goal of a modal epistemology would be to try to formulate maximally non-trivial sets of modal beliefs.
epistemology, the existence of mathematical objects plays no role; in particular, it plays no role in the formation of our mathematical beliefs. It's certainly not the case that mathematical objects somehow "constrain", or bring some "resistance to", our mathematical beliefs. We do have a primitive notion of consistency that we invoke in forming our mathematical beliefs: beliefs about what is logically possible and what follows logically from certain mathematical principles. But given that the existence of mathematical objects is not presupposed, the truth of the mathematical principles is not relevant either (see Field 1989).

The same point also holds in the modal case: given that the existence of possible worlds is not presupposed, the truth of our modal principles (as claims about possible worlds) is not required. And by not requiring the truth of such modal principles, the strategy suggested here doesn't face the problems we raised for Lewis. The issue regarding the reliability of the proposed strategy doesn't arise — certainly not in the sense that the principles invoked need to be true. But if modal principles need not be true, what norm should they satisfy? It's enough if the modal principles in question are conservative over the modal facts; that is, conservative over facts about what is possible and necessary (but without invoking possible worlds). This clearly provides a reason for the ultimate dispensability of possible worlds. And in this way, we challenge here the second premise of the indispensability argument (see section 2.2, above). Even if referring to possible worlds might be useful to express some modal claims, this use is not indispensable: possible worlds talk is a useful, but perfectly dispensable, fiction. Here is why (see also Bueno 1998).

The key idea is that, instead of being true, possible worlds talk only has to be conservative with respect to modal claims that do not involve possible worlds. (We call this $m$-conservativeness.) More clearly, and extending a proposal that Hartry Field originally formulated in the philosophy of mathematics (see 1989: 58) to the theory of modality, a possible worlds theory PW is $m$-conservative if it is consistent with every internally consistent set of modal claims. From this it follows that PW is $m$-conservative if and only if for any modal assertion $A$ and any body $M$ of such assertions, $A$ does not follow from $M + PW$ unless it follows from $M$ alone. (A proof of this result will be given below.) Thus, if the possible worlds theory is $m$-conservative, we would be entitled to use possible worlds talk only to facilitate inferences involving modal claims.

Here is the proof. First, let us introduce two definitions:

1. According to a possible worlds theory PW, the following valuations $v$ for a modal language hold:

   \[ v(\Diamond \alpha) = 1 \iff \exists w v(\alpha) = 1 \]
   \[ v(\Box \alpha) = 1 \iff \forall w v(\alpha) = 1 \]

   where the right-hand quantifiers range over non-actual worlds. These truth-conditions have been stated in the metalanguage of the modal language under consideration. But they have obvious object-language counterparts, namely:

   \[ \Diamond \alpha \leftrightarrow \exists w \alpha \]
   \[ \Box \alpha \leftrightarrow \forall w \alpha. \]

   A possible worlds theory PW is characterized here by these object-language axiom schemes.

2. A possible worlds theory PW is $m$-conservative iff for every consistent set $M$ of modal claims, there is a model of $M + PW$; in other words, for every set $M$ of modal claims, if $M$ is consistent, so is $M + PW$. (For simplicity, we are assuming here the model-theoretic notion of
consistency. Moreover, we will not consider the extensional connectives and quantifiers of the language, assuming that their treatment is given in the usual way.)

What we have to establish now is the following theorem. A possible worlds theory PW is \(m\)-conservative iff for every set \(M\) of modal claims and every modal claim \(\alpha\), if \(\alpha\) is a logical consequence of \(M+PW\), then \(\alpha\) is a logical consequence of \(M\) alone.

The left to right implication goes as follows. Suppose that PW is \(m\)-conservative and that \(\alpha\) is not a consequence of \(M\). We show that \(\alpha\) is not a consequence of \(M+PW\). Since \(\alpha\) is not a consequence of \(M\), there is a model of \(M\) (let us call it ‘Mod’*) according to which \(v_{\text{Mod}}(\alpha) = 0\). However, since PW is \(m\)-conservative, given the model Mod of \(M\), there is a model of \(M+PW\). Let us call such a model ‘Mod**’. We show that \(v_{\text{Mod}^*}(\alpha) = 0\). Indeed, since \(\alpha\) is a modal claim, either it is of the form \(\Box \beta\) or of the form \(\neg \Box \beta\). In the former case, we have that \(v_{\text{Mod}^*}(\alpha) = v_{\text{Mod}^*}(\Box \beta) = v_{\text{Mod}^*}(\exists \psi \beta)\). But \(v_{\text{Mod}^*}(\alpha) = v_{\text{Mod}^*}(\alpha)\), since PW is a conservative extension of \(M\), and \(v_{\text{Mod}^*}(\alpha) = 0\). Therefore, \(v_{\text{Mod}^*}(\alpha) = 0\). The latter case (in which \(\alpha\) is of the form \(\neg \Box \beta\)) is treated similarly. Thus, \(\alpha\) is not a consequence of \(M+PW\).

The right to left implication goes as follows. Suppose that PW is not \(m\)-conservative. We show that for some \(M\) and some \(\alpha\), \(\alpha\) is a consequence of \(M+PW\), but \(\alpha\) is not a consequence of \(M\). Since PW is not \(m\)-conservative, there is a consistent set \(M\) of modal claims such that there is no model of \(M+PW\). Since \(M\) is consistent, there is a model of \(M\). Let us call it ‘Mod’. Let \(\alpha\) be a modal sentence such that that \(v_{\text{Mod}}(\alpha) = 0\). Thus, \(\alpha\) is not a consequence of \(M\). But \(\alpha\) is a consequence of \(M+PW\), since there is no model of \(M+PW\) in which \(v(\alpha) = 0\). This concludes the proof.

This argument clearly establishes the dispensability of possible worlds in the analysis of modal discourse, and thus it challenges the adequacy of the second premise of the indispensability argument when applied to the metaphysics of modality. Given the dispensability of possible worlds, there’s no need to assert the truth of possible worlds talk to give an account of modal discourse. Moreover, given that possible worlds discourse is never asserted to be true, there’s no conflation of pragmatic and epistemic reasons in our proposal. After all, even if possible worlds were indispensable, this would establish, at best, the usefulness of the possible worlds hypothesis. However, as we argued above, this is not sufficient to establish the truth of that hypothesis, and hence the existence of possible worlds. As a result, given that possible worlds play no role in our account, this simplifies immensely the epistemology of modality. Modal knowledge is ultimately a matter of what follows from certain modal claims, and which claims are consistent with each other. So, it’s ultimately a matter of establishing logical relations among our modal beliefs.

5. CONCLUSION

As indicated above, we are not convinced that Lewis’ two strategies to develop a modal epistemology work. Neither the analogy with mathematics nor the theoretical utility argument succeed in reliably establishing how we can have knowledge of the existence of possible worlds. Instead of exploring the analogy with mathematics, or arguing for the existence of possible worlds via the theoretical utility argument, we explore the modalist intuition that we already have a primitive notion of modality in our language. We then use this notion as guide to our knowledge of modal claims, without presupposing the existence of possible worlds. Even though there’s more to be said, we hope we have said enough to indicate why the resulting epistemology is not as problematic as the one that the modal realist owes us.
REFERENCES