

CONTENTS

<i>Contributors</i>	ix
Introduction: Conceivability and Possibility TAMAR SZABÓ GENDLER AND JOHN HAWTHORNE	I
1. Modal Epistemology and the Rationalist Renaissance GEORGE BEALER	71
2. Berkeley's Puzzle JOHN CAMPBELL	127
3. Does Conceivability Entail Possibility? DAVID J. CHALMERS	145
4. Desire in Imagination GREGORY CURRIE	201
5. Essentialism versus Essentialism MICHAEL DELLA ROCCA	223
6. The Varieties of Necessity KIT FINE	253
7. A Study in Modal Deviance GIDEON ROSEN	283
8. On the Metaphysical Contingency of Laws of Nature ALAN SIDELLE	309
9. The Art of the Impossible ROY SORENSEN	337

10. Reliability and the A Priori	369
ERNEST SOSA	
11. What is it Like to be a Zombie?	385
ROBERT STALNAKER	
12. The Conceivability of Naturalism	401
CRISPIN WRIGHT	
13. Coulda, Woulda, Shoulda	441
STEPHEN YABLO	
<i>Index</i>	493

Introduction: Conceivability and Possibility

TAMAR SZABÓ GENDLER AND
JOHN HAWTHORNE

Overview

We have, it seems, a capacity that enables us to represent scenarios to ourselves using words or concepts or sensory images, scenarios that purport to involve actual or non-actual things in actual or non-actual configurations. There is a natural way of using the term ‘conceive’ that refers to this activity in its broadest sense.¹

When we engage in such conceivings, the things we depict to ourselves frequently present themselves *as possible*, and we have an associated tendency to judge that they *are possible*. Indeed, when invited to consider whether something is possible, we often engage in a deliberate effort to conceive of it; upon

For comments and conversation concerning this introduction, we are grateful to George Bealer, David Chalmers, Brendan Murday, Ted Sider, Zoltán Gendler Szabó, and Steve Yablo.

¹ The term ‘conceive’ shares a root with the term ‘concept’—the former is traceable to the Latin verb *concipere*, the latter to its past participle *conceptus*. But while the verb *concipere* is used frequently throughout antiquity, employment of the nominal form term *conceptus* does not seem to emerge until the third or fourth century CE; instead, the term *notio* (roughly ‘notion’) was employed. (Thanks to Charles Brittain for research into this question.) In this light, it seems reasonable to follow modern usage in allowing a broad sense for the term ‘conceive’ that permits as instances certain uses of (see *Roget’s Thesaurus*, 3rd edn.) ‘envisage, envision, fancy, fantasize, image, imagine, picture, see, think, vision, visualize’—that is, a use of the term that is non-committal on the relation between conceiving and concept-deployment. Imagining and conceiving in the narrow sense are special cases of conceiving in this broad sense.

finding ourselves able to do so, we conclude that it is. We may even decide that something is impossible on the basis of our apparent inability to conceive of it.

The technique just described is a pervasive feature of our mental life—both in day-to-day decision making and in philosophical reasoning. We might conceive of a scenario in which the couch fits through the door or the Democrats take control of the Senate, and conclude that such events could occur; if we cannot conceive of any such scenario involving the piano or the House of Representatives, we may conclude that such events could not. We might conceive of a scenario in which there is a golden mountain or a red square, and conclude that such entities could exist; if we cannot conceive of any scenario involving a mountain without a valley or a round square, we may conclude that such entities could not. We might conceive of a scenario in which there are exact physical duplicates of actual human beings who lack consciousness, and conclude that such beings are possible; if we cannot conceive of any such scenario, we may conclude that such beings are not.

Although there are numerous differences among the cases just described—in the sorts of conceiving in which they ask us to engage, the extent to which they involve sensory imagination, the sorts of possibility they apparently invoke, the generality of the conclusions they purport to establish,—there are also important similarities. In particular, each employs what is sometimes called a *conceivability–possibility* (or *inconceivability–impossibility*) move: from the fact that we are (or are not) able to depict to ourselves a scenario in which thus-and-such obtains, we take ourselves to have learned something about whether thus-and-such *could* (or could not) obtain. The very existence of such a practice raises a number of perplexing philosophical questions. Of these, four stand out as particularly central to traditional and contemporary discussions of the topic:²

- (a) What sorts of possibility are there, such that the conceivability of a scenario might be thought to be an indicator of that scenario's being possible?
- (b) What is it to conceive of something?
- (c) When is conceivability a reliable guide to possibility?
- (d) How, if at all, might conceivability–possibility reasoning be employed in particular cases to establish claims about the actual world?

In the remainder of this introduction, we discuss these questions from various perspectives. In section 1, we provide a broad overview of some of the general philosophical issues raised by consideration of them. Sections 2–4 are devoted to surveying in somewhat more detail some of the highlights of

² Note that in asking these questions, we are using 'conceive' in the broad sense described in the opening paragraph. There is an important tradition of trying to clarify the relevant notion of conceiving in terms of the notion of a priori (or rational) intuition; see Bealer, Ch. 1 below; see also Yablo (1993) and various of the papers in DePaul and Ramsey (1998).

traditional and recent discussions of these and related issues: section 2 describes accounts due to Descartes and Hume; section 3 describes an account due to Saul Kripke; and section 4 presents an overview of a family of accounts commonly referred to as *two-dimensionalist*. Finally, in section 5, we present brief summaries of each of the chapters in this volume.

I General Issues

I.1 Possibility

Our faculty of perception reveals to us what is actual. And there is a widely accepted explanation of why this is so: our perceptual mechanisms are sensitive to features of the actual world, which impinge on them causally to produce systematic patterns of stimulus and response. Likewise, it seems, our faculty of conception reveals to us what is possible. But here there is no widely accepted explanation of why and to what extent this is so.

There are two reasons for this. The first is that the term ‘possible’ is used in a number of different ways, resulting in a certain ambiguity in what is being claimed. But even when this is sorted out, a deeper problem in the epistemology of modality remains. We address these issues in turn.

If I claim that thus-and-such is possible, there are a number of things I might mean. One thing I might mean is that, for all I know, thus-and-such obtains. In so doing, I invoke one of the family of notions of *epistemic possibility*: notions of possibility that are defined relative to some subject (or set of subjects) in terms of some body of knowledge or evidence available to (or otherwise associated with) the subject(s) in question.³ So, for example, one might offer a *permissive* account of epistemic possibility, according to which P is epistemically possible for S just in case S does not know that not-P, or a *strict* account, according to which P is epistemically possible for S just in case P is consistent (metaphysically compossible) with all that S knows.⁴ There are important differences

³ For a systematic discussion of various established epistemic uses of ‘possible’ in English (and for a challenge to certain further alleged such uses), see Bealer, Ch. 1, sect. 1.3, below. See also Yablo (1993: 22–5).

⁴ In addition to these permissive and strict notions, intermediate characterizations might also be offered: for instance, P is epistemically possible for S just in case S’s evidence does not warrant S’s believing not-P; or P is epistemically possible for S just in case S could not reasonably be expected to ascertain not-P on the basis of what S knows. Spelling out precisely what these amount to requires fully characterizing what is meant by ‘warrant’ and ‘reasonable expectation’. Occasionally, the term is used in an even broader—and impersonal—sense: P is epistemically possible just in case it is not a priori that not-P (see, e.g., Chalmers, Ch. 3 below). Spelling out precisely what this amounts to requires fully characterizing what is meant by ‘a priori’.

between these characterizations: on the strict account, epistemic possibility entails metaphysical possibility; on the permissive account, it does not. But, regardless of the details of the characterization, conceivability is clearly *not* a general guide to epistemic possibility, in either of these senses. If I know that the cat is on the mat, then it is not epistemically possible for me that the cat is not on the mat, even in the permissive sense.⁵ Still, I can easily conceive a situation in which the cat is not on the mat—so I can easily conceive something epistemically impossible. Those who claim conceivability as a guide to possibility, then, presumably do not mean that it is a guide to *epistemic* possibility.⁶

So what they mean is that it is a guide to *non*-epistemic possibility. But in what sense? Consider three candidate notions of non-epistemic possibility with which philosophers have traditionally been concerned: (narrow) logical possibility, metaphysical possibility, and nomological (for example, physical or biological) possibility.⁷ On a standard sort of characterization, P is *logically possible* just in case no contradiction can be proved from P using the standard rules of deductive inference (in conjunction, perhaps, with certain definitions). On a similarly standard characterization, P is *nomologically possible* for a relevant body of *nomos* just in case P is consistent with the body of truths expressed by those laws. (For example, P is physically possible iff P is compossible with the laws of physics, biologically possible iff it is compossible with the laws of biology, and so on.) The notion of *metaphysical possibility*, meanwhile, is standardly taken to be primitive.⁸ It is taken as the most basic conception of ‘how things

⁵ Note that it *is* epistemically possible in the broad impersonal sense employed by Chalmers (since presumably my knowledge that the cat is on the mat is not a priori); for an argument that conceivability (of the requisite sort) is indeed a guide to epistemic possibility (in this sense), see Chalmers, Ch. 3 below.

⁶ One might object that the permissive notion of epistemic possibility is not a notion of *possibility* at all, since (as we noted in the text) P can be epistemically possible in this sense without being metaphysically possible.

⁷ We are not taking a stand on whether these notions are, on final analysis, distinct—only observing that they have been, at various times by various philosophers, treated as such. (A complicating terminological factor is the following: as George Bealer (Ch. 1) reminds us, before Kripke introduced the expression ‘metaphysical possibility’, many philosophers (including Kripke) used ‘logical possibility’ for what is now called ‘metaphysical possibility’. Where this consideration may lead to confusion, we flag divergent uses.) For discussion of the question of how many primitive forms of possibility there are, see Fine, Ch. 6 below; for additional discussion of the relation between nomological and metaphysical possibility, see Shoemaker (1998), Sidelle (Ch. 8 below and references therein); for additional general discussion of varieties of possibility, see also the chapters in this volume by Bealer, Chalmers, Della Rocca, Wright, and Yablo. For discussion of what is sometimes called *conceptual possibility*, see sections 3 and 4 below.

⁸ In contemporary discussions, at any rate.

might have been’—gestured at by talk of how ‘God might have made things’ or ‘ways it is possible for things to be’.⁹ Using the terminology of possible worlds, actuality and possibility in this sense can be characterized in parallel fashion: it is *actual* that P just in case P in the actual world, and (metaphysically) *possible* that P just in case P in some possible world.¹⁰ On the characterizations we have offered, then, it would appear that metaphysical possibility is more expansive than nomological possibility, less expansive than narrow logical possibility:¹¹ it is possible in none of the senses that something is both red and not red, logically but not metaphysically possible that something is both red and non-extended, metaphysically but not physically possible that something travel faster than the speed of light, and possible in all three senses that something travel faster than the space shuttle.

As a guide to logical possibility in the sense we have characterized it, conceivability seems somewhat superfluous; whether or not a contradiction can be derived from P seems better determined by proof procedures than by scenario depiction. If it is this notion of possibility that is at issue, the activity of conceiving seems largely irrelevant (or at least inessential) to the determination of possibility.¹² As a guide to nomological possibility, conceivability seems to confront many of the problems confronted in the epistemic case—namely, that it seems all too easy for us to conceive of situations that are not possible in the relevant sense. So if it is this notion of possibility that is at issue, the activity of conceiving seems largely ineffective to the task at hand.¹³ Rather, it is as a guide to metaphysical possibility that conceivability is typically taken as having a central role to play. On the standard view, our ability to conceive of a scenario

⁹ The distinction between essential and inessential properties is standardly glossed in terms of metaphysical possibility: a property is essential to a thing just in case it is not metaphysically possible that the thing lacks it.

¹⁰ Correspondingly, it is *necessary* that P just in case P in all possible worlds, *impossible* that P just in case P in no possible world, and *contingent* whether P just in case it is possible that P and not necessary that P.

¹¹ While the majority of the authors in this volume follow roughly this terminological practice, a few employ certain of the terms somewhat differently (e.g., Chalmers’s uses of (‘epistemic’)); in all cases where authors’ use differs significantly from that which we have introduced here, explicit characterizations of the terms in question are provided in the author’s main text. Note also that while several of the authors are directly interested in contrasting the various sorts of possibility introduced thus far (e.g., Fine, Sidelle), a number of others are primarily interested in the contrast (to be discussed in sections 3 and 4 below) between metaphysical and conceptual possibility (e.g., Chalmers, Della Rocca, Stalnaker, Wright, and Yablo), while still others are interested in additional uses of possibility: for instance, those relating to normative notions (e.g., Currie, Fine, and Yablo).

¹² For arguments that it is irrelevant in general, see Bealer, Ch. 1 below.

¹³ See, however, next paragraph and n. 15.

where P obtains is reckoned as constituting at least prima-facie reason for supposing that P is metaphysically possible. This issue is addressed, directly or indirectly, by nearly all the authors in the volume.

If conceivability is a good guide to metaphysical possibility, it is easy enough to see how—given the requisite additional nomological (or epistemic) information—it could in a derivative way be a good guide to various species of nomological and epistemic possibility. If what it means for P to be nomologically possible relative to some body of laws L (or epistemically possible relative to some body of knowledge K) is for P and L (or P and K) to be metaphysically compossible,¹⁴ then the question of whether P is L-nomologically (or K-epistemically) possible can be tested by attempting to conceive of a scenario where L and P hold (or where K and P hold).¹⁵ Of course, if we are misinformed about L (or K), we will go astray. But since conceivability can only work in this way as a guide to epistemic and nomological possibility on the assumption that it is a good guide to metaphysical possibility, it is the latter assumption that has been the primary object of philosophical attention.

While these clarifications dispel a certain amount of confusion, they do little to resolve an obvious puzzle: on the face of it, the idea that conceivability is a guide to metaphysical possibility is extremely problematic. According to current orthodoxy, metaphysical possibility can neither be reduced to, nor eliminated in favour of, linguistic rules and conventions; it constitutes a fundamental, mind-independent subject-matter for thought and talk. Given this picture, it is rather baffling what sort of explanation there could be for conceiving's ability to reveal its character. It seems clear that the causal explanation for the reliability of perception is unsuitable here—and it is profoundly difficult to see what to put in its place. A number of authors in the volume take up just this issue.¹⁶

¹⁴ Note while most of the notions of possibility described above can be characterized in this way, not all can; in particular, neither (narrow) logical possibility nor the permissive notion of epistemic possibility can be spelled out in such relative terms.

¹⁵ Matters may be more complex than this simple sketch allows. Perhaps, for instance, we have an evolutionarily instilled capacity for physical intuition that is psychologically distinct from what philosophers call 'conceiving' on the basis of which we are able to draw reliable conclusions about what is physically possible. At the same time, we sometimes set out to learn about the actual world by deliberately conceiving of idealized situations where certain laws of nature—say, laws governing friction—fail to hold. (These topics fall largely outside the scope of this volume. For an overview of such issues, see papers collected in Horowitz and Massey (1991), DePaul and Ramsey (1998), and additional works cited in the 'Bibliography on Experiment and Thought Experiment' in Gendler (2000: 229–50).)

¹⁶ See in particular the chapters by Bealer, Chalmers, Rosen, and Sosa.

1.2 *Conceiving and Imagining*

While the notion of possibility has received fairly thorough philosophical consideration, much less attention has been devoted to the second element in the supposed equation: conceiving.¹⁷ In the opening paragraph, we characterized the activity extremely broadly—broadly enough to include any sort of mental depiction of a scenario, whether in words or concepts or pictures.¹⁸ But, pressing a bit harder, one might wonder what sort of underlying psychological kind (or cluster of kinds) the notion is meant to capture. Consider the following rather diverse list of mental activities, each arguably related to the requisite notion of conceiving:

- Rationally intuiting that it is possible that P
- Realizing that not-P is not necessary
- Imagining (that) P
- Conjecturing that P
- Accepting that P for the sake of argument
- Describing to oneself a scenario where P obtains
- Telling oneself a coherent story in which P obtains
- Pretending that P
- Make-believing that P
- Supposing (that) P

¹⁷ Indeed, it is far from clear how its nature should even be investigated. One might, it seems, fruitfully explore it in light of recent work in (i) empirical psychology, (ii) general philosophy of mind, (iii) phenomenology, or (iv) general work on the nature of pretence, make-belief, and fiction—though it remains an open question how, if at all, such results could illuminate the relation between this capacity and our knowledge of modality. Regarding (i), one might consider the empirical psychological literature on (a) mental imagery (for an overview of these issues, see papers collected in Block (1981), Shepard and Cooper (1982); for developmental perspectives, see Piaget and Inhelder (1971)); (b) imagination and pretence (for a general overview of the psychological literature on imagination, particularly in children, see Harris (2000); for developmental perspectives, see papers collected in Lewis and Mitchell (1994), esp. part III, as well as numerous recent papers in the *British Journal of Developmental Psychology*, *Child Development*, *Cognition*, and *Cognitive Development*); (c) conceptions of possibility and necessity in developmental perspective (for classic discussion, see Piaget (1987a, 1987b); for more recent essays, see papers collected in Overton (1990)). Regarding (ii), one might consider in particular discussions of mental simulation and other sorts of non-belief-like attitudes (for a general philosophical overview of the issues relating to mental simulation, see papers collected in Davies and Stone (1995a, 1995b), as well as numerous recent papers in *Mind and Language*). Regarding (iii), one might begin with Casey (1976/2000), Sartre (1939/1962 and 1940/1963), or, for more general discussions of imagination, Brann (1991) and Warnock (1976). Regarding (iv), one might consider Walton (1990) and Currie and Ravenscroft (2002).

¹⁸ Cf. n. 1.

Understanding the proposition that P
Entertaining (that) P
Mentally simulating P's obtaining
Engaging in off-line processing concerning P

If conceiving is a natural psychological kind, then it presumably corresponds to something like one of these items (or to some natural cluster of them). But the wide variability among their features suggests that the notion in question may be highly elusive.¹⁹ Some, for example, are propositional attitudes; some are attitudes towards scenarios or states of affairs; and still others are activities. Some seem explicitly sensory; others explicitly non-sensory; still others are neutral on this question. Some are highly conceptual; others are strongly language-based; still others are, perhaps, non-conceptual. Some seem to take place primarily spontaneously, others only under our deliberate control; others in both ways. All seem capable of being directed both towards propositions (or states of affairs) involving particular individuals, as well as towards propositions (or states of affairs) that are general. And both within and among them there seem to be variations in the degree of privileged access associated with the attitude/activity and its content/object.²⁰ In light of these differences, one might reasonably wonder which, if any, of the features alluded to is required by conceivability in the sense we seek.

¹⁹ Cf. P. F. Strawson who writes, concerning the notion of imagination: 'the uses, and applications, of the terms "image", "imagine", "imagination" and so forth, make up a very diverse and scattered family. Even this image of a family seems too definite. It would be a matter of more than difficulty to identify and list the family's members, let alone their relations of parenthood and cousinhood' (Strawson 1970: 31). Or again, Brian O'Shaughnessy: 'What is the *imagination*? What is it to *imagine*? These perfectly natural questions already assume too much: the first question assumes there exists something that is the Imagination, presumably a distinctive faculty; the second that there is some one thing that is the phenomenon of Imagining, doubtless instantiated in diverse phenomenal forms. These assumptions may be valid, but they need not be. We ought not to prejudge these questions' (2000: 339–40). (After detailed discussion, O'Shaughnessy provisionally concludes that 'the doctrine of an essential common imagination agency is unacceptable, along with the theory of a common intrinsic imagining essence' (2000: 361), though he leaves room for looser uses of the term.) To the extent that we are taking 'conceiving' to be even broader than 'imagining', such problems are only magnified.

²⁰ Cf. Wittgenstein: 'Someone says, he imagines King's College is on fire. We ask him: How do you know that it's *King's College* you imagine on fire? Couldn't it be a different building very much like it. In fact, is your imagination so absolutely exact that there might not be a dozen buildings whose representation your image could be?—And still you say: "There's no doubt I imagine King's College and no other Building"' (Wittgenstein 1958/1965, 39). The question 'how do you know that it is really X that you imagine' may be absurd in the case of imagining King's College; but its analogue for certain other attitude-content pairs seems perfectly reasonable.

There is a traditional distinction made between (sensory) imagining on the one hand, and (non-sensory) non-imagistic conceiving on the other.²¹ But it is far from settled whether the distinction has a proper role to play in circumscribing the appropriate subject-matter for an investigation of conceivability as a guide to possibility.²² Certainly, it seems, there are things that we can (non-sensorily) conceive that we cannot (sensorily) imagine: but is the modal status of one or the other of these categories more or less apt to be illuminated by the associated mental act?²³ Perhaps there is an alternative distinction to be made between imagining and mere conceiving—wherein imagining is somehow perspectival or self-involving, whereas mere conceiving is not. If so, there may again be a difference between their relative capacities to illuminate certain sorts of modal and non-modal subject-matter.²⁴ More generally, one might wonder about the relation between imagination/conception, on the one hand, and perception/intellection, on the other: is the former parasitic on the latter, merely *re*-presenting perhaps in slightly modified form, what the latter presents, or is it somehow productive, enabling us to gain access to genuinely novel experiences or ideas?²⁵

While the general notion of conceiving and its connections to the notion of imagining have not been well investigated, a few attempts have been made to systematize what is meant by these terms,²⁶ and a number of the authors in this volume offer significant refinements on the traditional ‘placeholder’ use.²⁷

1.3 *Reliability Conditions*

Even if we lack a fully satisfying explanation for the link, conceivability does seem to provide at least a *prima-facie* guide to possibility; that something is conceivable is at least a good indicator that it is possible. At the same time, it is uncontroversial that there are cases where we are misled. Some Greeks found it conceivable (in some decent sense of ‘conceivable’) that stars are holes in the

²¹ This distinction is explored in more detail in our discussion of Descartes and Hume in section 2 below.

²² For an argument suggesting that failure to make such a distinction may lead to mistaken philosophical conclusions, see Hill (1997).

²³ For an exploration of the question of whether impossibilities that can be represented in non-sensory modalities can be depicted visually, see Sorensen, Ch. 9 below.

²⁴ For an exploration of this question with regard to morality, see Currie, Ch. 4 below.

²⁵ For a discussion of the contrastive roles of experience and mere representation, see Campbell, Ch. 2 below.

²⁶ For an influential attempt to provide such a taxonomy, see Yablo (1993). Some attention is also given to this question in van Cleve (1983) and Tidman (1994).

²⁷ In this regard, see especially the chapters by Bealer, Chalmers, and Yablo.

sky. George Berkeley found it conceivable that chairs are mere agglomerations of sensory experiences. Various mathematicians have found it conceivable that Goldbach's conjecture is wrong. Some philosophers have found it conceivable that seven hairs marks the borderline between being bald and not being bald, others that nominalism is true, that immanent universals exist,²⁸ or that there are exact physical duplicates of human beings who lack consciousness. Some find it conceivable that Hesperus is not Phosphorus, or that water is not H₂O. In each of these cases, the content that is conceived may well be metaphysically impossible. But if so, conceivability has failed as a guide to the relevant sort of possibility.

Can these cases be cordoned off in a principled way, so that one can explain the failure of conceivability in particular cases while maintaining the general reliability of the practice described? Or can nothing systematic be said in this regard?

In responding to this challenge, three families of strategies suggest themselves. The first cordons off on the basis of subject-matter. Perhaps certain classes of propositions—abstract metaphysical ones, ones concerning necessary beings, ones that turn on actual empirical matters of fact, and so on—are illegitimate targets for conceivability–possibility arguments. Perhaps our conceiving faculty (whatever that turns out to be) is simply ill-suited to the task of providing reliable guidance concerning such realms. At the same time, it might be suggested, other topics are such that conceiving can deliver reliable modal verdicts concerning them. As long as we circumscribe subject-matter properly, conceivability will be a reliable guide to possibility.

A second strategy cordons off by way of procedure. Perhaps certain sorts of conceivings—clear and distinct conceivings, or conceivings accompanied by rational insight, or conceivings that involve a detailed intellectual vision of a possible scenario—are able to deliver reliable modal verdicts. And perhaps there is a straightforwardly detectable difference between these sorts of conceivings and their unreliable counterparts. As long as we restrict ourselves to the relevant sorts of conceiving, conceivability will be a reliable guide to possibility.

Finally, one might combine the two strategies, either by relativizing types of conceiving to types of subject-matter or by restricting both domains on independent grounds. Perhaps for each sort of subject-matter, there is an associated form of conceiving that is possibility-revealing in the requisite sense. Or perhaps only certain sorts of subject-matter are tractable in this way, and only on the basis of certain sorts of conceiving.

²⁸ That is, that qualities exist that are actually located in space and time and are fully capable of bi-location.

In each case, the challenge is to come up with characterizations that are both sufficient to the task at hand and non-circular. For, of course, one might define ‘conceivable’ in such a way that P is not *really* conceivable unless P is possible. While this will solve the problem of reliability, unless there is an independent way of determining that we are conceiving in the relevant sense (and not merely seeming to conceive), the practical significance of the link will be negligible. And one might safely claim that one class of propositions for which conceivability implies possibility is the class of propositions that are possible—but here again, the practical significance of the observation is negligible.

How one goes about providing useful characterizations of the relevant notion of conceivability and the relevant class of propositions will depend on the sorts of answers one offers to the questions raised in the two previous sections—that is, on the views one has about the nature of possibility, the nature of conceiving, and the resultant reasons (if any) for expecting the latter to be a reliable guide to the former. Issues of this sort are addressed, directly and indirectly, by nearly all of the volume’s authors.

1.4 Applications

Assuming that one can successfully establish a link between (certain sorts of) conceivability and possibility (within certain realms), one might wonder what philosophical work the connection can do. Of course, it is of independent philosophical interest to know what is and is not metaphysically possible. But it is the strategy of employing conceivability arguments to establish claims of identity and distinctness that has been the primary target of such reasoning.

The idea behind such arguments is the following. If it is possible that a exist without b, it seems to follow that, as a matter of fact, a and b are distinct: after all, nothing can exist without itself, so if a *is* b, then presumably a cannot exist *without* b.²⁹ But if this line of reasoning is correct, and conceivability is a guide to possibility, then the mere conceivability of a’s existing without b will be sufficient to establish the possibility of a’s existing without b, which in turn will be sufficient to establish the actual distinctness of a and b.

This means that conceivability–possibility arguments are potentially quite powerful: if the conceivability of a’s existing without b is capable of revealing the metaphysical possibility of a’s existing without b, then conceivability would be able to reveal something about the actual lay of the land: namely, a’s distinctness from b. Faced with such a connection between actuality and possibility,

²⁹ For a challenge to this view, see Della Rocca, Ch. 5 below.

two attitudes towards the epistemic role of conceiving suggest themselves. The optimist will insist that since conceiving is a guide to possibility, it can reveal the actual distinctness of things; the pessimist will insist that since conceiving can never reveal actual facts of identity and distinctness, it is ill-suited to judging possibility in such cases.³⁰

The most famous of such arguments, of course, is that which attempts to establish some sort of mind–body dualism on conceivability grounds alone.³¹ If it is conceivable that mind exist without body, and conceivability implies possibility, then it is possible that mind exist without body, and thus actual that mind and body are distinct. The argument in question can be traced to Descartes, and it is for this reason that, as noted above, we devote the next section (2) of the introduction to a detailed presentation of Descartes’s version of it, in the context of a more general presentation of his (and Hume’s) views on conceivability and possibility. Contemporary interest in the question, however, is due largely to the work of Saul Kripke, and in the subsequent section (3), we introduce the Kripkean framework, connecting Kripke’s discussion to more general issues of conceivability and possibility. Even more recently, there has been a groundswell of interest in the relation between the viability of such arguments and so-called two-dimensionalist modal logics; in section 4, we present a framework for thinking about this most recent round of the debate between optimists and pessimists. With this background in place, we hope that even the previously uninitiated reader will be in a position to appreciate the volume’s papers, the contents of which are summarized in section 5.

³⁰ Note that on some metaphysical views, the connection between possibility and actuality runs so deep as to (apparently) render pessimism mandatory. Consider, for example, the causal theory of properties, according to which the causal role of a property is essential to that property. On such a conception, the shortcomings of conceivability as a guide to nomological possibility would seem to carry over to metaphysical possibility: since the (metaphysical) essence of a property is tied to the actual laws that govern it, our ability to mix and match qualities in thought and imagination will cause us to neglect certain necessary connections. Commitment to an optimistic perspective, meanwhile, would seem on the face of it to license dismissal of any such metaphysical picture. (But see Fine and Sidelle, Chapters 6 and 8 below.)

³¹ Such arguments are explicitly addressed in a number of chapters in this volume: see those by Bealer, Chalmers, Della Rocca, Sidelle, Stalnaker, Wright, and Yablo. Among the numerous discussions of the topic in the last decade or so, we direct readers who are particularly interested in conceivability–possibility issues to the following: (optimists) Bealer (1987, 1992), Chalmers (1996, 1999), Chalmers and Jackson (2002), Hart (1988), Jackson (1993, 1998); (pessimists, generally speaking) Balog (1999), Block and Stalnaker (1999), Levine (1998, 2000), Loar (1999), McLaughlin and Hill (1999), Yablo (2000). (Extensive lists of additional references can be found in the pieces mentioned.)

2 Descartes and Hume

As we have just noted, contemporary discussions of conceivability and possibility trace their ancestry to the early modern period—particularly to the writings of Descartes and, to some extent, Hume. In situating the volume’s papers in their historical context, then, it may be helpful for the reader to have some sense of how Descartes—and, where relevant, Hume—respond to the four questions posed in the introductory section.

Throughout, we will provide the reader with extensive quotations from the relevant texts. In this way, she may judge for herself to what extent these representative early modern discussions do and do not correspond to their contemporary analogues.³²

2.1 Possibility and Necessity

Like most philosophers before him, Descartes is concerned with a notion of possibility where what it is to be possible is to be non self-contradictory.³³ As he observes in the *Second Set of Replies*:

All self-contradictoriness or impossibility resides solely in our thought, when we make the mistake of joining together mutually inconsistent ideas; it cannot occur in anything outside the intellect. For the very fact that something exists outside the intellect manifestly shows that it is not self-contradictory, but possible. (CSM II, 108; AT VII, 152)³⁴

³² In the pages that follow references to the work of Descartes are to the three-volume edition *The Philosophical Writings of Descartes*, ed. John Cottingham, Robert Stoothof, Dugald Murdoch, and (in the case of vol. III) Anthony Kenny. References to vols. I and II respectively are of the form ‘CSM I’ or ‘CSM II’ followed by the relevant page number; references to volume III are in the form ‘CSMK’ followed by the relevant page number. The CSM(K) references are followed by references to the corresponding page in the standard twelve-volume edition of Descartes produced by Adam and Tannery; references to these pages are of the form ‘AT’ followed by the relevant page number. References to David Hume’s *Treatise of Human Nature* are made in the form [book, part, section] indicated respectively by [(large roman numeral), (small roman numeral), (arabic numeral)]. So, e.g., ‘*Treatise*, I. iii. 14’ refers to book I, part iii, section 14 of the *Treatise*. These references are followed by the corresponding page and paragraph in the Oxford Philosophical Texts (2000) edition of Hume’s *Treatise*, ed. David Fate Norton and Mary J. Norton. These are of the form ‘NN’, followed by [(page): (paragraph number)]. So, e.g., ‘NN 112: 23’ refers to the paragraph numbered 23 on p. 112 of the Oxford edition of the *Treatise*.

³³ For discussion of the historical background to Descartes’s conception of modality, see Alanen and Knuuttila (1988) and detailed bibliographic references provided therein.

³⁴ Descartes does not mean that impossibilities are explicitly contradictory. At CSM II, 108; AT VII, 151, he speaks of the relevant test as being whether or not a concept *implies* a contradiction (our emphasis). We will not pursue this topic further.

Descartes is also explicit about the subject-matter for such thoughts of possibility and necessity: such thoughts (unlike thoughts that encode impossible combinations of properties) concern a realm of essences that are external to our minds (though dependent on God's will³⁵). For Descartes, that is, eternal truths do not 'depend on the human intellect or on other existing things'; rather, they have reality whether or not anything actual possesses them. So writes Descartes in the *Fifth Meditation*:

When, for example, I imagine a triangle, even if perhaps no such figure exists, or has ever existed, anywhere outside my thought, there is still a determinate nature, or essence, or form of the triangle which is immutable and eternal, and not invented by me or dependent on my mind. (CSM II, 44–5; AT VII, 64)

On what grounds does he think this? Descartes's reasoning seems to be something like the following: (a) the properties of imagined entities outstrip those that we explicitly recognize in forming the image; (b) only those properties that we explicitly recognize in forming an image could be the products of our invention; therefore (c) at least some of the properties of imagined entities are not the products of our invention. Descartes presents this reasoning in the continuation of the passage just cited. The next sentence reads:

This is clear from the fact that various properties can be demonstrated of the triangle, for example . . . that its greatest side subtends its greatest angle . . . and since these properties are ones which I now clearly recognize whether I want to or not, even if I never thought of them at all when I previously imagined the triangle, it follows that they cannot have been invented by me. (CSM II, 45; AT VII, 64)

The move from this negative conclusion ('they cannot have been invented by me') to the positive claim ('there is . . . a determinate nature, or essence, or form . . . which is immutable and eternal') is left somewhat enthymematic. Still, it is clear that Descartes is committed to this stronger result. Later in the same *Meditation*, applying the principle just presented to the Ontological Argument for God's existence, Descartes writes:

from the fact that I cannot think of God except as existing, it follows that existence is inseparable from God. It is not that my thought makes it so, or imposes any necessity on the thing; on the contrary, it is the necessity of the thing itself, namely the existence of God, which determines my thinking in this respect. (CSM II, 46; AT VII, 67)

³⁵ For expressions of Descartes's notorious commitment to God's role in the creation of eternal truths, see, e.g., CSM II, 291, 294; AT VII, 432, 436. For the purposes of our discussion, we will set this complicated issue to one side.

For Descartes, that is, the direction of explanation runs from the necessities contained in the natures of things to the perceptions of the intellect, and not the other way around.

By contrast, an important theme in Hume's work is the mind-dependence of necessity:

Thus as the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones lies only in the act of the understanding, by which we consider and compare these ideas; in like manner the necessity or power, which unites causes and effects, lies in the determination of the mind to pass from the one to the other. (*Treatise*, I. iii. 14; NN 112: 23)³⁶

The modern reader will wonder why Hume does not say the same about possibility. If even logical necessity has its source 'in the act of understanding'—and, Hume insists, 'there is but one kind of necessity' (*Treatise*, I. iii. 14; NN 115: 33, italics in original) why not think the same is true of possibility? After all, for P to be possible is just for it not to be necessary that not-P. Though Hume never confronts the issue directly, one might extend his views on necessity as follows: the possibility of a given proposition is constituted by the capacity of the fancy to imagine its holding. As we noted in the prefatory section, this sort of projectivism about modality is currently out of favour—though (as we will discuss in section 3.2 below) it, like the full-blooded Cartesian picture, offers a sort of grounding for the conceivability–possibility link that is unavailable to those who hold certain sorts of post-Kripkean views.

2.2 *Conceiving and Other Faculties*

Descartes distinguishes sharply between intellection and understanding, on the one hand, and imagination and sensation, on the other. Whereas the former are general cognitive faculties that belong to us essentially *qua* thinking things, the latter are limited cognitive faculties that belong to us contingently *qua* embodied beings (cf. CSM II, 51; AT VII, 73). The faculties differ not only in their range of subject-matter—imagination being 'nothing but an application of the cognitive faculty to a body which is intimately present to it' (CSM II, 50; AT VII, 72)—but also in their phenomenology. As Descartes writes in the *Sixth Meditation*: 'When I imagine a triangle, for example, I do not merely understand that it is a figure bounded by three lines, but at the same time I also see the three

³⁶ Hume's ideas on this subject are, of course, most fully developed in his discussion of causality, where the appearance of necessity in the natural order is attributed to the mind's 'great propensity to spread itself on external objects'. See *Treatise*, I. iii. 14; NN 112: 25.

lines with my mind's eye as if they were present before me' (CSM II, 50; AT VII, 72). And again, in a July 1641 letter to Mersenne: 'whatever we conceive without an image is an idea of the pure mind, and whatever we conceive with an image is an idea of the imagination' (CSMK, 186; AT III, 395).

The primary task of metaphysical inquiry, according to Descartes, is the understanding of the immutable natures of things. And it is the intellect—not the imagination—that Descartes repeatedly credits with being suited to the task of revealing things as they actually are. Nonetheless, the faculty of imagination may serve as a useful supplement to the intellect in certain special cases. This ability to 'see [a shape] with my mind's eye as if [it] were present before me' (CSM II, 50; AT VII, 72) facilitates our grasping of certain simple truths about geometry and motion. As Descartes writes in a letter to Elizabeth on 28 June 1643:

body (i.e. extension, shapes and motions) can . . . be known by the intellect alone, but much better by the intellect aided by the imagination . . . and the study of mathematics, which exercises mainly the imagination in the consideration of shapes and motions, accustoms us to form very distinct notions of body. (CSMK, 227; AT III, 691–2)

Even in the case of geometrical figures, however, the imagination faces certain limitations: our finite representational capacities are restricted to the portrayal of fairly simple shapes. In his discussion of the chiliagon at the beginning of the *Sixth Meditation*, Descartes insists that while we can well understand what a chiliagon is, we cannot represent it in imagination:

if I want to think of a chiliagon, although I understand that it is a figure consisting of a thousand sides just as well as I understand the triangle to be a three-sided figure, I do not in the same way imagine the thousand sides or see them as if they were present before me. . . . I may construct in my mind a confused representation of some figure; but it is clear that this is not a chiliagon. For it differs in no way from the representation I should form if I were thinking of a myriagon, or any figure with very many sides (CSM II, 50; AT VII, 72)

And he is insistent that the faculty of imagination is an outright impediment to understanding when the subject-matter in question is metaphysical. In a letter to Mersenne dated 13 November 1639, he writes: 'The imagination, which is the part of the mind that most helps mathematics, is more of a hindrance than a help in metaphysical speculation' (CSMK, 141; AT II, 622).

In the case of body, the hindrance is due to the imagination's inability to grasp the relevant complexity in sufficient detail. In the *Second Meditation*, Descartes writes:

I can grasp that the wax is capable of countless changes of this kind, yet I am unable to run through this immeasurable number of changes in my imagination, from which

it follows that it is not the faculty of imagination that gives me my grasp of the wax as flexible and changeable. . . . I would not be making a correct judgment about the nature of wax unless I believed it capable of being extended in many more different ways than I will ever encompass in my imagination. I must therefore admit that the nature of this piece of wax is in no way revealed by my imagination, but is perceived by the mind alone. (CSM II, 20–1; AT VII, 31)

In the case of the soul, the hindrance is even more direct: since the soul cannot be depicted by imagery, the imagination is completely unsuited to reveal its nature. Earlier in the *Second Meditation*, Descartes writes:

it would . . . be a case of fictitious invention if I used my imagination to establish that I was something or other; for imagining is simply contemplating the shape or image of a corporeal thing . . . thus . . . none of the things that imagination enables me to grasp is at all relevant to this knowledge of myself which I possess. (CSM II, 19; AT VII, 28)

Thus, for Descartes, the proper vehicles for metaphysical inquiry are understanding and intellection, which are sharply distinguished from imagination. Hume, by contrast, blurs the distinction between imagining and conceiving. He writes:

‘Tis an established maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible.* (Treatise, I. ii. 2; NN 26: 8; italics in original)

His use of the expression ‘in other words’ to link the two italicized phrases suggests that he takes them to be equivalent. But the purportedly equivalent precepts differ in at least two crucial ways: the first concerns conceiving, the second imagining; and the first is concerned with the relation between bearing a certain attitude towards P and P’s *seeming* possible, whereas the second is concerned with the relation between bearing a certain attitude towards P and P’s as a matter of fact *being* possible.³⁷

The source of his indifference to the first is fairly straightforward: it is a consequence of his philosophical psychology (permitting only sensory experiences (‘impressions’) and their copies (‘ideas’) to serve as the building-blocks of the mind) coupled with an outright rejection of the rationalist distinction between imagination and understanding. (The source of his apparent indifference to the second is more controversial and would take us too far afield.)

While some attention has been paid in contemporary discussions to the question of how imagination and conception differ in their modality-revealing

³⁷ We assume here an equivalence between ‘nothing we imagine is (absolutely) impossible’ and ‘anything we imagine is (strictly) possible’.

character (see section 1.2 above), the question has not been explored fully. A number of the authors in this volume offer reasons for re-examining the issue.³⁸

2.3 *Reliability Conditions*

Descartes is explicit in endorsing the second of the strategies introduced in section 1.3 above: it is only certain sorts of conceiving that provide us with guidance concerning what is and is not possible. In particular, it is clear and distinct³⁹ understanding that provides us with the knowledge we seek. In the *Sixth Meditation*, he writes: 'I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it' (CSM II, 54; AT VII, 78). Or again, in *Comments on a Certain Broadsheet*:

the rule 'whatever we can conceive of can exist' is my own, [but] it is true only so long as we are dealing with a conception which is clear and distinct, a conception which embraces the possibility of the thing in question, since God can bring about whatever we clearly perceive to be possible. (CSM I, 299; AT VIII B, 352)⁴⁰

³⁸ For extended discussions of imagination, see chapters in this volume by Campbell, Currie, and Sorensen.

³⁹ Strictly speaking, the expression is redundant, as clarity is a special case of distinctness. In the *Principles of Philosophy*, Descartes explains: 'I call a perception "clear" when it is present and accessible to the attentive mind . . . I call a perception "distinct" if, as well as being clear, it is so sharply separated from all other perceptions that it contains within itself only what is clear' (CSM I, 207–8; AT VIII A, 22). Despite the redundancy, we follow Descartes in using the phrase 'clear and distinct' throughout.

⁴⁰ What is the relation between this and the truth rule of the *Third Meditation*: viz., that 'whatever I perceive very clearly and distinctly is true'? Prima facie, they differ markedly: the one is a rule connecting understanding with possibility, the other a rule connecting perception with truth. On a certain reading, however, the one can be seen as a special case of the other: if understanding involves a perception of possibility, then the conceivability–possibility rule is just the truth rule, applied to some particular subject–matter. Some of Descartes's writings support such an account. For instance, the *Sixth Meditation* claim that 'everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding' (CSM II, 54; AT VII, 78) is elsewhere presented in terms of perception of the possible: 'I boldly assert that God can do everything which I perceive to be possible' (CSM I, 363; AT V, 272, writing to More on 5 February 1649). Note that the notion of 'clear understanding' required is a strong one. In one sense we have an understanding of 'something is both round and square' clear enough to recognize it as necessarily false. But the sort of clear understanding required here is one allied to the intellectual perception of possibility.

In order for the clearness and distinctness condition to be a useful one, it must have two characteristics. First of all, clearness and distinctness of understanding must be *reliable*, in the sense that it serves as a fail-safe (or at least reasonably fail-safe) guide to the potentialities accruing to the immutable natures willed by God, and thereby to possibility. At the same time, clearness and distinctness of understanding must be *introspectively identifiable*, in the sense that we are able, as Descartes writes in the *Fifth Set of Replies* (to Gassendi), ‘to distinguish between the things that we really perceive clearly and those that we merely think we perceive clearly’ (CSM II, 260; AT VII, 379). As we noted above, one could introduce a notion of ‘clearness’ according to which, as a matter of stipulation, its being clear that *p* is a state that couldn’t obtain without *p* being true⁴¹—but unless we have some way of *telling* that we are in such a state, the notion will be of little epistemic value.

Two challenges might here be raised. First, why think that clearness and distinctness would be introspectively identifiable, let alone introspectively identifiable in the infallible way that Descartes seems to require? And second, even if we set aside the question of (infallible) introspective identifiability, why think that any such capacity would be a reliable guide to that which it is intended to illuminate?

The first challenge is a serious one—many of the objectors raise it in one form or another. Mersenne, for instance, asks (in the *Second Set of Objections*): ‘how can you establish with certainty that you are not deceived, or capable of being deceived, in matters which you think you know clearly and distinctly?’ (CSM II, 90; AT VII, 126). And Gassendi complains (in the *Fifth Set of Objections*) that ‘the difficulty does not seem to be about whether we must clearly and distinctly understand something if we are to avoid error, but about what possible skill or method will permit us to discover that our understanding is so clear and distinct as to be true and to make it impossible that we are mistaken’ (CSM II, 221; AT VII, 318).

Descartes is rather impatient with this line of attack—not only with Gassendi, but with the others as well—insinuating that it arises from a stubborn and deliberate unwillingness to engage in introspection. Such introspection, he suggests, would reveal to the objector his own obvious capacity to identify instances in which his understanding or perception is ‘transparently clear’ (*Fifth Meditation*, CSM II, 48; AT VII, 70)—indeed, cases where ‘perceptions are so transparently clear and at the same time so simple that we cannot even think of

⁴¹ Indeed, it is arguable that the English construction ‘it is clear that *p*’ is factive in this sense. If *p* turned out to be false, one would need to retreat to the claim that it merely *seemed* clear that *p*.

them without believing them to be true' (*Second Set of Replies*, CSM II, 104; AT VII, 145). Whether this answer is satisfactory is not a question we will attempt to adjudicate.

The second challenge is met as follows. As we noted at the outset, the sort of possibility to which clear and distinct understanding is supposed to be a guide is non-self-contradictoriness. But, as Descartes writes in the *Second Set of Replies*, 'Self-contradictoriness in our concepts arises merely from their obscurity and confusion' (CSM II, 108; AT VII, 152). And if self-contradictoriness arises only from obscurity and confusion, then if a concept is not obscure and confused, it will not be self-contradictory. But a concept that is not obscure and confused is a concept that is clear and distinct. And a concept that is not self-contradictory is a concept of something that is possible. So a concept that is clear and distinct is a concept of something that is possible.

Not only does clear and distinct understanding of a concept guarantee that the concept is of something possible; clear and distinct understanding is understanding *as possible*. As Descartes writes in the *First Set of Replies*, 'Possible existence is contained in the concept or idea of everything that we clearly and distinctly understand' (CSM II, 83; AT VII, 116). For Descartes, then, there is a straightforwardly circumscribed class of cases for which a certain sort of mental act provides us with reliable knowledge of possibility: things that are clearly and distinctly understood both *seem* possible and *are* possible.

Still, one might wonder what grounds the link between obscurity and confusion, on the one hand, and self-contradictoriness, on the other: why mightn't self-contradictoriness arise in the case of thoughts that are clear and distinct? (Of course, one could stipulate that thoughts that entailed a contradiction were thereby not clear. But, to rehearse a point made already, this would serve to problematize the claim of introspective identifiability.) At the very least, Descartes has a theological answer available. While acknowledging that 'In the case of our clearest and most careful judgments . . . if such judgments were false they could not be corrected by any clearer judgments or by means of any other natural faculty' (CSM II, 102–3; AT VII, 143–4). Descartes nonetheless maintains that we have reason to be sanguine. For, Descartes contends, it is incoherent to suppose that God would allow us to be deceived under such circumstances, since it would be contradictory to suppose 'anything should be created by him which positively tends towards falsehood' (CSM II, 103; AT VII, 144). As we noted at the end of section 1.1 above, for those of us unwilling to appeal to divine benevolence, the problem is not so easily escaped.

2.4 Applications

2.4.1 General Issues: Cut and Paste

While there are a number of conspicuous differences between Descartes and Hume at the level of substance⁴² and at the level of practice,⁴³ there are also important points of agreement. Of these, the one most central to contemporary discussions derives from Hume's use of the Cartesian thesis that when a pair of things are distinct,⁴⁴ each can exist without the other. Such arguments make direct use of conceivability–possibility reasoning: when we conceive (or imagine) one of two non-overlapping things without the other, we establish the thing's possible distinct existence, thereby establishing their actual distinctness. In the context of Hume's philosophical psychology, our ability to separate distinguishable objects in thought is a corollary of one of the most basic traits of the faculty in question, namely, '*the liberty of imagination to transpose and change its ideas*' (*Treatise*, I. i. 3; NN 12:4; italics in original), which has as a straightforward corollary that: 'Where-ever the imagination perceives a difference among ideas, it can easily produce a separation' (*Treatise*, I. i. 3; NN 12: 4).

From (a) the liberty of imagination and (b) the imagination–possibility link, we can extract from Hume something like a 'cut and paste' story about possibility. The liberty of imagination underwrites the following two principles:

Cut: If we can imagine a region that is (intrinsically) F adjacent to a (non-overlapping) region that is (intrinsically) G, then we can imagine a region that is F and withhold imagining a region that is G (where the regions can be either spatial or temporal).

⁴² Most striking, perhaps, are the sharply different conclusions they draw when it comes to questions of space and finitude. Descartes uses the principles that whatever is clearly and distinctly understood is possible to argue that the idea of an extended indivisible atom is incoherent (cf. CSMK, 202–3; AT III, 477–8; cf. also CSM I, 231–2; AT VIIIA, 51) and that a finitely bounded space is unimaginable (cf. CSM I, 232; AT VIIIA, 51); whereas Hume argues from facts about imaginability to the possibility that space is composed of a finite number of extended atoms—and thus to the conclusion that it could be only finitely divisible and of finite extent (cf. *Treatise*, I. ii. 2; NN 25:2).

⁴³ In general, Hume is quite ready to conclude that a thing is possible on the basis of his capacity to imagine it sensorily. See, e.g., his arguments '*that an object may exist, and yet be no where*' (*Treatise*, I. iv. 5; NN 154: 10, italics in original); '*that any object may be . . . annihilated in a moment*' (I. iv. 5; NN 164: 35); '*that any material particle can exist in the absence of any other, distinct quantity*' (I. ii. 5; NN 40: 3); or '*that it is possible for any given thing to exist without any cause at all*' (I. iii. 3; NN 56–8: 1–9). To the extent that Descartes would reject such reasoning, he would presumably do so on the grounds that the supposedly imagined content was not clearly and distinctly understood.

⁴⁴ Where this means non-overlapping.

Paste_p: If we can imagine a region that is (intrinsically) F, and we can imagine another region that is (intrinsically) G, then we can imagine adjacent F and G regions (where the regions can be either spatial or temporal).

If we accept these principles, along with the inference from imaginability to possibility, we now get:

Cut_p: If we can imagine a region that is (intrinsically) F adjacent to a (non-overlapping) region that is (intrinsically) G, then it is possible that there is a region that is F in a world where no non-overlapping region is G.

Paste_p: If we can imagine a region that is (intrinsically) F, and we can imagine another region that is (intrinsically) G, then it is possible that there is a region that contains two adjacent sub-regions, one F, another G.

Coupled with facts about identity, Cut_p and Paste_p—and their conceivability-based analogues—can serve as extremely powerful tools.⁴⁵ For example, it is on the basis of Cut_p, that Hume derives the conclusion that any material particle can exist in the absence of any other, distinct quantity. And one can readily reconstruct much of his reasoning about causation on the basis of Cut_p and Paste_p—for instance, his commitment to the thesis that any ‘cause’ may be followed by any ‘effect’ or that it is possible for any given thing to exist without any cause at all.

2.4.2 *Mind–Body Dualism*

By far the most famous application of conceivability–possibility reasoning is Descartes’s effort to establish the ‘real distinction’ between mind and body. The crucial thought here is that since we can clearly and distinctly understand the mind being apart from the body, it is possible that the mind be apart from the body; hence mind and body are possibly distinct; hence mind and body are actually distinct.

In somewhat more detail, the argument runs as follows. Using the principle—call it the *conceivability–possibility principle*—put forth in the *Sixth Meditation* and cited above, namely that ‘everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it’ (CSM II, 54; AT VII, 78). Descartes applies it to a particular case: ‘the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things

⁴⁵ For a modern descendant, see David Lewis: ‘I suggest we look to the Humean denial of necessary connections between distinct existences. To express the plenitude of possible worlds, I require a principle of recombination according to which patching together parts of different possible worlds yields another possible world’ (1986: 87–8).

are distinct, since they are capable of being separated, at least by God' (CSM II, 54; AT VII, 78). The argument for this principle—call it the *distinctness principle*—relies on the conceivability–possibility principle coupled with the thesis that for any x and y , if x and y can exist apart from each other, x and y are in reality distinct. Descartes goes on to apply the distinctness principle to a particular case of clear and distinct understanding of one thing apart from another: namely, the case of mind and body. The *Sixth Meditation* text continues:

I have a clear and distinct sense of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And, accordingly, it is certain that I am really distinct from my body, and can exist without it. (CSM II, 54; AT VII, 78)

Here, as elsewhere, Descartes places great weight on the clarity and distinctness restriction. As he notes in an August 1641 letter to 'Hyperaspites', our ability confusedly and obscurely to understand one thing apart from another would enable us to draw no such conclusion:

Of course someone whose eyes are unsteady may take one thing for two, as people often do when drunk; and philosophers may do the like . . . when in the same body they make a distinction between the matter, the form and the various accidents as if there were so many different things. In such cases . . . if they paid more careful attention they would notice that they do not have completely distinct ideas of the things they thus suppose to be distinct. (CSMK, 197; AT III, 435)

2.4.3 *Arnauld's Objection*

But how is Descartes so sure that his argument for the real distinction between mind and body is not trading off a similarly inadequate conception? Arnauld raises such a worry in the *Fourth Set of Objections*:

Suppose someone knows for certain that the angle in a semi-circle is a right angle, and hence that the triangle formed by this angle and the diameter of the circle is right-angled. In spite of this, he may doubt, or not yet have grasped for certain, that the square on the hypotenuse is equal to the squares on the other two sides. . . . [He] clearly and distinctly understand[s] that this triangle is right-angled, without understanding that the square on the hypotenuse is equal to the squares on the other sides. It follows on this reasoning that God, at least, could create a right-angled triangle with the square on its hypotenuse not equal to the square on the other sides. (CSM II, 141–2; AT VII, 201–2)

Moreover, Arnauld continues:

although the man in the example clearly and distinctly knows that the triangle is right-angled, he is wrong in thinking that the aforesaid relationship between the

squares on the sides does not belong to the nature of the triangle. Similarly, although I clearly and distinctly know my nature to be something that thinks, may I, too, not perhaps be wrong in thinking that nothing else belongs to my nature apart from the fact that I am a thinking thing? Perhaps the fact that I am an extended thing may also belong to my nature. (CSM II, 142–3; AT VII, 202–3)

Arnauld anticipates the obvious retort—that ‘the person in this example does not clearly and distinctly perceive that the triangle is right-angled’ (CSM II, 142; AT VII, 202). But, he responds, ‘how is my perception of the nature of my mind any clearer than his perception of the nature of the triangle? He is just as certain that the triangle in the semi-circle has one right angle (which is the criterion of a right-angled triangle) as I am certain that I exist because I am thinking’ (CSM II, 142; AT VII, 202). And he goes on to underscore the fact that we do not have a ‘complete and adequate conception’ of mind: ‘I conceive of it only inadequately, and by a certain intellectual abstraction’ (CSM II, 143; AT VII, 203).

But when a substance is thought about in abstracted terms, there may well be unnoticed elements of the nature of the substance to which the abstracted conception applies. Arnauld again offers an analogy from geometry:

Geometers conceive of a line as a length without breadth, and they conceive of a surface as length and breadth without depth, despite the fact that no length exists without breadth and no breadth without depth. In the same way, someone may perhaps suspect that every thinking thing is also an extended thing . . . although simply in terms of this power [thought], it can by an intellectual abstraction be apprehended as a thinking thing, in reality bodily attributes may belong to this thinking thing. (CSM II, 143; AT VII, 203–4)

So Arnauld’s concern is the following: Even if I have a clear and distinct sense of myself as simply a thinking, thing, I have no way of ruling out that my conception of thinking substance is the result of intellectual abstraction; if this is so, I have no way of ruling out that the nature(s) of which I have a conception include(s) a good deal more than thought, and hence no grounds for applying the distinctness principle to the case of mind and body.

2.4.4 *Descartes’s Reply*

Arnauld’s worry is widely echoed in contemporary discussions, and in this light, it is instructive to consider Descartes’s various replies.⁴⁶ Importantly, he does not try to defend the view that our conception of substance is ‘adequate’

⁴⁶ Because they are somewhat orthogonal to the central issue, we do not present in detail Descartes’s discussion of the disanalogies between the geometrical cases that Arnauld presents, and the particular case they are intended to illuminate; see CSM II, 157–9; AT VII, 224–6.

in the sense of containing ‘absolutely all the properties that are in the thing which is the object of knowledge’, since a ‘created intellect . . . can never know it has such knowledge unless God grants it a special revelation of the fact’ (CSM II, 155; AT VII, 220). Rather, he insists, what is required is that I ‘understand the thing well enough to know that my understanding is *complete*’—that is, that I understand the thing in question to be a ‘complete thing . . . a substance endowed with the forms or attributes which enable me to recognize that it is a substance’ (CSM II, 156; AT VII, 221–2), where ‘the notion of a *substance* is just this—that it can exist by itself, that is without the aid of any other substance’ (CSM II, 159; AT VII, 226; italics in original).

But, one might object, even if I know that everything that has the attribute of being a thinking thing is a substance, my conception of the nature of that substance may be deficient in exactly the ways at issue. For the sceptical hypothesis is not that mental substance needs the aid of another substance to exist; rather, the worry is that mental substance, while independent of all other substance, is such that a fully adequate conception of it reveals corporeality to be part of its nature.

Descartes’s discussion elsewhere is more helpful in advancing the dialectic. In the letter to Gibieuf dated 19 January 1642, he offers the following gloss on ‘completeness’:

the idea of a substance with extension and shape is a complete idea, because I can conceive it entirely on its own, and deny of it everything else of which I have an idea. Now it seems to me very clear that the idea which I have of a thinking substance is complete in this sense. (CSMK, 202; AT III, 475)

The idea seems to be that we have an insight that it is possible that there be a self-standing thing enjoying some determinate version of extension and shape and no other intrinsic properties⁴⁷ (or at least nothing else in our repertoire of determinables). By contrast, perhaps, we realize that no self-standing thing could have merely colour and nothing else, or motion and nothing else, or smell and nothing else. For a conception of a substance to be complete, then, is for it to specify some determinable(s) such that a thing could exist with just some version of those determinables.⁴⁸ And, insists Descartes, our conception of thinking substance is complete in just this way.

The point can be put even more straightforwardly, employing terminology from the *Principles of Philosophy*. To each substance, Descartes explains, there belongs one principal attribute ‘which constitutes its nature and essence, and to

⁴⁷ That it may have various relations to other things is beside the point.

⁴⁸ Among those determinables of which we have some notion.

which all its other properties are referred' (CSM I, 210; AT VIII A, 25). In the case of mind, this is thought; in the case of body, extension: 'Everything else which can be attributed to a body . . . is merely a mode of an extended thing; and . . . whatever we find in the mind is simply one of the various modes of thinking (CSM I, 210; AT VIII A, 25). Thus, concludes Descartes, 'we can easily have two clear and distinct notions or ideas, one of created thinking substance, and the other of corporeal substance, provided we are careful to distinguish all the attributes of thought from the attributes of extension' (CSM I, 211; AT VIII A, 25).

Readers may well wonder why Descartes is entitled to move from the possibility that there be a thinking thing that exists without corporeality to the general conclusion that *all* thinking things can exist without corporeality: after all, perhaps some thinking things are essentially embodied, others not. Descartes's thought here seems to be that, since the attribute of thought is sufficient for yielding a self-standing substance, then once one has a thing that *inter alia* thinks, one can strip away all other attributes and still be left with the same self-standing substance. After all, if the addition of corporeality to a thinking thing would not destroy it, why should subtraction of corporeality have a different effect?

While there is much more to be said on these matters, further exploration of the Cartesian argument lies beyond our current purview.⁴⁹ Instead, we turn to its distinguished contemporary revival (and refinement) in the writings of Saul Kripke.⁵⁰

3 Kripke

Even authors who disagree with Kripke's fundamental picture tend to present their arguments against an implicitly or explicitly Kripkean backdrop—including most of the authors in this volume.

3.1 *Key Distinctions*

Most readers will be aware of three key distinctions that drive Kripke's discussion: the distinction between rigid and non-rigid designators, the distinction

⁴⁹ For discussion of Descartes's argument with particular attention to conceivability-possibility questions, see van Cleve (1983) and Yablo (1990).

⁵⁰ While there are numerous important moments in the discussion of conceivability-possibility arguments from Hume to the present—to mention but three, consider Kant's Copernican turn, positivism's conventionalist approach to modality, and Quine's scepticism about the coherence of modal discourse—it is Kripke's *Naming and Necessity* that sets the stage for most contemporary discussions of our topic.

between reference-fixers and definitions, and the distinction between apriority and necessity (and, correspondingly, between aposteriority and contingency). But, given the centrality of these distinctions to Kripke's diagnosis of how and when conceivability–possibility arguments can be successfully deployed, it is worth reviewing them before continuing.

3.1.1 *Rigid versus Non-Rigid Designators*

A *rigid designator* is a term that picks out the same thing 'across possible worlds'.⁵¹ Names ('Pierre', 'London'), demonstratives ('that', 'I'), and natural kind terms ('water', 'light') are rigid designators—they designate the same individual or kind in all possible worlds (where that individual or kind exists). A term that designates *non-rigidly*, by contrast, may pick out different objects in different worlds. Definite descriptions are generally treated as non-rigid⁵² ('the most interesting person in the room', 'the oldest line in the book')—these terms designate different individuals or categories in different possible worlds.

Often, we pick out the same individual in the actual world by employing diverse descriptions and names: 'Cicero', 'Tully', 'the prosecutor of Cataline', and 'the man who was consul in 63BC' are all ways of referring to the same individual. In many contexts (so-called transparent contexts), which of the terms we choose will make no difference to the truth-value of the sentence in which it appears. The following sentences, for example, stand or fall together:

- (1) Cicero was wise.
- (2) Tully was wise.
- (3) The prosecutor of Cataline was wise.
- (4) The man who was consul in 63BC was wise.

By contrast, suppose we wish to consider how what is expressed by (5) and (6) respectively might not have been true:

- (5) Cicero is wise.
- (6) The man who was consul in 63BC is wise.

Because 'Cicero' is rigid, it picks out the same individual in all possible worlds. So to entertain a possible world where what is said by (5) is false, we need to think of a possible situation in which a certain actual individual—Cicero—is not wise. By contrast, because 'The man who was consul in 63BC' is non-rigid,

⁵¹ More precisely, in all possible worlds where it picks out anything at all. For the sake of simplicity, we will generally omit this caveat in our discussion below. And we will set to one side the issues raised by reference failure and its cousins.

⁵² There are exceptions—for instance, 'the smallest prime number'.

it picks out different individuals in different possible worlds. So to entertain a possible world where what is said by (6) is false, we need, for example, only to think of a possible situation in which someone or other—never mind whether he is Cicero—is, in that situation, both non-wise and consul in 63BC.

With these considerations in mind, we can see that the second of the following sentences expresses a proposition that might have been true, whereas the first does not:

- (7) Cicero is not Tully.
- (8) Cicero is not the prosecutor of Cataline.

Because ‘Cicero’ and ‘Tully’ are rigid designators, each picks out the same individual in all possible worlds. So, since Cicero is Tully in the actual world, Cicero is Tully in all possible worlds, hence (7) does not express a proposition that might have been true.⁵³ Identity claims flanked by a pair of rigid designators are, if true, necessarily true, and if false, necessarily false.⁵⁴ But the same does not hold in the case of (8). ‘The prosecutor of Cataline’ may pick out different individuals in different possible worlds. So, even though Cicero is the prosecutor of Cataline in the actual world, he is not the prosecutor of Cataline in all possible worlds; hence (8) expresses a proposition that might have been true.

3.1.2 *Reference-Fixers versus Definitions*

To the extent that names are treated as rigid designators and descriptions as non-rigid designators, a description cannot be used to *define* a name. That is, we cannot give a rule for determining the name’s meaning across possible worlds that coincides with the rule for determining the meaning of the description, for the (rigid) name picks out the same individual across possible worlds, whereas the (non-rigid) description picks out different ones. There is a second way, however, that a name and a description might be associated: the description might be used to ‘fix the reference’ of the name. In this sense, we can give a rule for determining that name’s meaning across possible worlds that *makes use of* the rule for determining the meaning of the description; in particular, we can let the name pick out across possible worlds whatever the description picks out in the actual world. Here again the (rigid) name will pick out the same individual across possible worlds, whereas the (non-rigid) description will

⁵³ This follows from the nature of rigid designation, coupled with the highly intuitive theses of the necessity of identity and diversity (implicit in Descartes’s musings about separability): $\Box [(x) (y) (x = y) \supset \Box (x = y)]$; $\Box [(x) (y) (x \neq y) \supset \Box (x \neq y)]$

⁵⁴ For a challenge to this thesis, see Della Rocca, Ch. 5 below.

pick out different ones. But what the name picks out across possible worlds will, in some important intuitive sense, *depend on* what the description picks out in the actual world. Kripke writes:

suppose we say, 'Aristotle is the greatest man who studied with Plato'. If we used that as a *definition*, the name 'Aristotle' is to mean 'the greatest man who studied with Plato'. Then of course in some other possible world that man might not have studied with Plato and some other man would have been Aristotle. If, on the other hand, we merely use the description to *fix the referent* then that man will be the referent of 'Aristotle' in all possible worlds. The only use of the description will have been to pick out to which man we mean to refer. But then, when we say counterfactually 'suppose Aristotle had never gone into philosophy at all', we need not mean 'suppose a man who studied with Plato, and taught Alexander the Great, and wrote this and that, and so on, had never gone into philosophy at all', which might seem like a contradiction. We need only mean, 'suppose that *that man* had never gone into philosophy at all'. (1980: 57, italics in original)

What goes for proper names goes for other rigid terms, such as natural kind terms. Here again, we might introduce a term that is intended to be rigid by use of a description that is not: 'let *jooce* name the liquid that is in that test-tube', 'let *woozle* name the species of animal that raided the hen-house', and so on. Because we are generally ignorant of the 'real definition' or 'real essence' of the kinds that comprise the joints in nature, our access to the referents of natural kind terms frequently proceeds in this way: we (rightly or wrongly) reckon some (set of) manifest qualities as indicating the presence of a natural kind, and we use those qualities to fix the reference of what we take to be a rigid term.⁵⁵ Thus it is plausible that the reference of 'light' was fixed by the visual appearance it produces, 'heat' by the sensation it engenders, and so on. Idealizing somewhat for heuristic purposes, we might even suppose that the term 'heat' was explicitly introduced via the dictum: 'Let "heat" denote whatever kind it is in the actual world that produces such-and-such sensations.'

Recognizing that manifest qualities can serve as mere reference-fixers allows us to explain our apparent capacity coherently to conceive of situations where the relevant kind is present but produces different manifest qualities (light produces sensations of darkness, heat sensations of chill) and situations where the relevant kind is absent but where some other kind(s) produce(s) the qualities that serve to fix the reference of the kind term in the actual world (something other than light produces sensations of brightness, something other than heat produces sensations of warmth).

⁵⁵ For discussion of complications related to this strategy, see Wright, Ch. 12 below.

3.1.3 *A Priority versus Necessity*

On one traditional conception of their relation, apriority and necessity coincide, as do their counterparts, aposteriority and contingency.⁵⁶ On this traditional conception, all truths knowable a priori are necessary truths, and all necessary truths are (at least in principle) knowable a priori. Correlatively, all truths knowable only a posteriori are contingent truths, and all contingent truths are knowable only a posteriori. One of the central accomplishments of *Naming and Necessity* was explicitly to challenge this purported equivalence.

Whether a proposition is a priori or a posteriori, observes Kripke, is an epistemic question: it turns, roughly, on whether we can know the proposition to be true independently of any experience. Whether a proposition is necessary or contingent is a metaphysical question: it turns on whether its truth is independent of what the world is like. The two concepts deal ‘with two different domains, two different areas, the epistemological and the metaphysical’ (Kripke 1980: 36). Kripke continues:

More important than any particular example of something which is alleged to be necessary and not *a priori* or *a priori* and not necessary, is to see that the notions are different, that it’s not trivial to argue on the basis of something’s being something which maybe we can know only *a posteriori*, that it’s not a necessary truth. It’s not trivial, just because something is known in some sense *a priori*, that what is known is a necessary truth. (1980: 38–9)

Not only do the concepts differ in intension; they also differ in extension: there are necessary a priori and contingent a posteriori truths, but there are also contingent a priori and necessary a posteriori truths.

Why might one have thought otherwise? ‘I guess it’s thought that . . . if something is known *a priori* it must be necessary, because it was known without looking at the world. If it depended on some contingent feature of the actual world, how could you know it without looking?’ (Kripke 1980: 38). The answer to this latter question exploits the reference–fixer/definition distinction. Suppose I introduce ‘Bob’ by the reference–fixer ‘the number of the planets’. Then the sentence

(9) Bob is the number of the planets

will be a priori knowable—but it will not be necessary. It will be a priori knowable because I can know without looking at the world that the numbers

⁵⁶ For example, on the positivist picture, both ‘a priori’ and ‘necessary’ were to be explicated in terms of analyticity (which in turn, for the positivists, was to be explained in terms of convention). On the Kantian picture, necessity is to be explained in terms of apriority (though apriority outstrips analyticity). Both are species of this ‘traditional conception’.

designated by ‘Bob’ and ‘the number of the planets’ are the same,⁵⁷ and hence that (9) is true; but it will not be necessary because ‘Bob’ is a rigid term, whereas ‘the number of the planets’ is not, so there will be worlds where ‘Bob’ picks out one number (namely, the number of planets in the actual world), whereas ‘the number of the planets’ picks out another (namely, the number of planets in that world).⁵⁸ By similar reasoning, if ‘heat’ has its reference fixed by ‘the phenomenon that produces sensations of warmth’, then the statement

(10) If heat exists, it produces sensations of warmth

will likewise be contingent a priori.

The same goes for the other direction of the supposed necessity–apriority equivalence. Initially, it may appear that all necessary truths are knowable (at least in principle) a priori. The thought behind this, suggests Kripke, is something like the following: ‘if something not only happens to be true in the actual world but is also true in all possible worlds, then, of course, just by running through all the possible worlds in our head, we ought to be able with enough effort to see, if a statement is necessary, that it is necessary, and thus know it *a priori*’ (1980: 38). Kripke mentions in passing possible counter-examples from mathematics: it is far from trivial that every mathematical truth is a priori knowable, while it is fairly clear that every mathematical truth is necessary. But the best-known class of counter-examples proceeds via the observation that there are necessary truths that, while unknowable a priori, are knowable a posteriori. The most straightforward examples are provided by identities expressed by statements in which the copula is flanked by rigid designators, but where empirical investigation is required to determine whether the terms co-refer. For example,

(11) Hesperus is Phosphorus

(12) Water is H₂O

are necessary but a posteriori. They are necessary because the rigid designators on each side of the copula co-refer in the actual world, and hence in all possible worlds; but they are knowable only a posteriori because the co-reference of the terms is itself contingent. Other examples can be generated by appeal to the real essence of any given thing or kind, in cases where reference to the thing

⁵⁷ Of course, for all I know a priori, that number may be zero.

⁵⁸ While Kripke’s examples of contingent a priori identities involve one rigid and one non-rigid designator, there are examples of contingent a priori identities that do not. Suppose Saul is a book in a row of books, and suppose I introduce ‘David’ by the reference-fixer ‘Let “David” be the book two to the right of Saul’. Then the statement ‘If Saul and David exist, then the book to the right of Saul is the book to the left of David’ expresses a contingent a priori truth.

or kind does not require epistemic access to that essence.⁵⁹ Thus the following sentences are necessary a posteriori:

- (13) Water has hydrogen as a constituent.⁶⁰
- (14) Prince Charles is the son of Queen Elizabeth.⁶¹

3.2 *Conceivability and Possibility*

3.2.1 *The Problem*

On the traditional picture adverted to above, there is a straightforward, if idealized, way to explain the connection between conceivability and possibility. Recall that P is possible iff it is not necessary that not-P. Let us introduce, as a term of art, that P is *Conceivable* iff it is not a priori that not-P. If all and only a priori truths are necessary truths, then all and only Conceivable truths are possible truths. For the Conceivable truths are just those whose negations are not a priori, and the possible truths are just those whose negations are not necessary. And since the latter two classes coincide, so do the former two.

As we noted above, some version of this form of reasoning is implicit both in Descartes and in Hume, though the two accounts differ in their order of explanation. Put crudely, in Descartes, the direction of explanation runs from the metaphysical to the epistemic: something is knowable by reflection because it is necessary; in Hume, the direction of explanation runs from the epistemic to the metaphysical: something is necessary because the mind treats it as such.⁶² Yet in each case the metaphysical and epistemic categories coincide.

⁵⁹ So whether an identity is a priori or a posteriori knowable will depend on how the references of its terms are fixed. If we can know that humans are essentially humans—let's not worry about Jesus—then if a's reference is fixed by the attribute 'human', one can know a priori that if a exists, a is essentially human. But if a name b refers to a human without having its reference fixed by the attribute of humanity, then the corresponding conditional may be knowable only a posteriori.

⁶⁰ Whether we can know a priori the necessary statement that H₂O has hydrogen as a constituent is a trickier question; on this issue, Kripke offers little explicit guidance.

⁶¹ Assuming that our parental origins are necessary.

⁶² This oversimplifies the views of the actual historical figures. For example, Descartes is explicitly sensitive to the possibility of truths that transcend our cognitive powers altogether, so unless a priori means something like '*a priori* knowable by God', then what we described as the traditional equation will not apply straightforwardly to him. But it remains plausible that Descartes believes that *for any proposition we can grasp*, it is necessary iff it is a priori knowable, and thus possible iff it is conceivable. Note, though, that the fact that there may be ungraspable propositions suggests that Conceivability (as defined above) and conceivability (in the intuitive sense) may come apart fairly radically. These issues are discussed at some length in Chalmers's contribution to this volume (Ch. 3).

On the post-Kripkean picture, however, no such explanation is available. For if there are a posteriori necessities and a priori contingencies, no such grounds can be appealed to in establishing a conceivability–possibility link. On the post-Kripkean picture, even if it is not necessary that not-P, it may still be a priori that not-P (contingent a priori); and even if it is not a priori that not-P, it may still be necessary that not-P (necessary a posteriori). But then, by substitution, it may be possible that P but not Conceivable that P, or Conceivable that P but not possible that P. Thus the contingent a priori seems to guarantee that there will be cases of possibility without Conceivability; the necessary a posteriori seems to guarantee that there will be cases of Conceivability without possibility.

In the face of such discrepancies, some have been inclined to give up on the link between conceivability and possibility. As Hilary Putnam writes in ‘The Meaning of “Meaning” ’:

we can perfectly well imagine having experiences that would convince us (and that would make it rational to believe that) water isn’t H₂O. In that sense, it is conceivable that water isn’t H₂O. It is conceivable but it isn’t logically possible! Conceivability is no proof of logical possibility . . . Human intuition has no privileged access to metaphysical necessity. (1975b: 233)⁶³

3.2.2 *The Diagnosis*

Kripke, however, is more sanguine. While it is not his ambition to offer some general epistemology of modality—in favour of his positive modal claims, he is content to appeal to the fact that they are highly intuitive—he does offer a strategy for re-establishing a link between intuitions of possibility and what is in fact possible. There will, he acknowledges, be cases where it seems (at least to some of us) to be possible that not-P, but where, in fact, P is necessary: a posteriori necessities provide us with a class of such cases.⁶⁴ In such cases, we will be faced with an *illusion of possibility*—not-P will *seem* possible, though in fact it is not. At the same time, there will also be cases where it seems (at least

⁶³ In earlier writings, for instance, in his (1962) ‘It ain’t necessarily so’, Putnam is similarly dubious of the possibility- and necessity-revealing powers of conceivability and apriority respectively. But there his motivation is essentially Quinean: ‘the traditional philosophical distinction between statements necessary in some eternal sense and statements contingent in some eternal sense is not workable’ (1975a: 248).

⁶⁴ Does Kripke think that it seems to ordinary people to be metaphysically possible that Hesperus is not Phosphorus and metaphysically necessary that the metre stick is a metre long, or is this an illusion that has occurred only to philosophers in the grip of a picture? We shall not take a stand on this finer point of exegesis. For more on this topic, see Bealer, Ch. 1 below, and Della Rocca, Ch. 5 below.

to some of us) to be necessary that P, but where, in fact, not-P is possible: a priori contingencies provide us with a class of such cases. In such cases, there will be an *illusion of necessity*—P will seem necessary, though in fact it is not. But Kripke also offers a general strategy for dealing with such cases. We can explain these illusions by adverting to certain conflationations that systematically give rise to them, notably:

- (i) between reference-fixing descriptions and the rigid terms they introduce
- (ii) between the possibility of a community in an epistemically analogous situation saying something true by a sentence and the possibility of what a sentence says being true.

3.2.3 *Reference-fixing Surrogates*

Recall the following examples:

- (9) Bob is the number of the planets.
- (11) Hesperus is Phosphorus.

The contingent a priori (9) has seemed to some to express a necessary truth, the necessary a posteriori (11) a contingent one. But, of course, these appearances are misleading: the negation of (11) expresses a necessary falsehood, whereas the negation of (9) expresses a possible truth. (There is no possible world in which Hesperus is not Phosphorus, but there is a possible world in which Bob is not the number of the planets.) In both cases, the explanation for the modal status of the proposition relies on the distinction between rigid and non-rigid designation. ‘Hesperus’ and ‘Phosphorus’ are rigid designators that pick out the same object in the actual world, and hence in all possible worlds. But whereas ‘Bob’ is rigid, ‘the number of planets’ is not, so there are worlds where ‘Bob’ picks out one object and ‘the number of planets’ another.

Why, then, does the negation of (11) seem to express a possible truth, whereas the negation of (9) does not? In this case, the explanation relies on the fact that our modal intuitions will go astray in so far as we conflate a reference-fixer with the term it introduces.

Suppose that ‘Hesperus’ and ‘Phosphorus’ were introduced by the reference-fixers ‘the heavenly body that appears at thus-and-such location in the morning sky’ and ‘the heavenly body that appears at thus-and-such location in the evening sky’, respectively. There are certainly worlds where these descriptions pick out distinct objects. If we are careless in distinguishing the terms in question from the descriptions by which their reference is fixed, we may confusedly think that the question whether it is possible that Hesperus is not Phosphorus is the same as the question whether it is possible that the heavenly

body that appears at thus-and-such location in the morning sky is not the heavenly body that appears at thus-and-such location in the evening sky.⁶⁵ A parallel explanation can be offered for the apparent necessity of (9), where 'Bob' has its reference fixed by the description 'the number of planets'. Here again, if we are careless in distinguishing terms from reference-fixers, we may find ourselves thinking that the question whether it is possible that Bob is not the number of planets is the same as the question whether it is possible that the number of planets is not the number of planets.

Let us generalize the point. Call a statement s_2 the *reference-fixing surrogate* of s_1 when each term in s_1 that was introduced by a reference-fixing description is replaced in s_2 by the reference-fixing description itself. The reference-fixing surrogate of an a posteriori necessary truth will be contingent, and the reference-fixing surrogate of an a priori contingent truth will be necessary. It is no wonder that in so far as statements are conflated with their reference-fixing surrogates, modal illusions arise.⁶⁶

3.2.4 *Epistemic Duplicates*

There is a second, though related, strategy deployed by Kripke for explaining modal illusion. He does not, after all, think that all rigid designators are introduced by reference-fixing descriptions: sometimes such terms are introduced by a baptismal act whereby an individual or kind is ostended.⁶⁷ And even when a term is originally introduced by a reference-fixing description, a competent user of the term need not be familiar with the description that introduced it (and so may lack grounds for conflating the term with its associated reference-fixing description). Return to the case of 'Hesperus' and 'Phosphorus': What is certainly possible (whether or not the terms were originally introduced by reference-fixing description or ostension) is that there be a community meeting two conditions: (a) their epistemic situation is identical to that of our

⁶⁵ Cf. Kripke: 'Of course, it is only a contingent truth . . . that the star seen over there in the evening is the star seen over there in the morning . . . But that contingent truth shouldn't be identified with the statement that Hesperus is Phosphorus' (1980: 105).

⁶⁶ Kripke writes: 'Let " R_1 " and " R_2 " be the two rigid designators which flank the identity sign . . . The references of " R_1 " and " R_2 ", respectively, may well be fixed by nonrigid designators " D_1 " and " D_2 " . . . Then although " $R_1 = R_2$ " is necessary, " $D_1 = D_2$ " may well be contingent, and this is often what leads to the erroneous view that " $R_1 = R_2$ " might have turned out otherwise' (1980: 143–4).

⁶⁷ 'Usually a baptizer is acquainted with some sense with the object he names and is able to name it ostensively' (1980: 96 n. 42).

term-introducing community's, and (b) they express something false by the sentence 'Hesperus is Phosphorus':

There certainly is a possible world in which a man should have seen a certain star at a certain position in the evening and called it 'Hesperus' and a certain star in the morning and called it 'Phosphorus'; and should have concluded—should have found out by empirical investigation—that he names two different . . . heavenly bodies. . . . And so it's true that given the evidence someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two heavenly bodies 'Hesperus' and 'Phosphorus', without their being identical. (1980: 103–4)

Generalizing, let us call a community C_1 an *epistemic duplicate* of a community C_2 just in case for every proposition known by someone in C_1 , someone in C_2 knows that proposition (or an analogous proposition), and vice versa.⁶⁸ Even when an utterance expresses a necessary truth in our mouths, there may nevertheless be some possible epistemic duplicate community for whom it or its analogue would express a (necessary) falsehood. Insofar as we conflate the latter possibility with the possible falsity of the original statement, modal illusion will arise. Meanwhile, even when an utterance expresses a contingent truth in our mouths, there may nevertheless be no possible epistemic duplicate community for whom it or its analogue would express a falsehood. Insofar as we conflate the latter impossibility with the necessary truth of the original statement, modal illusion will similarly arise.⁶⁹

3.2.5 *General Morals*

The modal illusions that we have been trying to make sense of arise from some sort of genuine modal insight, distorted by some sort of conflation. Once the conflation is ironed out, the modal insight can be expressed in a trouble-free way. Thus:

The loose and inaccurate statement that gold might have turned out to be a compound should be replaced (roughly) by the statement that it is logically possible that

⁶⁸ Of course, if the members of C_1 and C_2 are wholly distinct, no one in C_1 will know quite what a member of C_2 knows when that member utters 'I exist'. But that member's 'counterpart' will, in some intuitive sense, know a counterpart singular proposition—hence the 'analogous proposition' clause. The task of making the notion of 'analogous' proposition precise in this sense is a daunting one—which we leave to others. For discussion of these issues, see the contributions to this volume by Bealer, Chalmers, and Yablo.

⁶⁹ Note that in the case of a term introduced by a reference-fixing description, epistemic duplicate confusion will induce reference-fixing surrogate confusion.

there should have been a compound with all the properties originally known to hold of gold. (1980: 142–3)⁷⁰

Similarly, the loose and inaccurate statement that water might not have been H₂O is a faulty attempt to convey the perfectly acceptable thought that there might have been a community that was an epistemic duplicate of our predecessors that rigidly denoted some stuff other than H₂O by ‘water’.⁷¹ Those who conflate this possibility with the possibility that water is not H₂O may be duped into thinking that any possible stuff with those manifest qualities originally known to hold of water counts as water. They may also be induced to suppose that in possible worlds where H₂O does not generate those manifest qualities originally known to hold of water, it will not count as water. (After all, in a world where XYZ has those properties and H₂O does not, the epistemic duplicate will not refer to H₂O by ‘water’.) The latter style of mistake is particularly tempting, Kripke notes, in the case of heat. While the sensation of heat fixes the reference of ‘heat’, it does not define it. Heat, after all, is identical to molecular motion:

Suppose we imagine God creating the world; what does He need to do to make the identity of heat and molecular motion obtain? Here it would seem that all He needs

⁷⁰ And again: ‘it could have turned out that P entails that P could have been the case. What, then, does the intuition that the table might have turned out to have been made of ice or of anything else . . . amount to? I think that it means simply that there might have been a *table* looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. In other words, I (or some conscious being) could have been *qualitatively in the same epistemic situation* that in fact obtains, I could have the same sensory evidence that I in fact have, about a *table* which was made of ice. . . . Something like counterpart theory is thus applicable to the situation, but it applies only because we are *not* interested in what might have been true of *this particular table*, but in what might or might not be true of a *table* given certain evidence’ (1980: 141–2, italics in original).

⁷¹ Note that the strategy just described will not work for certain sorts of statements: for instance, those concerning mathematics. (Wright (Ch. 12 below) discusses potential implications of this fact.) If Goldbach’s conjecture is true, then the statement that it might have turned out to be false should presumably *not* be replaced, even roughly, by the statement that one might, in an epistemically identical situation, encounter an analogous mathematical conjecture that is false. Rather, the statement that Goldbach’s conjecture might have turned out to have been false is to be replaced, roughly, by an acknowledgement of our uncertainty about the conjecture’s truth or falsity. Kripke writes: ‘there’s one sense in which things might turn out either way, in which it’s clear that that doesn’t imply that the way it finally turns out isn’t necessary. For example, the four color theorem might turn out to be true and might turn out to be false. It might turn out either way. . . . Obviously, the “might” here is purely “epistemic”—it merely expresses our present state of ignorance, or uncertainty’ (1980: 103). Note that certain constructions are better suited than others to express the ‘might’ of ignorance: it is relatively difficult to think of the sentence ‘It might not have been the case that water is H₂O’ being used in this way (unless the utterer wishes to express the thought that perhaps in the past water had a different chemical constitution). For a general strategy for distinguishing uses of the ‘epistemic might’, see Bealer, Ch. 1 below.

to do is to create the heat, that is, the molecular motion itself. . . . How then does it appear to us that the identity of molecular motion with heat is a substantive scientific fact, that the mere creation of molecular motion still leaves God with the additional task of making molecular motion into heat? This feeling is indeed illusory, but what *is* a substantive task for the Deity is the task of making molecular motion felt as heat. (1980: 153, italics in original)

Once our thinking has been purged of such conflations, Kripke holds, we can happily rely on the modal intuitions that remain intact.

3.3 *Dualism*

One such case, Kripke argues, is the case of our modal intuitions concerning the distinctness of mental and physical phenomena. Here, he maintains, the conceivability of their distinctness *is* a guide to possibility—and hence to actuality.

Consider some candidate materialist claim of property identity:

(15) Pain is C-fibre stimulation.

As with heat, it seems that the mere creation of C-fibre stimulation leaves God with an additional task: namely, that of bringing pain into the world. Moreover, it seems that God could have brought pain into the world without creating C-fibres. But ‘pain’ and ‘C-fibre stimulation’ both seem to be rigid designators, so it would seem that the identity theorist cannot coherently maintain that ‘pain is C-fibre stimulation’ is a mere contingent truth (1980: 148–9). A natural strategy would be to defend it as a necessary truth—and explain away the appearance of contingency using one of the strategies just presented. So, for instance, the defender of materialism might say: ‘What you call an intuition to the effect that C-fibre stimulation could exist without pain and vice versa is just a loose and misleading way of saying that C-fibre stimulation could fail to satisfy the description that is used to fix the reference of “pain”.’ Or he might say: ‘What you call an intuition to the effect that C-fibre stimulation could exist without pain and vice versa is just a loose and misleading way of saying that there is some possible epistemic duplicate community (of ourselves or our predecessors) such that “pain” in their mouths does not apply to C-fibre stimulation.’

But, Kripke maintains, these strategies for explaining the appearance of contingency will not work for ‘pain’.⁷² ‘In the case of molecular motion and heat there is something, namely, the sensation of heat, which is an intermediary between the external phenomenon and the observer. In the mental-physical

⁷² Kripke (1980: 151–3); Bealer (Ch. 1 below) argues that Kripke is too quick here.

case no such intermediary is possible' (1980: 151). So, while there is a distinction between the manifest qualities that can be used to fix the reference of a natural kind term like 'heat' and the reference of the natural kind term itself, there seems to be no such distinction available between the quality that is used to fix the referent of 'pain' and pain itself (cf. 1980: 152–3). Likewise in the case of the second speech:

Someone can be in the same epistemic situation as he would be if there were heat, even in the absence of heat, simply by feeling the sensation of heat; and even in the presence of heat, he can have the same evidence as he would have in the absence of heat simply by lacking the sensation *S*. No such possibility exists in the case of pain and other mental phenomena. To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of a pain *is* not to have a pain. (1980: 152, italics in original)⁷³

So, maintains Kripke, neither of the strategies available for explaining modal illusion can be applied in this case: the appearance of possibility is a reflection of genuine possibility. (15) is not merely a posteriori, but contingent.

4 Two-Dimensionalism

Crucial to Kripke's diagnosis of our tendency to modal illusions, is the contrast between the modal status of the proposition expressed by a sentence and modal facts about what that sentence would have expressed in the mouths of other possible communities. This makes for two different perspectives upon any given assertoric utterance. On the one hand, we might wish to consider which possible worlds are in accord with how the sentence *does* represent things to be. On the other hand, we might wish to consider how that sentence would have represented things as being had it been uttered in a different setting (and to consider which possible worlds are in accord with how the sentence *would have* represented things to be). 'Two-dimensional' approaches to modal discourse attempt to accommodate both perspectives, using the machinery of two-dimensional modal logic.⁷⁴

⁷³ Though Kripke does not emphasize the point, the heat/molecular motion strategy has more promise for explaining the appearance that I can exist without my body. Assuming that I am identical to a certain organism, there may be a possible being for which Cartesian dualism is true, whose evidential situation is analogous to mine, and for which an analogous claim of separability is correct.

⁷⁴ For conversations and advice concerning this section, we are grateful to David Chalmers, Ted Sider, and Brian Weatherson. In thinking about these issues, we have been much influenced by Stalnaker (2001).

4.1 *Standard ‘One-Dimensional’ Modal Logic*

There are many ways that the world could have been. If the world had been some of those ways—ways such that snow is white—then what is said by the assertoric utterance ‘Snow is white’ would have been true. If the world had been others of those ways—ways such that snow is not white—then what is said by the utterance would have been false.

Facts such as these are represented within standard modal logic using a familiar formal framework.⁷⁵ Speaking at a maximal level of abstraction, a modal propositional language consists of the symbols of propositional logic plus two monadic sentential operators \Box and \Diamond . The usual semantics for this language consists of a set of indices or ‘points’, a valuation function that assigns to each atomic sentence of the language one of the values 1,0 at each point, and an accessibility relation that specifies which points are accessible from any given point. Complex formulas are then assigned values at points in ways that depend on the values of atomic sentences at various points. The value of a Boolean combination of formulas depends only on the values of those formulas *at that point*, in the usual way. (For example, the formula that results from prefixing a formula ϕ with \sim gets the value 1 at a point if ϕ has the value 0 at that point, 0 if ϕ has the value 1.) But the value of modal formulas $\Box \phi$ and $\Diamond \phi$ at a point depends on the value of ϕ *at other points*. $\Diamond \phi$ has the value 1 at a point i iff ϕ has the value 1 at some point j accessible from i . $\Box \phi$ has the value 1 at i iff ϕ has the value 1 at all points j accessible from i .

For the purposes of shedding light on discourse about possibility and necessity, there is a natural interpretation of this formal framework. We think of \Box and \Diamond as representing necessity and possibility, respectively. We think of the points in the intended interpretation as being an array of possible worlds (with one of the worlds earmarked as the actual world). We think of ‘1’ and ‘0’ as standing for the truth-values true and false, respectively. And we think of the formulas as representing assertoric sentences of some natural language. The valuation function will thus be in the business of assigning truth-values to assertoric sentences of natural language relative to possible worlds.

Suppose that one such model is correct for the set of assertoric utterances of English.⁷⁶ To simplify, let us suppose that the model is one according to which

⁷⁵ How one understands general issues of modality will turn to some extent on fundamental philosophical questions of priority—between worlds and propositions, between sentential truth and propositional truth, between sentential operators and propositional operators. Our discussion here does not explore these foundational issues. For critique of some widely accepted views on these matters, see Bealer, Ch. 1 below.

⁷⁶ Of course, if we are simply interested in determining which modal formulas are valid (given, say, certain constraints on accessibility), we will be concerned with which formulas are true at

every world is accessible from every other (so that from the perspective of any world, every world is a possible world).⁷⁷ Each assertoric utterance will now have, associated with it, a function from possible worlds to truth-values. Associated with the English sentence ‘Snow is white’ is one such function: a function that delivers the value ‘true’ for all worlds where snow is white, and ‘false’ for all other worlds. This function specifies a condition that a world must satisfy in order for this English sentence to be true relative to it: the world must be such that snow is white in that world. Call this function the utterance’s *content*. The content of an utterance will associate two sets of worlds with it: a set of worlds that are the way the utterance represents the world as being, and a set of worlds that are not. Necessary truth and possible truth will be explicable in terms of the \Box and \Diamond operators described earlier. Given the simplifying assumptions about accessibility, an utterance is necessarily true if the second set is empty (that is, if every world is the way the utterance represents the world as being), necessarily false if the first set is empty (that is, if no world is the way the utterance represents the world as being), possibly true if the first set is non-empty (that is, if some world is the way the utterance represents the world as being), and true if the first set contains the actual world (that is, if the actual world is the way the utterance represents the world as being).

In so far as an assertoric utterance is neither necessarily true nor necessarily false, the utterance effects a cut in possibility space. It is natural to suppose that the truth-conditions of the utterance are given by this cut. And it is standard to suppose that the meaning of an utterance is intimately linked to its truth-conditions. Suppose, for example, you ask me what my utterance of ‘pachyderms are macrotous’ means, and I answer by conveying to you that the utterance is true in all and only those worlds where elephants, rhinoceroses, hippopotami, and the like have big ears. It seems reasonable to say that I have done a tolerably good job in answering your question. What we have called the ‘content’ of an assertion is thus closely associated with what in ordinary English would be called its *meaning*. And it has seemed to many philosophers that this might serve as a promising centre-piece for an analysis of the ordinary notion of what is said.⁷⁸

every point in every model (that obeys those accessibility constraints), and not with which formulas come out true on some particular intended interpretation.

⁷⁷ This is the framework of S5 modal logic.

⁷⁸ The most obvious stumbling-block for content-based analyses of meaning is, of course, the account they offer of necessary truths and falsehoods. Intuitively, I have fallen far short of conveying the meaning of a complex mathematical expression if I simply convey to you that the utterance is true/false in all possible worlds. Related problems of coarse-grainedness arise in

4.2 *The Two-Dimensional Framework*

Let's return to our old friend Bob, who, you will recall, was introduced by the reference-fixer 'the number of the planets'. Now suppose that I utter the following sentence:

(16) Bob is odd.

An utterance of (16) is true just in case the number designated by 'Bob' is odd. Since 'Bob' is rigid, the utterance will express a necessary truth if it expresses a truth at all (assuming that any truth of mathematics is necessary).⁷⁹ Reflecting on how the name 'Bob' is introduced, however, it is readily apparent that there are possible worlds where the name 'Bob' is introduced in just the way we actually introduced it, but where the content of the sentence 'Bob is odd' is different. Although the sentence actually expresses a necessary truth, there will be possible tokens of 'Bob is odd' that express necessary falsehoods—for in some worlds, 'Bob' picks out a different (even) number that in that world numbers the planets. So the content associated with 'Bob is odd' depends on contingent facts about the world at which it is uttered. Our competence with various sorts of indicative conditionals—such as 'If, to our great surprise, it turns out that there are in actual fact only four planets, then in actual fact Bob is identical to four, and so Bob is not odd'—seems to turn on an appreciation of something like this dependency.⁸⁰

This discussion makes obvious two kinds of interest that we might have in an utterance. On the one hand, we might be interested in its content. On the other hand, we might be interested in what content that utterance might have had if it had been made in different circumstances. Two-dimensional approaches attempt to accommodate facts corresponding to both kinds of interest within the formal framework of two-dimensional modal logic.

Speaking at a maximal level of abstraction, while one-dimensional modal semantics deploys a valuation function that evaluates formulas relative to a single index, a two-dimensional modal semantics deploys a valuation function that evaluates formulas relative to an ordered pair of indices. Thus the valuation

other cases of extensional equivalence. Addressing such issues would take us beyond the scope of this introduction. Bealer (Ch. 1 below) argues that two-dimensional semantics cannot deal with a number of outstanding problems of 'fine-grained content' relevant to modal epistemology (Frege's puzzle, Mates's puzzle, Kripke's puzzle, etc.), and he outlines an algebraic semantical account designed to handle such phenomena.

⁷⁹ Whether it expresses a truth, of course, will depend on the ultimate classificatory fate of poor Pluto.

⁸⁰ A satisfying account of what such appreciation comes to will require, *inter alia*, some decision as to which of the versions of two-dimensionalism that we describe in sect. 4.3 is best suited to the job.

function will assign one of the values 1,0 to each atomic formula relative to each ordered pair $\langle i,j \rangle$. How a formula gets evaluated now depends upon two dimensions of variation, not one. The truth-value of a Boolean combination at a pair depends on the truth-values of its parts at that same pair, in the usual way. (For example, the formula $\sim \phi$ has the value 1 at a pair $\langle i,j \rangle$ if and only if ϕ has the value 0 at $\langle i,j \rangle$.) The truth-values of the modal formulas $\Box \phi$ and $\Diamond \phi$ at a pair $\langle i,j \rangle$ depend on the truth-values of ϕ at pairs $\langle i,k \rangle$: the modal operators \Box and \Diamond are in this way ‘tied to’ the second index. For example, $\Diamond \phi$ gets the value 1 at $\langle i,j \rangle$ iff ϕ gets the value 1 at some pair $\langle i,k \rangle$ where k is accessible from j . But the language of propositional modal logic may be enriched by other operators as well, operators that are tied to the first index, or even operators that are tied to each index. For example, one might introduce the operator ‘Act’ with a corresponding valuation condition: Act ϕ gets the value 1 at pair $\langle i,j \rangle$ iff ϕ has the value 1 at $\langle i,i \rangle$.

For the purposes of shedding light on discourse about possibility and necessity, there is a natural interpretation of this formal framework. We think of \Box and \Diamond as representing necessity and possibility, respectively. We think of the points in the intended interpretation as being an array of possible worlds (with one of the worlds earmarked as the actual world). We treat ‘1’ and ‘0’ as standing for the truth-values true and false. And we think of the formulas as representing assertoric sentences of some natural language. The valuation function will thus be in the business of assigning truth-values to assertoric sentences of natural language relative to possible worlds.

For the purpose of representing the facts that we are interested in, there is, once again, a natural interpretation of this formal framework. As before, we think of formulas as sentences of a natural language, and the values 1 and 0 as truth-values. A formula ϕ will have a truth-value relative to an ordered pair of indices $\langle i,j \rangle$. We can think of i as representing a possible occasion of use⁸¹ of ϕ . To this end we might usefully think of i as what many philosophers, following Quine, call a ‘centred world’⁸²—roughly, an ordered pair of a possible world and a location in that world. Then ϕ ’s content on that possible occasion of use delivers a truth-value relative to j . We thus think of j as a possible world. \Box and \Diamond , as before, may be thought of as representing the standard notions of necessity and possibility; and ‘Act’ can be thought of as representing the English operator ‘It is actually the case that’.⁸³ To think of ϕ as being true relative to

⁸¹ This oversimplifies things a little: not everyone who uses the two-dimensional framework in this area deploys centred worlds for representing possible occasions of utterance. (For more, see sect. 4.3.)

⁸² Quine 1969.

⁸³ This definition secures that if P is true, then it is necessarily actually the case that P . The best-known discussion of the ‘Actuality’ operator and its relation to the contingent a priori and

$\langle i, j \rangle$, then, is to think of the content that it would have at i delivering the value true at world j .⁸⁴

Suppose that one such model is correct for the assertoric sentences of English. Return now to our ‘Bob is odd’ example. Suppose an index i is a centred world where ‘Bob is odd’ is uttered and the number of planets in the world associated with i is five. Suppose further that the same reference-fixing mechanism is employed for ‘Bob’ as in the actual world. Then, for every pair $\langle i, j \rangle$, the valuation function will assign the value 1 to ‘Bob is odd’. Suppose, meanwhile, an index k is a centred world where ‘Bob is odd’ is uttered and where the number of planets is six. And suppose, as before, that the same reference-fixing mechanism is employed for ‘Bob’ as in the actual world. Then, for every pair $\langle k, j \rangle$, the valuation function will assign the value 0 to ‘Bob is odd’.

For each assertoric sentence, our model will determine what might be called a *two-dimensional* or *2D* function. The arguments of that function will be centred worlds. The values of the function will be contents (which are themselves functions from worlds to truth-values). The variation in contents that the 2D function assigns to a sentence relative to different centred worlds as arguments will represent the way that the content of a sentence can vary in different contexts of use—how its content is determined by contingent facts about the world. Meanwhile, the value of the function given some centred world based on⁸⁵ the actual world—corresponding to some particular deployment of that sentence—will describe the content of some actual assertion. If one is interested in whether that actual assertion is true, or possibly true, or necessarily true (in the sense of ‘possibly’ and ‘necessarily’ that has traditionally interested those concerned with modality), it is the content assigned by the 2D function relative to the centred world corresponding to that speech act that will be of interest. If one is interested in how the truth-value of that sentence might have

necessary a posteriori is Davies and Humberstone (1980). While noting the adequacy of two-dimensional modal logic to explain the behaviour of ‘Actually’, they deploy a slightly different formal framework that relies on considering sets of models that differ only in which world is tagged as actual. Clearly, the notation of two-dimensionalism is not the only way to encode the sorts of philosophical ideas that it is standardly used to express.

⁸⁴ The contrast between considering a world as occupying the first index and considering it as occupying the second corresponds to what Jackson and Chalmers call the difference between considering a world ‘as actual’ and considering it ‘as counterfactual’.

⁸⁵ Each centred world can be mapped on to a world that is the world for which it provides a centre. Assuming a centred world is an ordered pair of world and location, let us say that a centred world is ‘based on’ a world w iff w is the first member of the ordered pair that is that centred world.

been different had the circumstances of utterance been different, then further aspects of the 2D function will be what is of interest.

In some contexts of inquiry, what Stalnaker calls the ‘diagonal proposition’ will be of special interest. This can be thought of as a function from centred worlds to truth-values, where the value true is assigned to a centred world just in case the sentence expresses a truth at that world (that is, at the world upon which the centred world is based).⁸⁶ In effect, the diagonal proposition tells us, for a given sentence, which possible and actual speakers succeed in saying something true by means of it. (The term ‘diagonal’ is used because the function is determined by the values along the diagonal of a two-dimensional matrix representing the sentence’s two-dimensional function.⁸⁷) Return to our ‘Bob is odd’ example. Suppose ‘Bob is odd’ picks out a necessary truth. Still, there is an evident epistemic risk to an utterance of ‘Bob is odd’, one that is highlighted by the diagonal proposition: there will be many possible contexts in which ‘Bob is odd’ expresses a falsehood, and thus many centred worlds for which its diagonal proposition gives the value false. Meanwhile, there are sentences with contingent contents that appear to enjoy a special epistemic security—such as Kripke’s examples of contingent a priori truths. Return to our earlier example, ‘Bob is the number of the planets’. At any world where ‘Bob’ is introduced by the same reference-fixing device as we have specified, and ‘is the number of the planets’ expresses the property that it does at the actual world, the diagonal proposition will deliver the value true.⁸⁸

4.3 *Sub-semantic, Semantic, and Epistemic Two-Dimensionalism*

The discussion so far has deliberately masked a number of issues that divide those who deploy the two-dimensionalist framework. It is time to bring those to the surface, and to explain their significance.⁸⁹

⁸⁶ The diagonal proposition will be determined by how values are assigned to a particular class of ordered pairs: those where the very same world occurs in both places (or as the first member of the pair, supposing that centred worlds are themselves ordered pairs of worlds and locations).

⁸⁷ See Stalnaker (1978). Stalnaker’s discussion there streamlines matters by using worlds rather than centred worlds for the purposes of both indices.

⁸⁸ Can we say that a claim is a priori when its associated diagonal proposition is necessary, and a posteriori when its diagonal proposition is contingent? Such questions cannot be settled prior to deciding which version of two-dimensionalism one is deploying. For example, on the sub-semantic version of two-dimensionalism, such equations are hopeless.

⁸⁹ We are indebted to David Chalmers for a number of the ideas in this section, both taxonomic and substantive. In addition, our discussion in this section has been very much influenced by Stalnaker (2001).

Of special note is the fact that the abstract 2D framework can be combined with a number of different ways of defining the 2D evaluation of statements at worlds. Correlatively, 2D approaches can be conducted with or without a specialized notion of meaning, and with or without an idealized notion of apriority. These variations are of great significance in the current context: for, as we shall see, the extent to which the 2D framework has any prospect of illuminating the connections between apriority and necessity, and between conceivability and possibility, depends very much on the kinds of two-dimensional evaluation that are available.

To begin, the reader may have noticed that the 2D function, as just described, was characterized with regard to some sentence: arguments for which the 2D function delivers values correspond to possible occasions of use of that sentence. But what counts as a use of *that* sentence rather than some other sentence? What criterion of sameness of sentence type is to govern the philosophical interpretation of the framework?

Return to 'Bob is odd'. On one approach, any possible occurrence of a contentful string of sounds that is suitably similar by phonetic (or graphemic) standards (or some other relatively superficial linguistic criterion) will serve as a suitable argument for the two-dimensional function. Some possible use of the sounds 'Bob is odd' express the content that Plato was tall. The two-dimensional function will take such a value as argument and yield as content the function that assigns truth to all worlds in which Plato was tall. (Presumably, the 2D function will be a partial function: given a centred world as argument that does not correspond to a possible use of that string of noises, the 2D function will deliver no value. From the current perspective, its job will be to codify what content that string would have on possible occasions of use. The 2D function for 'Bob is odd' will have nothing to say regarding occasions of use of a different string.) From this perspective, no diagonal proposition will be such as to always deliver the value true for any centred world (where it delivers a value at all), since any phonemically or graphemically individuated sentence type could have been used to express a falsehood. Call this the *sub-semantic 2D function* associated with an utterance.⁹⁰

There are, however, other ways of thinking about the 2D approach. Intuitively, there is a notion of meaning that is orthogonal to the notion of content adumbrated earlier. Begin with David Kaplan's notion of the *character* of an indexical.⁹¹ There is an obvious sense in which you and I mean the same thing when each of us says 'I am hungry'—even though the contents of our utterances are different. We both grasp a rule associated with the lexical

⁹⁰ For exploration of one such approach, see Stalnaker (2001).

⁹¹ See Kaplan (1989).

expression 'I', a rule that generates different contents in different contexts of use. This commonality of rule grounds our sense that the English pronoun 'I' in some sense means the same thing on different occasions of use.

It seems plausible to think that this notion of meaning can be generalized beyond standard indexicals. Consider 'Bob is odd'. Suppose someone in another possible world—call him my 'counterpart'—introduces a name with the same reference-fixing description (which may or may not bear a superficial resemblance to 'Bob'), but rigidly designates a different number thereby. There is an obvious sense in which there is a common rule to my use of 'Bob' and his use of the term he introduces, so that it is not implausible to say that in some sense, my counterpart and I mean the same by our respective terms. Moreover, it is *prima-facie* natural to think that it is meaning in this sense that is more closely tied to the intuitive notion of understanding. Understanding the expression 'Bob' appears to have little to do with possessing a substantive conception of which thing it is that 'Bob' refers to. Indeed, if—due to astronomical ignorance—I mistakenly believe that Bob is identical to the number 6, that is quite consistent with my understanding the term 'Bob' perfectly well. On the face of it, my understanding of the expression is determined by my knowledge of the reference-fixing description that serves as the rule by which a referent is determined and my knowledge that the term so introduced functions rigidly. But this seems to be knowledge that I share with my counterpart.⁹²

While reserving the term 'content' for the notion previously explained (see section 4.1), let us henceforth use the term 'meaning' for the phenomenon just identified. If one finds something like this conception of meaning compelling, one might well find it useful to build some two-dimensional apparatus around it. In particular, the notion of meaning will provide us with a way to type utterances that serve as arguments for the 2D function. Return to 'Bob is odd'. While we have little interest in uses of 'Bob is odd' that are phonetically similar but different in meaning, we may very well be interested in uses that share its meaning but differ in content.

A 2D function will nicely represent how content varies according to the interaction between contingent facts and fixed meanings. In this connection,

⁹² We might allow that the rule is sometimes deferential: In my mouth, the referent of 'quark' might be given by the rule 'Whatever the experts refer to by "quark"', and in that sense, a possible duplicate of me embedded in a community that used 'quark' to pick out a different kind of fundamental particle would mean the same by 'quark' as me in the relevant 'character-theoretic' sense (cf. Chalmers's discussion of Neptune, Ch. 3 below). More generally, the notion of meaning that will maximally illuminate epistemic matters may differ from ordinary notions of linguistic meaning in being far less public.

the function's arguments will be possible utterances with the same meaning, rather than, for instance, possible utterances typed by superficial similarity. Call this the *semantic two-dimensional function*.⁹³

(Perhaps we can generalize even further: for every possible space-time location, we associate a content with the meaning of 'Bob is odd', regardless of whether there is at that location a speaker who makes an utterance with that meaning.⁹⁴ The content will be a function from worlds to truth-values that delivers True at all worlds iff the number of planets at the time associated with the location is odd, and delivers False at all worlds otherwise. Note that it is now no longer clear that the 2D function need be a partial function.⁹⁵)

Not everyone will think that the notion of meaning that is required for the semantic 2D function is coherent. One might in particular worry whether there are disciplined standards for sameness and difference of meaning—as opposed to content—of natural language expressions.⁹⁶ The formal machinery of two-dimensionalism cannot serve to explicate any notion of meaning, since it is presupposed by the very idea of a function from meanings to contents: the concept of meaning is thus used to introduce the notion of a semantic 2D function, not the other way around.⁹⁷ But if there is such a coherent notion of meaning, one might reasonably hope that semantic two-dimensionalism would shed some useful light on our main topics.

In this connection, note that while the diagonal proposition of a sub-semantic 2D function is always contingent, it is natural to suppose that the diagonal proposition of a semantic 2D function is sometimes necessary. Consider 'Bob is the number of the planets' (where 'Bob', as before, is introduced by the reference-fixer 'the number of the planets'). The content is contingent. But

⁹³ Note that our use of 'semantic' here differs from Stalnaker's.

⁹⁴ After all, if someone had actually introduced 'Bob' into English in this way, it wouldn't really have been crucial for anyone to actually utter 'Bob is odd' in order that that sentence get a truth-value relative to the actual world. And in a world where we never even introduce 'Bob' into English (or any other name like it), it seems that we can nevertheless think of that enrichment of English (English + 'Bob') as being available at that world.

⁹⁵ This is the version of semantic two-dimensionalism that comes closest to Chalmers's epistemic version of two-dimensionalism. Note that on this version of semantic two-dimensionalism what is crucial is how meanings (suitably conceived) deliver different contents relative to various hypotheses about which world is actual, not what contents are enjoyed by particular utterances relative to various hypotheses about which world is actual (holding fixed the meaning).

⁹⁶ Cf. in this regard Yablo's discussion (Ch. 13 below) of the connection of these issues to the issue of narrow content.

⁹⁷ It would similarly be absurd to suppose that the machinery of sub-semantic two-dimensionalism is a useful way of explicating what it is for two possible utterances to be phonetic or graphemic duplicates.

any possible utterance that means the same as ‘Bob is the number of the planets’ will, it seems, express a truth on that occasion of use.⁹⁸ Thus the diagonal proposition of the semantic 2D function for ‘Bob is the number of the planets’ will be necessary. While the necessity of the content does not appear sufficient or necessary for apriority, there appears to be a tighter connection between a statement’s apriority and the diagonal proposition of its semantic 2D function.⁹⁹ After all, if its diagonal proposition is contingent, then there will be possible worlds where someone who means the same thing will express something false.¹⁰⁰ Relatedly, where ‘S’ has a contingent diagonal proposition, there would seem to be true indicative conditions of the form ‘If such-and-such a situation is actual, S is false’, where we have no way of telling a priori whether the situation described in the antecedent is actual. So it is hard to see how a statement with a contingent semantic diagonal proposition could be a priori.¹⁰¹ Correlatively, while conceivability may be an insecure guide to whether the content of a statement is possibly true,¹⁰² it may be a much better guide to whether or not the semantic diagonal proposition is contingent. Note in this connection that ‘Hesperus is not Phosphorus’ does enjoy a possibly true diagonal proposition—that is, there is some circumstance where ‘Hesperus is not Phosphorus’ has the same meaning (though not the same content) as it

⁹⁸ We assume for now that the concept of meaning is well enough understood to permit such judgements. Not everyone will agree. Here is one potential source of obscurity: Does sameness of meaning of the term introduced require merely that the reference-fixing descriptions used to introduce the term have the same meaning? Or must these descriptions also have the same content?

⁹⁹ One might even think that the ordinary ‘necessity’ is ambiguous between what we have been calling metaphysical necessity and necessity of diagonal proposition.

¹⁰⁰ There are complications here. We noted above that a semantic 2D function might evaluate formulas at locations where they are not uttered. If we make use of this conception, we may well allow the sentence ‘I exist’ to count as a contingent semantic 2D diagonal proposition. But no possible utterance of ‘I exist’ with the same meaning as its meaning in English will express a falsehood. Call a sentence ‘weakly contingent’ if semantic diagonal proposition assigns it falsity only at centred worlds that are not associated with possible utterances with the same meaning. One now has to classify weak contingency in relation to apriority and conceivability—a task we leave to others.

¹⁰¹ Quite obviously, the contingency of the diagonal proposition associated with a sub-semantic 2D function will not tell against apriority. If the sub-semantic two-dimensionalist is to explain apriority and conceivability, it will have to be by supplemental theory.

¹⁰² Of course, one might stipulate that ‘conceive’ is a success term, so that one can only conceive of the possibility of some content if that content is possible: thus one cannot conceive that Hesperus is not Phosphorus, but only seem to conceive it. But, as before, such a stipulation will restore a tight connection between conceivability and possibility (and may even have some plausibility as a matter of capturing the ordinary language use of ‘conceive’), at the expense of making one’s success at conceiving inscrutable to oneself.

does for us, yet expresses a truth. So someone wishing to defend the view that conceivability is a guide to semantic diagonal possibility¹⁰³ need not be embarrassed to concede that we can conceive that Hesperus is not Phosphorus.¹⁰⁴

In addition to the two just canvassed, a third version of two-dimensionalism has recently been advanced.¹⁰⁵ Begin by noting that the evaluation of such indicative conditionals as

- (17) If the star seen in the morning is not in actual fact the star seen in the evening, then Hesperus is not Phosphorus¹⁰⁶

cannot proceed simply via the contents of the antecedent and consequent. But there does seem to be an epistemic connection between the antecedent and the consequent, and it seems that it is this connection that grounds our confidence in the conditional. Why not, then, build a framework directly around this notion of epistemic connection?¹⁰⁷ *Epistemic two-dimensionalism*¹⁰⁸ does just this. Here—in extremely sketchy form—is the basic picture. Associated with each world w is a world description d . Our two-dimensional modal logic now evaluates claims relative to pairs of world descriptions. We proceed as follows. Take an ordered pair of world descriptions $\langle d_1, d_2 \rangle$ and a statement s . Assume that d_1 describes the actual world. Now determine, on a priori grounds alone, whether on this assumption s is true at the world described by d_2 . For example, assume d_1 is a description that says, *inter alia*, that the lakes and rivers are filled with a transparent, flavourless liquid that is XYZ. Suppose d_2 is a description that says, *inter alia*, that Dave has XYZ in his pool.

¹⁰³ Once again, though, we need to get clearer about what ‘conceivability’ is supposed to be. It is intuitively possible that Bob is not the number of the planets. Yet ‘Bob is the number of the planets’ is diagonally necessary. In so far as conceiving is a guide to diagonal possibility, we need a notion of conceiving according to which our finding it intuitively possible that Bob is not the number of the planets doesn’t count as a case of conceiving. (See Chalmers (Ch. 3 below) for a detailed treatment of such issues, noting especially his distinction between primary and secondary conceivability.)

¹⁰⁴ A number of philosophers have suggested that the necessary a posteriori can be demystified via the 2D framework, although the underlying idea can be expressed without the formalism that the framework provides. The basic idea is that a necessary a posteriori truth of the form $\Box P$ can be factorized into an empirical premiss about the underlying structure of the actual world, together with an a priori conditional whose antecedent is that empirical premiss and whose consequent is $\Box P$. See, e.g., Sidelle, Ch. 8 below.

¹⁰⁵ Cf. Chalmers, Ch. 3 below, and other papers at <http://www.u.arizona.edu/~chalmers/>.

¹⁰⁶ Yablo (Ch. 13 below) argues that it is in fact conditionals of the form ‘If it had turned out that $P \dots$ ’ that best capture the explanandum here. For critique see Chalmers, Ch. 3 below.

¹⁰⁷ One could in principle develop versions of epistemic two-dimensionalism that exploit a notion other than that of a priori connection: relevant here is Yablo, Ch. 13 below.

¹⁰⁸ Thanks to David Chalmers for suggesting this label.

And suppose s is the sentence ‘Dave has water in his pool’. Now determine, on a priori grounds alone, whether Dave has water in his pool at the world described by d_2 on the assumption that d_1 describes the actual world. (Intuitively, our verdict will be ‘yes’: if the lakes and rivers are filled with XYZ in the actual world, then water is XYZ at every world, so at the world described by d_2 , Dave has water in his pool.)

Of special interest will be the epistemic diagonal proposition for a given statement, s .¹⁰⁹ This proposition will be determined, for each value of d , by whether the material conditional ‘ $d \supset s$ ’ is a priori true or a priori false. The a priori evaluation of such conditionals will fix the value of s at all pairs $\langle d_1, d_2 \rangle$ where $d_1 = d_2$. (To inquire a priori whether or not s is true is at a world described by d_1 on the assumption that d_1 is the actual world is just to inquire a priori whether or not $d_1 \supset s$.) The epistemic diagonal will assign to s at a pair $\langle d_1, d_1 \rangle$ the value true or false depending on whether $d_1 \supset s$ is a priori true or a priori false (and will be indeterminate to the extent that there is no a priori decision either way¹¹⁰).¹¹¹

Obviously the details of such a picture will depend a good deal upon what the ‘world descriptions’ look like. A few points to bear in mind:¹¹² (1) To

¹⁰⁹ Chalmers calls the content of a statement its ‘secondary intension’ and the diagonal proposition associated with epistemic two-dimensionalism its ‘primary intension’. Note that what is common among the three sorts of diagonal is purely formal: in each case the diagonal proposition is determined by how the valuation function operates on those ordered pairs of indices $\langle i, j \rangle$ where $i = j$.

¹¹⁰ If the descriptions are framed so that there will always be an a priori decision with regard to any statement evaluated with respect to them, they will be what Chalmers calls ‘epistemically complete’.

¹¹¹ There is no analogously simple procedure for assigning values to the values of the epistemic 2D function for a pair of indices associated with different worlds. One promising way to think about the general case is as follows: The evaluation of some claim s at some index d_2 given that d_1 is actual turns on the a priori evaluation of the more complex conditional:

$$d_1 \supset s \text{ (if } d_2 \text{ were true, } s \text{ would be true).}$$

Where $d_1 = d_2$, this reduces to the a priori evaluation of

$$d_1 \supset s.$$

Given the extra cognitive demands imposed by pairs of indices based upon different worlds, it should not be assumed that our facility with the primary intension associated with an epistemic 2D proposition is cognitively derivative from a facility with the 2D function itself. (Chalmers himself is emphatic that, on his view, primary intension is not to be thought of as the by-product of an epistemic 2D function, and for that reason holds that a primary intension is not fundamentally the diagonal of a 2D function.)

¹¹² Our discussion in this paragraph owes much to conversations with David Chalmers.

the extent that *d* is descriptively thin, the epistemic diagonal proposition for certain claims may yield highly indeterminate results. For example, if the continuum hypothesis is true as a matter of metaphysical necessity, but a priori undecidable, then in so far as the ‘world description’ for each world is silent concerning the truth of the continuum hypothesis, the epistemic diagonal for ‘The continuum hypothesis is true’ will be correspondingly indeterminate. (Note that, intuitively, the semantic diagonal will yield the value True at every world in this case.) (2) To the extent that *d* is descriptively thick, epistemic diagonal possibility will not correspond to anything like conceivability. For example, if the world description for each H₂O world contains the sentence ‘Water is H₂O’, then ‘Water is not H₂O’ will not be an epistemic diagonal possibility.¹¹³ (3) The view requires as an epistemic primitive a highly idealized notion of apriority.¹¹⁴

It is an open question which, if any, versions of a ‘conceivability entails epistemic diagonal possibility’ thesis are true. Take the following toy example: Suppose space, if it exists at all, is necessarily non-Euclidean. Then, relative to any world description *d*, it will presumably be a priori false that ‘If *d*, then a Euclidean space exists’. But it may none the less be conceivable that space is Euclidean. One who wishes to maintain that conceivability entails epistemic diagonal possibility will presumably wish to maintain that this kind of a posteriori necessity cannot arise.¹¹⁵

Sub-semantic two-dimensionalism is, quite obviously, very different from epistemic two-dimensionalism. There appears to be a far greater affinity between semantic and epistemic two-dimensionalism, though we leave an exploration of their relationship to others. We also leave it to readers to explore other possible applications of the formal framework of two-dimensionalism. After all, the versions presented above hardly exhaust the conceptual space. We have in effect indicated a variety of axes along which applications of the

¹¹³ On Chalmers’s view, the water/H₂O problem is avoided by requiring canonical descriptions to be couched in an appropriately neutral vocabulary. The continuum hypothesis problem will not arise if Chalmers’s substantive claims about conceivability and possibility are correct; but if such cases can arise, they are handled by requiring that canonical descriptions are epistemically complete.

¹¹⁴ For example, (i) it applies to infinitary descriptions; (ii) it applies to descriptions involving vocabulary expressing concepts alien to the human mind; (iii) the inference from a given world description to a given statement will often require a good deal more than narrow logical competence, since the vocabulary of the statement may not appear in the world description (as should be clear from the example just given).

¹¹⁵ Chalmers calls purported necessities like these ‘strong necessities’, and argues against their existence (Ch. 3 below).

framework might differ: whether they deploy worlds, centred worlds, or some style or other of world description; whether the evaluation is by truth-value or a priori recognizable truth-value; whether the two-dimensional function can yield values relative to centred worlds where no speech act or thought token occurs; whether the selection of the functions (perhaps partial) from pairs of indices to truth-values is governed by considerations of superficial properties of utterances, or considerations about meaning, or interest in some kind of epistemic connections, and so on. We leave it to readers to consider how these parameters might be juggled, and how yet further interests or parameters might be brought into play.¹¹⁶

4.4 *Two-Dimensionalism and the Mind–Body Problem*

Both the semantic and epistemic versions of two-dimensionalism have prima-facie application to the mind–body problem. By way of example, we offer in what follows a reconstruction of Kripke’s anti-materialism argument through the lens of semantic two-dimensionalism.

We begin by noting that, arguably, there are cases where content cannot vary while meaning is preserved. In such cases, meaning determines content. For example, one might think that any possible person who means what I mean by ‘2 is odd’ will express the same content. Where meaning and content march in step in this way, the content of an utterance will be necessary iff the diagonal proposition is necessary as well. Content and meaning seem to collapse in precisely this way for predicates describing phenomenal experience: Any possible person that means what I do by ‘there is pain’ will (it would seem) express the same content by it. It is this that Kripke’s argument trades on. In the terminology of our semantic two-dimensionalist, while there are possible uses

¹¹⁶ For example, a more austerely Kripkean implementation of the 2D formal framework might associate a partial 2D function with some claim *c* such that pairs of indices for which the function gives a value will have a first index that is either (a) a centred world where someone issues an utterance (or thought token) which expresses the proposition actually expressed by *c* or (b) a world centred upon someone who is an epistemic duplicate—here we have in mind Kripke’s notion of an epistemic counterpart—of someone at the centre of one of the worlds captured by (a). The value delivered by the function will depend upon whether the proposition expressed by the utterance associated with the first index is true at the world that serves as the second index. This version of the framework does not appear to require that we bifurcate notions of meaning. Arguably, something like it can implement a Kripkean account of the sense in which it is conceivable that ‘Hesperus is not Phosphorus’: there is possible being who is (i) an epistemic duplicate of someone who thinks (falsely) that Hesperus is not Phosphorus, and (ii) thinks something true.

of ‘There is heat’ that have the same meaning (via the same reference-fixing mechanisms) but different contents, no such separation of meaning and content can be effected for descriptions of experience.

Let us now re-articulate Kripke’s line of thought within the semantic 2D framework.¹¹⁷ We appear to be able to conceive of the possibility that there is C-fibre stimulation but no pain—a state of affairs incompatible with the identity of C-fibre stimulation and pain. We can concede that, in general, the ability to conceive of ‘a is not b’, where ‘a’ and ‘b’ are rigid, is evidence only for the diagonal possibility of the statement. But in the case of ‘Pain is not C-fibre stimulation’, it would seem, the statement is diagonally possible if and only if it is possibly true (there being no gap between meaning and content).¹¹⁸ If one concedes the diagonal possibility of ‘Pain is not C-fibre stimulation’, it seems, one is forced to relinquish materialism.

The argument is successful only if our conceiving P succeeds in establishing P’s diagonal possibility. How secure is that assumption? An ambitious project in this area is to find some suitable sense of ‘conceiving’ according to which conceiving that P is sufficient for the possible truth of the diagonal proposition of the semantic 2D function—that is, diagonal possibility—for each utterance.¹¹⁹ Even if we grant that the notion of meaning required for semantic two-dimensionalism is coherent, this project faces considerable difficulties. If one is to maintain the thesis that conceiving is an infallible (or near-infallible) guide to diagonal possibility, one needs to deploy a highly idealized notion of conceiving. For example, there is an obvious sense in which we can conceive of the possible truth of necessary falsehoods of mathematics of whose truth-values we are ignorant—but such claims are not, it would seem, diagonally

¹¹⁷ Note that what follows is a semantic two-dimensionalist reconstruction of Kripke, not of more recent extant dualist arguments.

¹¹⁸ Though, as Bealer urges (1994 and Ch. 1 below) there is a worry here that we are forgetting the gap between meaning and content associated with the expression ‘C-fibre stimulation’. One response is to point out that even if there is a gap between meaning and content for ‘Pain is C-fibre stimulation’, there is no such gap for ‘Pain is had by an unextended individual’, and that the diagonal possibility of the latter is enough to make trouble for at least some versions of materialism. (Reconstructing modal arguments this way is an instance of a general strategy—defended by Bealer (1994, 1996, and Ch. 1 below)—according to which ‘semantically unstable’ terms—‘C-fiber’, ‘water’, ‘heat’, etc.—are replaced with associated ‘semantically stable’ terms—‘pain’, ‘individual’, ‘nested functional part’, etc. Bealer argues that, since the resulting sentences are semantically stable, their epistemic possibility entails their metaphysical possibility, therefore yielding the desired outcome in a significant range of cases, but without commitment to any particular controversial semantical scheme.) Further explorations of these issues are offered below by Chalmers, Stalnaker, Yablo, and Wright.

¹¹⁹ An even more ambitious project is to find some sense of ‘conceivable’ according to which being conceivable is necessary for diagonal possibility.

possible: any possible mathematical utterance with that meaning is false.¹²⁰ For the thesis to be maintained, then, there must be a demanding conception of conceiving—which, to be interesting, must not be tied to diagonal possibility by mere stipulation—according to which what we do in these mathematical cases will not count as conceiving.¹²¹ Call this stronger notion of conceiving—conceiving that entails diagonal possibility—*superconceiving*. If superconceiving were (a) internally scrutable, and (as defined) (b) such as to guarantee diagonal possibility, then one could, potentially, justify anti-materialism in the philosophy of mind using the following style of argument:

- (1) A statement is superconceivable iff it is diagonally possible.
- (2) ‘Pain is not C-fibre stimulation’ is superconceivable.
- (3) ‘Pain is not C-fibre stimulation’ is diagonally possible iff it is possible
- (4) Therefore ‘Pain is not C-fibre stimulation’ is possibly true.
- (5) Where a claim of distinctness flanked by two rigid designators is possibly true, it is necessarily true.
- (6) Therefore ‘Pain is not C-fibre stimulation’ is actually true.

Can two-dimensional reflection ultimately serve to deliver on the Cartesian dream of refuting materialism in the philosophy of mind through a priori reflection? On this matter, an introduction like this should remain silent.

5 Summaries of Papers

In this section, we provide brief summaries of each of the papers in the volume, so as to permit the reader to identify those most likely to be of interest to her.

5.1 *George Bealer*

In ‘Modal Epistemology and the Rationalist Renaissance’, George Bealer defends a view he calls *moderate rationalism*: that for a certain distinguished class of propositions—those that are *semantically stable*¹²²—necessity and a priori knowability coincide. Because many recent discussions of modal epistemology have misidentified or mischaracterized their subject-matter, he suggests, the existence of such an equivalence (with proper scope) has often been challenged or supported for philosophically unsound reasons. By way of corrective, Bealer

¹²⁰ But see Rosen, Ch. 7 below.

¹²¹ Chalmers (Ch. 3 below) is helpful here; see also the chapters by Bealer, Wright, and Yablo.

¹²² A proposition is semantically stable if it is invariant across communities whose epistemic situations are qualitatively identical.

offers positive characterizations of apriority/aposteriority, possibility/necessity, the nature of propositions, and the nature of understanding that together serve to ground an alternative picture.

Many discussions of the epistemology of modality, he suggests, founder at the outset through their failure to recognize that the source of all (non-stipulative) a priori knowledge is the *sui generis* propositional attitude of *rational intuition* (and not, as many people suppose, conceivability). Recognizing this makes possible a positive characterization of a priori knowledge: knowledge that is directly intuitive, or stipulative, or based wholly upon intuition and/or stipulation.

Apriority in this sense is, Bealer claims, a guide to metaphysical possibility and necessity. In contrast to the latter notions, epistemic possibility and necessity concern a different phenomenon: namely, the quality and character of our evidential relation to *p*. Proper attention to the details of these distinctions, Bealer suggests, permits proper appreciation of Kripke's contribution to modal epistemology, and points towards ways in which two-dimensionalist accounts are doomed to inadequacy.

Such accounts misfire, Bealer suggests, in failing to recognize that notions of possibility should be accommodated not by distinguishing different kinds of meaning (primary, secondary, etc.), but by distinguishing different kinds of modal operators that can be applied to a fixed domain of propositions. Because they do not, he argues, two-dimensionalist approaches to epistemic possibility, meaning, and understanding such as those proposed by Frank Jackson and David Chalmers fall short in a number of ways: they run a foul of Frege's puzzle, posit ambiguities where there are none, wrongly reckon only sentences (never propositions) to enjoy the property of being necessary a posteriori, and overestimate the centrality of reference-fixing descriptions to linguistic understanding.

As a corrective, Bealer develops a modal epistemology that takes propositions as more basic than possible worlds and that attempts to ground the reliability of intuition by locating it within an account of what it is to understand a concept or proposition. With a sketch of this theory in place, Bealer presents a taxonomy of sources of modal error, and applies his analysis to recent versions of anti-materialist arguments in the philosophy of mind.

5.2 *John Campbell*

In 'Berkeley's Puzzle', John Campbell offers an interpretation of Berkeley's famous argument that it is impossible to conceive of existence unperceived, suggesting that the argument is best understood as making appeal to a principle

that Campbell calls *the explanatory role of experience*: that concepts of physical objects, and of their observable characteristics, are made available by our experience of the world. The puzzle that confronts Berkeley, Campbell suggests, is that it is difficult to see how experience could provide us with such concepts if the objects it concerns are mind-independent. Hence, concludes Berkeley, we do not have concepts of mind-independent objects.

Campbell accepts that the puzzle raises a challenge, but rejects Berkeley's solution. Instead, Campbell contends, we need to take what he calls a *relational view* of experience, according to which experience is a primitive cognitive relation that explains, rather than depends upon, the possibility of thoughts about objects. The view is *disjunctivist* in holding that there is no experiential factor in common between a case where one sees an object and a case where one has a hallucination as of such an object. On such a view, experience of objects *can* provide us with concepts of mind-independent objects in ways that mere imagination cannot, for experience has mind-independent objects as its constituent ingredients.

5.3 David Chalmers

In 'Does Conceivability Entail Possibility?', David Chalmers presents a detailed taxonomy for exploring the relationship between conceivability and possibility, examines and defends the most promising versions of the thesis that conceivability entails possibility, and applies these results to the question of mind-body dualism.

Chalmers's taxonomy of types of conceivability employs three (independent) distinctions: between (a) finding oneself *prima facie* able to conceive of P and (b) being able to conceive of P on ideal rational reflection (*prima-facie* versus *ideal* conceivability); between (c) being unable to rule out P and (d) being able to form a detailed positive conception of a situation in which P (*negative* versus *positive* conceivability); and between (e) its being conceivable that P is actually the case and (f) its being conceivable that P might have been the case (*primary/epistemic* versus *secondary/subjunctive* conceivability). This final distinction is paired with an associated distinction between two corresponding notions of possibility: P is *primarily* possible iff there is a world *w* such that if *w* is actual, P is the case; P is *secondarily* possible iff there is a world *w* such that if *w* had been actual, P would have been the case. (So, for example, it is *primarily*, though not *secondarily*, possible that Hesperus is not Phosphorus.)

Of these, it is *primary* conceivability that is of greatest interest to modal rationalism, *primary* possibility to which *primary* conceivability offers the most promising guide, and *ideal* conceivability that provides a credible

candidate for reliable guidance. The conceivability–possibility theses of principal interest, then, are (1) that ideal primary positive conceivability entails primary possibility, and (2) that ideal primary negative conceivability entails primary possibility. If (1) and (2) are true (as Chalmers believes they are), then instances of modal error will be traceable to cases where (non-possibility entailing) prima-facie conceivability is (mistakenly) taken as a guide to possibility, and cases where (non-secondary possibility entailing) primary conceivability is (mistakenly) taken as a guide to secondary possibility.

Chalmers points out that for (1) to be true but (2) to be false, there would need to be statements that fall in what he calls the *twilight zone*: statements that are negatively but not positively conceivable. Such statements, he suggests, would need to be either *inscrutable*, in the sense that they are not epistemically accessible on the basis of a complete qualitative description of the world, or *openly inconceivable*, in the sense that they cannot be ruled out a priori, but are verified by no positively conceivable situation. Chalmers suggests that both of these classes are empty, so that if (1) is true, (2) is true as well. But for (1) to be false, there would have to be instances of *strong necessities*—statements true in all possible worlds that are falsified by some positively conceivable situation considered as actual. But, Chalmers argues, there are no such instances, so (1) (and hence (2)) is true.

Chalmers then applies this result to the mind–body problem, offering an argument against materialism that moves from the premiss that zombies are ideally primarily positively (or negatively) conceivable to the conclusion that they are primarily possible, and thence to the conclusion that materialism (of the standard sort) is false.

5.4 *Gregory Currie*

In ‘Desire in Imagination’, Gregory Currie contends that we need to distinguish between two sorts of imagining, belief-like imagining, on the one hand, and desire-like imagining, on the other. In imagining of the first sort, one imaginatively projects into the situation of one who believes P; in imagining of the second sort, one imaginatively projects into the situation of one who desires Q. Recognizing these two categories, contends Currie, allows one to account for a phenomenon—first identified by Hume—that has come to be known as the puzzle of *imaginative resistance*: whereas we have no systematic difficulty imagining non-moral facts to be otherwise than they are—even to the point of imagining straightforward impossibilities—we tend to find ourselves unwilling or unable to imagine alien moral facts.

The explanation for this, on Currie’s account, is that imaginative resistance arises primarily in cases of desire-like imagining, and only derivatively in cases

of belief-like imagining. For, whereas belief-like imagining is carried out in a largely ‘off-line’ fashion, largely unconstrained by the beliefs that one holds about the actual world, desire-like imagining, according to Currie, engages the imaginer’s moral character; as a result, desire-like imagining is constrained by factors that belief-like imagining is not, and hence is more restricted in its range.

5.5 *Michael Della Rocca*

In ‘Essentialism versus Essentialism’, Michael Della Rocca challenges an assumption generally taken as common ground in contemporary debates about the relation between conceivability and possibility: namely, the assumption that if \underline{a} and \underline{b} differ in their possible features, then \underline{a} and \underline{b} are non-identical. As Della Rocca observes, Cartesian-style conceivability–possibility arguments for the dualism of phenomenal and physical properties (versions of which are advanced by Kripke and Chalmers) proceed in two steps: (i) from the conceivability of \underline{a} ’s having some feature F and the inconceivability of \underline{b} ’s having that feature F , it is concluded that \underline{a} and \underline{b} differ modally (that they have different possible features); (ii) from the fact that \underline{a} and \underline{b} differ modally, it is concluded that \underline{a} and \underline{b} are non-identical. In challenging the second step of such arguments, Della Rocca aims to call into question the metaphysical significance of the conceivability–possibility debate; for if modal difference does not imply distinctness, then even granting the conceivability–possibility move will not allow proponents of Cartesian-style arguments to establish non-identity claims on conceivability grounds.

Della Rocca suggests that the move from modal difference to non-identity presupposes essentialism, which he glosses as the two-part thesis that (i) objects have their modal properties independently of how they are described, and (ii) have certain properties essentially. Working with largely Kripkean materials, he attempts to raise doubts concerning the first of these theses.

It is characteristic of essentialists, he contends, to suppose that the modal properties of an individual are not to be explained in terms of facts of similarity between that individual and various possibilia. By contrast, anti-essentialists—and especially ‘counterpart theorists’—frequently make use of facts of similarity to account for the truth and falsity of modal property ascriptions (and for the description dependence of such ascriptions). Della Rocca is struck by the fact that there are certain contexts in which Kripke himself is more than willing to deploy facts of similarity in order to make sense of modal intuitions. For example, in explaining away intuitions of contingency accruing to ‘Hesperus is Phosphorus’, Kripke invokes epistemically similar scenarios where

there are two objects in the sky. If facts of similarity are to be invoked there, why should they not be invoked more generally to account for modal property ascriptions (as many anti-essentialists would recommend)? Della Rocca suggests that the essentialist has no good answer to this query.

5.6 *Kit Fine*

In 'The Varieties of Necessity', Kit Fine argues that there are three main forms of necessity—the metaphysical, the natural, and the normative—none reducible to either of the others, or to any other form of necessity. Metaphysical necessity is logical necessity 'in the broad sense'—necessity that obtains in virtue of the identity of things, broadly conceived. So, for instance, the truths of logic are metaphysically necessary (it is metaphysically necessary that anything red is red), as are mathematical truths (there are prime numbers less than 10), conceptual truths (nothing red is green), and truths of identity broadly understood (2 is a number). Natural necessity is the form of necessity standardly associated with laws of nature. So, for instance, certain propositions describing the motions of objects or kinds of objects express natural necessities (it is naturally necessary that billiard-ball B moves in thus-and-such manner when impacted by billiard-ball A; it is naturally necessary that bodies attract one another according to an inverse square law). Normative necessity is the form of necessity that pertains to moral claims—and perhaps to normative claims more generally. So, for instance, certain propositions expressing judgements of rightness and wrongness express normative necessities (it may be normatively necessary that war or murder or lying is wrong, or that if I make a promise I am obliged to keep it).

After providing reasons for thinking that metaphysical necessity (as opposed to some more restricted notion such as mathematical necessity or logical necessity 'in the narrow sense') is properly taken to be a primitive notion of necessity, Fine devotes the bulk of his discussion to showing that neither natural nor normative necessity can be assimilated—either by subsumption or by definition—to metaphysical necessity. Natural necessity cannot be subsumed under metaphysical necessity, for the laws that govern the natural world cannot be fully explained by essentialist truths arising from the identity of things. As a result, attempts to subsume the natural under metaphysical will mislocate the source of the necessity. Nor can natural necessity be defined in terms of metaphysical necessity by characterizing it as a form of metaphysical necessity *relative to* some set of assumptions about the natural world. For such a strategy renders the necessity of the propositions with respect to which the necessity is relative a trivial matter, whereas we are aiming for an account of

their necessity that does not trivialize it. Parallel arguments are offered for the case of normative necessity: it cannot be subsumed under metaphysical (or conceptual) necessity on pain of mislocating the source of its normative force, and it cannot be defined in terms of metaphysical necessity on pain of trivializing its content.

Fine closes by considering whether all three forms of necessity might be defined as restrictions of some more comprehensive form of necessity, but rejects this on the grounds that such a characterization would fail to capture the sense in which these three forms of necessity seem, at base, to describe three fundamentally different sorts of constraints upon the world.

5.7 *Gideon Rosen*

In 'A Study in Modal Deviance', Gideon Rosen describes an imaginary tribe—the Q—who hold that the existential truths of number theory—and with them, the related truths of arithmetic—are contingent: according to the Q, there might have been no numbers, in which case it would be false, for instance, that $7 + 5 = 12$, or that there are 10 prime numbers less than 30.

Rosen provides the Q with three arguments in favour of their position: an argument from 'vivid conceivability', a 'Humean argument', and an argument from 'strong coherence'. According to the first, we can vividly conceive a world in which there are no numbers, and hence, barring some special reason for thinking the conception misleading, we are entitled to conclude that such a world is possible. The recipe for such conception goes as follows: start by conceiving a world in which standard arithmetic is true—a world containing both concreta and an infinite number of abstracta. Now eliminate the abstracta, and you have conceived a world in which there are no numbers. The second argument employs the ingredients of the first, without direct appeal to conceiving. Assuming with Hume that there are no necessary connections between distinct existences, it follows that there is a world of concreta without associated abstracta—and hence a world in which arithmetic is false. The third argument makes appeal to the principle that any hypothesis that is 'strongly coherent' is possibly true, where a strongly coherent hypothesis is any hypothesis whose truth in some possible world is consistent with all non-modal facts about the actual world. The hypothesis that there are no numbers seems strongly coherent in this sense, and hence, according to the principle in question, possibly true.

Rosen goes on to describe another imaginary tribe—the Z—who defend the necessary truth of sentences couched in their arithmetical vocabulary on the basis of a 'modal structuralist' construal of such claims. Because of

the ontological innocence of arithmetical claims in their mouths, the Z are untroubled by the arguments of the Q. Whether we can follow suit in this regard depends crucially, Rosen suggests, on subtle issues concerning the meaning of ‘exists’.

5.8 *Alan Sidelle*

In ‘On the Metaphysical Contingency of Laws of Nature’, Alan Sidelle aims to defend the traditional role of imagining and conceiving as a means to modal knowledge: while there are, he concedes, necessary a posteriori truths, their existence should not undermine our general reliance on imagination and conception as guides to possibility.

Sidelle makes his case by focusing on the particular example of laws of nature, which some philosophers have recently maintained to be—‘in the strongest sense’—necessary. Against this, Sidelle tries to show that the laws of nature are contingent—or, if necessary, necessary in metaphysically uninteresting ways.

Sidelle begins by enumerating a number of prima-facie reasons for thinking the laws of nature to be contingent: namely, that it seems fully imaginable that they might have been otherwise. Kripke- and Putnam-style arguments cannot, Sidelle contends, provide a metaphysically interesting challenge to such imaginative exercises, for necessary a posteriori truths are the result of filling ‘gaps’ in analytic principles of individuation with particular empirical findings. For example, water is the substance that has the same deep explanatory feature as the stuff we call ‘water’, and that deep explanatory feature is: being H_2O . As a result, to say that such facts are necessary—that water is necessarily H_2O , or that the laws of nature are necessarily thus-and-such—turns out not to describe an interesting constraint on ways it is possible for the world to have been, but rather to describe an uninteresting constraint on ways it is possible for us to *describe* the world as having been.

Sidelle considers in detail two main arguments in favour of the necessity of laws of nature. According to the first, it is only by taking laws of nature to be metaphysically necessary that we can explain their modal force and their capacity to support counterfactuals; according to the second, it is only by taking laws of nature to be metaphysically necessary that we can understand them as governing the properties that they do. Neither succeeds in motivating metaphysically robust claims of necessity for laws of nature, he argues. Generalizing these results, Sidelle concludes that we have categorical grounds for rejecting challenges to the overall reliability of conceiving and imagining as guides to possibility.

5.9 Roy Sorensen

In ‘The Art of the Impossible’, Roy Sorensen explores the question of what would be required for the visual depiction of a logical impossibility. (If such impossibilities are visually representable, then presumably they are imaginable; if they are not visually representable, then their imaginability may be more difficult to maintain.)

Biological, physical, and conceptual impossibilities, Sorensen contends, can be straightforwardly depicted visually: mermaids, floating rocks, and Escher staircases provide examples of each in turn. In exploring the question of whether one might similarly depict a logical impossibility, Sorensen articulates a set of standards for what such a depiction would require. In particular, the depiction must be (a) open to inspection, in the sense that it places no limit on potential detail concerning the item represented; (b) equivocation-free, in the sense that contradiction depicted does not exploit an ambiguity in language or representational convention; (c) perceptual, in the sense that the impossibility is not a consequence of how the depiction is labelled or described; (d) not merely adverbially inconsistent, in the sense that the impossibility is not the result of treating inaccuracies as inconsistencies; (e) not merely inconsistent at the level of infrastructure, in the sense that the impossibility is not merely the result of the interplay between two distinct perceptual capacities; and (f) not merely ambiguous, in the sense that the impossibility is not a consequence of treating the depiction as simultaneously characterized by two of its ambiguous readings.

Though Sorensen does not present an example of a visual depiction of a logical impossibility—indeed, he offers a reward of \$100 to the first reader who provides him with one—he is optimistic about the prospects of there being such. After all, he insists, logical impossibilities can be depicted narratively. So, unless there is an in principle consistency constraint governing perceptual representation that does not govern its linguistic and conceptual cousins, then, presumably, logical inconsistencies can be depicted visually as well.

5.10 Ernest Sosa

In ‘Reliability and the A Priori’, Ernest Sosa considers various responses to what he terms *the Platonist’s dilemma*—the problem of explaining our reliable knowledge of mathematical truths, assuming that these truths concern a realm of abstracta that lies beyond space-time. The dilemma is this: according to the Platonist, (i) mathematical objects are acausal and mind-independent. But, it seems that (ii) causal explanations of our mathematical reliability are

incompatible with the acausality of mathematical objects, and that (iii) non-causal explanations of our mathematical reliability are incompatible with their mind-independence. So, given Platonist commitments, it seems that (iv) there is no explanation of our mathematical reliability.

Sosa considers two families of response to this dilemma. The first sort of response involves rejecting (iii) (and in its extreme form, the second clause of (i)) on the grounds that mathematical truths are *judgement-dependent*, in the sense that the truths about mathematical objects are determined by our best judgements about them. Sosa presents various refinements of this view, but ultimately rejects them because he thinks they cannot account for the purported necessity of mathematical truths. The second sort of response rejects (iii) by exploiting an analogy between thoughts about mathematics and *cogito*-like thoughts. Sosa suggests that where the content-determining conditions of a thought imply that to have such a thought is to have a thought that is true, we may have a non-causal explanation of the reliability of that thought without abandoning our commitment to the mind-independence of its content. Mathematical facts may be of this sort: in the case of mathematical truths, understanding and knowledge go hand in hand. While conceding that such responses leave certain sorts of questions unanswered, Sosa is sanguine about the general strategy. We can, it seems, offer a fairly satisfactory explanation of our knowledge of certain necessary truths.

5.11 *Robert Stalnaker*

In ‘What is it Like to be a Zombie?’, Robert Stalnaker seeks to throw light on general questions about the nature of modal claims and the relations among metaphysical, semantic, and empirical questions by examining the debate between those who think zombies are (metaphysically) possible and those who think they are not. (Zombies are creatures that physically duplicate ordinary people, but lack phenomenal consciousness.) Stalnaker frames the debate as a conversation among three imaginary philosophers with real-world counterparts (Dave, Patricia, and Sydney) and a fourth (Anne) who steps in at the end for clarification. Dave (whose real-world counterpart is David Chalmers) holds that zombies are possible but non-actual; Patricia (whose real-world counterpart is Patricia Churchland) holds that zombies are possible and, indeed, actual; and Sydney (whose real-world counterpart is Sydney Shoemaker) holds that zombies are neither actual nor possible. Anne (who lacks a particular real-world counterpart) holds that *if* materialism is false (as Dave holds), then zombies are possible though non-actual, but *if* materialism is true (as Sydney holds), then they are neither actual nor possible.

Patricia, Sydney, and Anne agree on the empirical question of whether materialism is true; Dave demurs. Sydney, Anne, and Dave agree on the empirical question of whether there is (actually) phenomenal consciousness; Patricia demurs. Sydney holds that phenomenal consciousness is a functional property that would be (and is) present in what Stalnaker calls a *z-world*: a world exactly like the actual world in all physical respects, containing nothing that does not supervene on the physical. Moreover, he believes it is built into the meanings of experiential vocabulary that its terms denote functional properties. Dave, by contrast, holds that phenomenal consciousness is an irreducible non-physical property that would be absent in a *z-world*; the world *we* live in, he maintains, is what Stalnaker calls an *a-world*: a world physically just like a *z-world* that has, in addition, certain properties not instantiated in the *z-world*—in particular, properties of conscious beings that are the properties we refer to when we talk about phenomenal consciousness. So, while Dave and Sydney agree that *z-worlds* are both conceivable and metaphysically possible, Dave holds that *z-worlds* are zombie worlds, whereas Sydney holds that they are not.

Anne regards experiential concepts as less theoretically loaded than either Dave or Sydney. According to Anne, if the actual world is a *z-world*, then phenomenal consciousness is a functional property of the kind Sydney takes it to be; if the actual world is an *a-world*, then phenomenal consciousness is an irreducible non-physical property of the kind Dave takes it to be. But this means that such concepts cannot carry with them either an a priori requirement that they refer to irreducible non-physical properties (as Dave requires) or an a priori requirement that they refer to functional properties (as Sydney requires). For Anne, the prima-facie conceivability of zombies establishes their metaphysical possibility only if we know on independent grounds that our world is not, in fact, a *z-world*.

5.12 Crispin Wright

In ‘The Conceivability of Naturalism’, Crispin Wright explores Kripke’s well-known argument concerning materialism, generalizes it, and then raises questions concerning its adequacy. He begins by pointing out that Kripke’s argument depends on what Wright calls the *Counter-Conceivability Principle*, that if we can clearly and distinctly conceive of a scenario in which not-*P*, then it is not necessary that *P*. An apparently lucid conception of a scenario in which *a* and *b* are distinct is thus sufficient to establish the non-identity of *a* and *b*—except where the conception can be discounted as misleading. Wright notes that Kripke’s method for identifying such deficiencies relies on a particular insight: an apparently lucid conception of a situation where some kind

(or individual) is F may latch on to nothing more than a possible situation where a symptomatic counterpart of that kind (or individual) is F. Central to Wright's presentation of Kripke's argument, then, is the observation that there is no distinction between a symptomatic counterpart of pain and pain itself. Thus the general strategy for blocking deployment of the Counter-Conceivability Principle cannot be invoked in this case.

Wright defends Kripke's argument against objections by Boyd (to the effect that Kripke's argument neglects the relevance of the distinction between C-fibre stimulation and a symptomatic counterpart of it) and McGinn (to the effect that a token-identity version of physicalism remains intact). He then generalizes the argument, suggesting that analogous anti-materialist arguments can be run for any concept that is what he calls *Euthyphronic*—such that it is a priori that (when certain specified conditions hold) if it seems to a thinker that the concept is instantiated, then it is.

None the less, Wright maintains, the generalized Kripke argument is unsuccessful. For, while Kripke offers one recipe for blocking conceivability-possibility arguments, he is insensitive to other ways to defuse such reasoning. In particular, Wright contends, there are cases where an apparently lucid conception that purports to be of a *possible* scenario in which not-P is in fact a conception of what it would be like if—*per impossibile*—P were (found to be) false. In such cases, he argues, the Counter-Conceivability Principle cannot be employed to establish claims of non-identity.

5.13 *Stephen Yablo*

In 'Coulda, Woulda, Shoulda', Stephen Yablo offers a critique of modal rationalism through careful examination of a notion he calls *conceptual possibility*, which can be characterized in contrast to its more familiar metaphysical counterpart: it is metaphysically, but not conceptually, possible that the metre stick exist without being a metre long, conceptually, but not metaphysically, possible that Hesperus exist without Phosphorus. In general, suggests Yablo, we can express the difference between the two notions as follows: S is metaphysically possible iff it could have been that S (iff some world *w* is such that it would have been that S, had *w* obtained), and conceptually possible iff it could have turned out that S (iff some world *w* is such that it would have turned out that S, had *w* turned out to be actual). To consider a world as metaphysically possible is to consider it as *counterfactual*; to consider a world as conceptually possible, he suggests, is to consider it as *counteractual*.

Using this terminology, modal rationalism can be characterized as the thesis that (although metaphysical necessity and apriority may come apart in certain

cases) conceptual necessity and apriority coincide: P is knowable a priori iff P holds in all counterfactual worlds. But, contends Yablo, the thesis fails in both directions: there are a priori knowable statements that are false at some counterfactual worlds, and statements true at all counterfactual worlds that are knowable only a posteriori. This can be seen, for example, by considering how we come to know the truth or falsity of conditionals of the form: 'If w had turned out to be actual, it would have turned out that Q', where w is some detailed world description, and Q is some candidate conceptual possibility. For modal rationalism to be correct, we would need in all such cases to be able to deduce (or recognize that one cannot deduce) Q from w by a priori methods. But, argues Yablo, in a wide range of cases, we cannot. In particular, we cannot do so for Qs that involve predicates that are *recognitional* (like 'painful'), otherwise *observational* (like 'jagged'), *evaluative* (like 'wrong'), or *theoretical* (like 'energy'). In these cases, he argues, we can determine whether Q given w only by means of engaging in a sort of 'off-line' simulation that allows us to exercise our (perceptual or non-perceptual) sensibilities—and judgements formed on such a basis do not provide us with knowledge that is a priori.

For a statement S to be known a priori by me, Yablo suggests, is for me to possess some information G such that (a) I grasp the meaning of 'S' in part by knowing that 'S' is G, and (b) that 'S' is G conceptually necessitates that 'S' is true. Since there may be statements that are conceptually necessary for which there is no such G possessable by me, there will be conceptually necessary truths that I cannot know a priori. And since the information G that conceptually necessitates that 'S' is true may itself not be conceptually necessary, there will be a priori truths that are conceptually contingent.

REFERENCES

- Alanen, Lilli, and Knuuttila, Simo (1988), 'The Foundations of Modality and Conceivability in Descartes and his Predecessors', in Simo Knuuttila (ed.), *Modern Modalities* (Dordrecht: Kluwer), 1–69.
- Balog, Katalin (1999), 'Conceivability, Possibility and the Mind–Body Problem', *Philosophical Review*, 108(4): 497–528.
- Bealer, George (1987), 'Philosophical Limits of Scientific Essentialism', *Philosophical Perspectives*, 1: 289–365.
- (1992), 'The Incoherence of Empiricism', *Aristotelian Society*, Supp. Vol. 66: 99–138.
- (1994), 'Mental Properties', *Journal of Philosophy*, 91: 185–208.
- (1996), 'A Priori Knowledge and the Scope of Philosophy', *Philosophical Studies*, 81(2–3): 121–42.

- Block, Ned, (1981) (ed.), *Imagery* (Cambridge, Mass.: MIT Press).
- Block, Ned and Stalnaker, Robert (1999), 'Conceptual Analysis, Dualism, and the Explanatory Gap', *Philosophical Review*, 108(1): 1–46.
- Brann, Eva (1991), *The World of the Imagination: Sum and Substance* (Lanham, Md.: Rowman and Littlefield).
- Casey, Edward S. (1976/2000), *Imagining: A Phenomenological Study*, 2nd edn. (Bloomington, Ind.: Indiana University Press).
- Chalmers, David (1996), *The Conscious Mind* (New York: Oxford University Press).
- (1999), 'Materialism and the Metaphysics of Modality', *Philosophy and Phenomenological Research*, 59(2): 473–96.
- and Jackson, Frank (2001), 'Conceptual Analysis and Reductive Explanation', *Philosophical Review*, 110: 315–61.
- Currie, Gregory, and Ravenscroft, Ian (2002), *Recreative Minds: Image and Imagination in Philosophy and Psychology* (Oxford: Oxford University Press).
- Davies, Martin, and Humberstone, Lloyd (1980), 'Two Notions of Necessity', *Philosophical Studies*, 38: 1–30.
- and Stone, Tony (1995a) (eds.), *Folk Psychology: The Theory of Mind Debate* (Oxford: Blackwell).
- — (1995b) (eds.), *Mental Simulation: Evaluations and Applications* (Oxford: Blackwell).
- DePaul, Michael, and Ramsey, William (1998) (eds.), *Rethinking Intuition* (Lanham, Md.: Rowman and Littlefield).
- Descartes, René ([1619–64] 1990), *The Philosophical Writings of Descartes*, vol. I, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press).
- ([1641–2, 1701] 1989), *The Philosophical Writings of Descartes*, vol. II, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press).
- ([1619–50] 1991), *The Philosophical Writings of Descartes*, vol. III, trans. John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny (Cambridge: Cambridge University Press).
- Gendler, Tamar Szabó (2000), *Thought Experiment: On the Powers and Limits of Imaginary Cases* (New York: Garland Routledge).
- Harris, Paul (2000), *The Work of the Imagination* (Oxford: Blackwell).
- Hart, W. D. (1988), *Engines of the Soul* (Cambridge: Cambridge University Press).
- Hill, Christopher (1997), 'Imaginability, Conceivability, Possibility and the Mind–Body Problem', *Philosophical Studies*, 87(1): 61–85.
- Horowitz, Tamara, and Massey, Gerald (1991) (eds.), *Thought Experiments in Science and Philosophy* (Savage, Md.: Rowman and Littlefield).
- Hume, David ([1739–40] 2000), *A Treatise of Human Nature*, ed. David Fate Norton and Mary Norton (Oxford: Oxford University Press).
- Jackson, Frank (1993), 'Armchair Metaphysics', in Michaelis Michael and John O'Leary-Hawthorne (eds.), *Philosophy in Mind* (Dordrecht: Kluwer), 23–42.

- (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (New York: Oxford University Press).
- Kaplan, David (1989), 'Demonstratives', in Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan* (New York: Oxford University Press), 481–564.
- Kripke, Saul (1980), *Naming and Necessity* (Cambridge, Mass.: Harvard University Press).
- Levine, Joseph (1998), 'Conceivability and the Metaphysics of Mind', *Nous*, 32(4): 449–80.
- (2001), *Purple Haze: The Puzzle of Consciousness* (New York: Oxford University Press).
- Lewis, Charles, and Mitchell, Peter (1994) (eds.), *Children's Early Understanding of Mind* (Hillsdale, NJ: Laurence Erlbaum).
- Lewis, David (1986), *On the Plurality of Worlds* (Oxford: Blackwell).
- Loar, Brian (1999), 'David Chalmers's "The Conscious Mind"', *Philosophy and Phenomenological Research*, 59(2): 465–72.
- McLaughlin, Brian, and Hill, Christopher (1999), 'There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy', *Philosophy and Phenomenological Research*, 59(2): 445–54.
- O'Shaughnessy, Brian (2000), *Consciousness and the World* (Oxford: Clarendon Press).
- Overton, W. F. (1990) (ed.), *Reasoning, Necessity and Logic: Developmental Perspectives* (Hillsdale, NJ: Laurence Erlbaum).
- Piaget, Jean (1987a), *Possibility and Necessity: The Role of Necessity in Cognitive Development*, trans. Helga Feider (Minneapolis: University of Minnesota Press).
- (1987b), *Possibility and Necessity: The Role of Possibility in Cognitive Development*, trans. Helga Feider (Minneapolis: University of Minnesota Press).
- and Inhelder, Bärbel (1971), *Mental Imagery in the Child: A Study of the Development of Imaginal Representation*, trans. P. A. Chilton (New York: Basic Books).
- Putnam, Hilary (1975a), *Mathematics, Matter and Method, Philosophical Papers*, i (Cambridge: Cambridge University Press).
- (1975b), *Mind, Language and Reality, Philosophical Papers*, ii (Cambridge: Cambridge University Press).
- Quine, Willard van Orman (1969), *Ontological Relativity and Other Essays* (London: Columbia University Press).
- Sartre, Jean-Paul (1939/1962), *Imagination: A Psychological Critique*, trans. Forrest Williams (Ann Arbor: University of Michigan Press).
- (1940/1963), *The Psychology of Imagination*, trans. Forrest Williams (New York: Citadel Press).
- Shepard, Roger, and Cooper, Lynn (1982), *Mental Images and their Transformations* (Cambridge, Mass.: MIT Press).
- Shoemaker, Sydney (1998), 'Causal and Metaphysical Necessity', *Pacific Philosophical Quarterly*, 79: 59–77.
- Stalnaker, Robert (1978), 'Assertion', in Peter Cole (ed.), *Syntax and Semantics*, ix: *Pragmatics* (New York: Academic Press), 315–32.

- (2001), 'On Considering a Possible World as Actual', *Proceedings of the Aristotelian Society*, supp. vol. 65: 141–56.
- Strawson, P. F. (1970), 'Imagination and Perception', in L. Foster and J. W. Swanson (eds.), *Experience and Theory* (Amherst, Mass.: University of Massachusetts Press), 31–54.
- Tidman, Paul (1994), 'Conceivability as a Test for Possibility', *American Philosophical Quarterly*, 31(4): 297–309.
- van Cleve, James (1983), 'Conceivability and the Cartesian Argument for Dualism', *Pacific Philosophical Quarterly*, 64: 35–45.
- Walton, Kendall (1990), *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Cambridge, Mass.: Harvard University Press).
- Warnock, Mary (1976), *Imagination* (Berkeley and Los Angeles: University of California Press).
- Wittgenstein, Ludwig ([1958] 1965), *The Blue and Brown Books* (New York: Harper Torchbacks).
- Yablo, Stephen (1990), 'The Real Distinction Between Mind and Body', *Canadian Journal of Philosophy*, supp. vol. 16: 149–201.
- (1993), 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research*, 53(1): 1–42.
- (2000), 'Textbook Kripkeanism and the Open Texture of Concepts', *Pacific Philosophical Quarterly*, 81(1): 98–122.