

Jackson's Classical Model of Meaning

LAURA SCHROETER AND JOHN BIGELOW

I. Conceptual analysis and implicit knowledge

In his book *From Metaphysics to Ethics*, Jackson's defence of what he calls "conceptual analysis" proceeds in two stages. In the first stage, he argues for the existence of common knowledge shared by all speakers of a public language. He then brings this postulated common knowledge to bear on an array of philosophical problems, ranging from puzzles about secondary qualities like *colour* to meta-ethical concerns about moral appraisals of things as *right* or *bad* or *immoral*. The method of a priori conceptual analysis, he thinks, can yield important metaphysical conclusions by unearthing the commitments implicit in our shared linguistic norms.

It is plausible to suppose that speakers of a language have implicit knowledge of the way in which the reference of pronouns like 'I' and 'you' shift, depending on empirical facts concerning who has uttered them, who else was nearby, and so on. Jackson argues that we have similar implicit knowledge of the way in which the reference of a noun like 'water' or an adjective like 'red' will depend on empirical facts about the world around us. This sort of knowledge can be conveniently laid out in 2-D matrixes, in a manner that has been explained at length by Kaplan, Stalnaker, and others.¹ Such matrixes encode amazingly rich and complex patterns of common knowledge. This common knowledge is important, Jackson believes, because it underwrites the philosophical method of conceptual analysis.

Suppose Jackson were right to think that we have a rich stock of common knowledge concerning the ways in which word-world relationships

¹ See Kaplan 1989 and Stalnaker 1978 for seminal work on the 2-D semantic framework.

depend on complicated matters of empirical fact. Suppose he were right to think that common knowledge of this kind is an essential precondition for our speaking a common language. Suppose he were right to think that this common knowledge is in some important sense available to speakers of the language on careful reflection. Then it would be reasonable to conclude with Jackson that we can establish significant results in philosophy that are in some sense “analytic” (that is, their truth is determined solely by facts about the meanings of words) and that we could establish these results by a method that is in some sense “a priori”, a method that is conducted almost entirely by *careful reflection* rather than observation and experiment.

Thus, at the opening of chapter 4 of *From Metaphysics to Ethics*, we find the framework that has been articulated and illustrated in earlier chapters is brought to bear on philosophical problems concerning what are called *secondary qualities*, like colour. As Jackson says, we already know—near enough—roughly what are the relevant sorts of properties of material objects, light, eyes and brains; and what we know already is enough to present us with what he calls the *location problem*:

Which of those properties in the world are the ones that have colour names as well as scientific descriptions?

Jackson thinks his conceptual analysis supports an unequivocal verdict:

My answer is the ‘Australian’ view that colours are physical properties of objects: certain physical properties of objects have colour names as well as their physical property names. (Jackson 1998a: 87)

This is a substantial philosophical result. Jackson claims to establish this result by drawing on semantic theories for colour words modelled on the semantic theories he has sketched for words like ‘water’, which in turn are modelled on the 2-D semantic theories for words pronouns like ‘I’ and ‘you’. The upshot is that he claims to establish the “Australian” theory of colours as an analytic truth, which can be established a priori. “Them’s fightin’ words!” This is one of many substantial results that Jackson’s semantic theory promises to deliver.

You might contest the details of some of Jackson’s applications of his semantic theories. Yet even if he were to have made significant errors in application, we could still see the potential that the method would have if we were to grant his initial semantic framework. There would surely be some significant applications of the broad kind that he illustrates with examples like that of the “Australian” theory of colours, even if the right

applications were to differ a little from the sketches that Jackson actually offers.

The vulnerable point in Jackson’s project is his opening move. Jackson maintains that language use rests on a deep bed of *common knowledge*. The coordination required for a natural language does not, he assumes, come about by magic, and he is surely right about that. There must be something that explains it. Jackson maintains that this “something” is speakers’ implicit knowledge of *what it takes to be the reference of a word*. This implicit knowledge plays two important causal roles in Jackson’s account. The first role is to help causally explain how members of a linguistic community achieve coordination by means of public language symbols. The second role is to causally explain individual speakers’ verdicts about how to apply a word to particular hypothetical cases—thus, grounding the armchair methodology of conceptual analysis. But is shared implicit knowledge of the putative subject matter really the *only* way to explain these phenomena?

Jackson often writes as if his account of public language meanings in terms of descriptivist conventions were just plain common sense. How else are we to explain how different speakers manage to communicate using a public language? And how else can we explain how individuals arrive at confident judgments about the reference of their words in hypothetical scenarios? Our aim in this paper is to show just how controversial the psychological assumptions behind Jackson’s semantic theory really are. First, we explain how Jackson’s theory goes well beyond the commonsense platitudes he cites in its defence. Second, we sketch an alternative explanation of those platitudes, the improvisation model of meaning, which seems psychologically more realistic. Our conclusion is that the psychological picture presupposed by Jackson’s semantic theory stands in need of a more substantial defence than he has so far offered.

2. The role of the 2-D framework

Before we examine the psychological underpinnings of Jackson’s account, it’s important to get clear about the role played by his two-dimensional semantic framework. The core claim in Jackson’s account is that speakers implicitly know what it takes to be the reference of their own words. But this claim seems to directly contradict the standard lessons drawn from externalist thought experiments offered by Kripke and Putnam. The moral standardly drawn from Putnam’s Twin Earth case and Kripke’s Gödel/Schmidt case is that our words

'water' and 'Gödel' can have a determinate reference even if we have no accurate knowledge of what it takes to be the reference.²

Jackson rejects this moral. He argues that the real lesson to be learned from the externalist thought experiments is that, although we don't know the *essential nature* of what's picked out by names and natural-kind terms, we do know an indirect *reference-fixing criterion*. In the case of 'water', for instance, competent speakers don't necessarily know that water is H₂O, but they do know (roughly) that water is the actual stuff that fills lakes and oceans, runs from taps, and is good for drinking.³ Since we all agree what would count as water if the actual world turned out to be like Twin Earth, he argues, we must have some shared criterion for what it takes to be the reference of 'water'. In a slogan: we know how the reference depends on empirical facts about one's actual environment. So if Jackson is right, our grasp of names and natural-kind terms resembles our grasp of indexicals: in all three cases, competent speakers implicitly know a criterion for identifying the reference on the basis of contingent empirical facts about their environment.

A number of externalists have challenged Jackson's story at this point. The problem they see is that as soon as we actually specify a candidate reference-fixing criterion, we can think up counterexamples—situations in which the proposed reference-fixing criterion would yield the intuitively wrong verdict about the reference of 'water' or 'Gödel'. The worry, then, is that the ease with which counterexamples emerge every time a proposal is actually articulated suggests that there just *isn't* any uniform reference-fixing criterion that determines our verdicts for every way the actual world could turn out. (See, in particular, Block and Stalnaker 1999; Byrne and Pryor 2004; and Schroeter (forthcoming).)

Jackson is not too worried about this sort of objection. In Jackson's view, all this sort of objection establishes is that a particular proposed definition does not perfectly capture the criterion we actually rely on. But the fact that we haven't yet gotten the definition right so far doesn't show that there is no right answer to be had. Indeed, Jackson thinks that the very fact that we can generate convincing counterexamples to the proposed definitions shows that we *must* share some implicit reference-fixing criterion. (See Jackson 1994: 33; 1998a: 37–41.)

Jackson can even admit that there is no precise, non-homophonic definition of a word like 'water' or 'red' in English. It's very hard, for instance, to

precisely characterize what it takes to fall into the extension of 'red' without using the term 'red', for instance in some such phrase as 'looks red'. Even so, this doesn't establish that we have no shared implicit reference-fixing criterion. The impossibility of formulating a precise definition might simply be due to the limitations of our current natural language. Alternatively, the lack of a precise definition might be traceable to facts about our psychology: perhaps our ability to identify water involves a psychologically basic type of pattern recognition that cannot be decomposed in an illuminating way. The point is that the success or failure of Jackson's descriptivism does not turn on there being short, handy descriptions that sum up the putative reference-fixing criteria we associate with 'water'.

2-D matrices provide a convenient way of making explicit the sort of implicit knowledge Jackson thinks we associate with words like 'water'. Along the vertical axis of the matrix, we can array all the possible empirical contexts in which the speaker's practice with a word might be embedded. And along the horizontal axis, we can array all the possible worlds considered as a circumstance of evaluation. The 2-D matrix can then be used to record the speaker's ideal, fully informed verdicts about the reference of her word relative to different empirical contexts of use. Such a 2-D matrix would display the speaker's ultimate commitments about how the reference of her word varies depending on her empirical environment.

Jackson has emphasized, however, that this 2-D framework is not essential to his account of meaning. (See, e.g., Jackson 2003: section 2. iv.) Jackson starts with the conviction that we must have implicit knowledge of reference-fixing criteria for words like 'water', and he takes the 2-D framework as a convenient way of elucidating the structure of that knowledge and of tacking down its precise content. But nothing important to his account would be lost if we cast aside the 2-D framework and characterized Jackson's account of meaning directly in terms of the speaker's implicit criteria for specifying the reference of her words.

3. Lockean descriptivism

As we've seen, Jackson insists that English speakers *must* share an implicit reference-fixing criterion for 'water', even if we can never actually formulate a failsafe definition. Why is he so confident? The answer lies in his attraction to a broadly Lockean approach to meaning. The core Lockean idea is that linguistic communication is explained by tacit social conventions that require speakers

² For the classical statement of the thought experiments see Kripke 1972 and Putnam 1975.

³ See Jackson 1994: 37–40 and Jackson 1998a: 40–52. For an explanation of this distinction between grasping the essential nature of the reference and grasping a reference-fixing criterion for recognizing its essential nature, see Schroeter (forthcoming).

to associate certain reference conditions (and ultimately truth conditions) with words.⁴ This constitutes Jackson's basic working hypothesis about the nature of linguistic meaning.⁵

Almost everyone agrees that the connection between a word and its reference is established by conventions.⁶ The controversial bit concerns the precise content of these conventions. This is where Lockean descriptivism provides a simple and powerful explanatory hypothesis. The Lockean account starts with the bold claim that all competent speakers share a core set of assumptions about what it takes to be the reference of 'water'—an implicit "folk theory" of water—and that this fact is mutually obvious to everyone. If this is right, then the standard "folk theory" can ground a tacit linguistic convention governing how speakers are to understand the reference of 'water'. Given speakers' interest in coordinating their linguistic usage with those of others in their community, the fact that *most* of them associate certain assumptions with 'water' gives *everyone* a strong motive for accepting those assumptions themselves. Since these facts are mutually obvious to everyone, the standard set of assumptions will tend to become conventionally entrenched within the community: everyone will be motivated to accept the standard "folk theory" for 'water', but their motivation will be conditional on others' acceptance of the same theory.

Lockean descriptivism is a bottom-up approach to linguistic communication. It starts with an account of individual speakers' substantive understanding of what it is a word represents, and then claims that this understanding *also* provides the content of tacit linguistic conventions. Thus the very same theoretical posit—implicit knowledge of a "folk theory"—simultaneously helps explain both how individual speakers use words to categorize things and how different speakers manage to coordinate their understanding of words. This is a very natural and attractive account of what different speakers of a common language must share if they are to communicate.

To appreciate the explanatory role of the Lockean account, it's important to be clear about the phenomena to be explained. Jackson is right to emphasize that speakers of a common language share a great many attitudes and dispositions

⁴ This talk of the *reference conditions* of a word is intended as a neutral way of specifying the reference, either by specifying its *essential nature* or by giving an indirect *reference-fixing condition*.

⁵ Jackson advertises himself as a Lockean in a number of places. See, for instance, Jackson 1998a: 114; 1998b: 201–2. It's worth noting that this Lockean conception of language is *not* shared by Jackson's sometime co-author David Chalmers. Although Chalmers advocates a similar type of a priori conceptual analysis, he does not think it captures anything like the conventional linguistic meaning of words. See Chalmers 2002: section 8, for Chalmers' explicit rejection of this idea.

⁶ Quine once queried this platitude "Truth by Convention" but Lewis has convincingly defended it in *Convention* (Lewis 1969), and now it can be assumed without fear of significant dispute.

regarding the references of their words. Indeed, everyday coordination would be impossible if we did not. When Monima tells Dirk the sink is leaking water, he immediately understands something about the mess that's being created. And Monima knows that Dirk knows this: that's why she told him in the first place. And he knows that she knows that he'll know this: otherwise he wouldn't trust her utterance to be a reliable guide to the facts. And so on. Thus, some sort of common knowledge of the reference of words seems a precondition for ordinary communication.

We'll focus on three commonsense observations about language that need to be accounted for. Like Jackson, we're inclined to regard these as Moorean facts about language.

First, individual speakers' substantive understanding of a word puts them in a position to know about its reference, if provided with sufficient empirical information about their actual environment. For instance, if we describe for Dirk different ways the actual world might turn out to be like in sufficient detail, he will be able to come to correct verdicts about what would count as water in those scenarios. It's controversial, of course, what exactly the epistemological status of these judgments is. (Are they a priori? Are they analytic?) But it's *not* controversial that, given enough empirical information, speakers can come to correct verdicts about many cases. This simple psychological fact is what we're focusing on here.

Second, two interlocutors who share the same language will normally have substantially overlapping sets of assumptions about *what it is* their words represent—and they are generally aware of the rough extent of this overlap. It's common knowledge between Dirk and Monima, for instance, that water is clear, odourless drinkable liquid, that it fills rivers and oceans, that it's useful for washing clothes and making tea, and so on. Notice that the extent of this overlap will vary depending on the interlocutors involved. Since Dirk and Monima are both academics of a certain persuasion, it's also common knowledge between them that the chemical composition of water is H₂O and that water figures in one of Hilary Putnam's most famous thought experiments; but these assumptions are not common knowledge between Monima and the barista at the local café. In general, the list of assumptions that interlocutors treat as common knowledge about water will vary depending on how well they know each other and on what the conversation has previously included. What's undeniable, however, is that competent speakers are normally in a position to know a great deal about their interlocutors' conception of the objects, kinds, and properties their words refer to.

Third, for there to be successful linguistic communication, members of a linguistic community must know that others in the community use words to

refer to the very same object, kind or property. In short, competent speakers must know *that their words co-refer*. It's common knowledge between Dirk and Monima, for instance, that they each use the word 'water' to refer to the very same stuff. In fact, the phenomenon is a bit more specific than simple knowledge of co-reference. Intuitively, there is a sharp contrast between the kind of knowledge of co-reference involved when two words are taken to share a meaning (e.g. Dirk's use of 'water' and Monima's use of 'water') and when the words merely share their reference (e.g. Dirk's use of 'water' and Monima's use of 'H₂O'). Sharing a meaning makes words co-referential *de jure* rather than merely *de facto*. An account of linguistic meaning must be able to explain how this knowledge of *de jure* co-reference is possible.

Lockean descriptivism offers a powerful unified explanation of these three points. The first point concerns speakers' ability to know the reference of 'water' in different empirical scenarios. On the Lockean account, a speaker's "folk theory" determines the reference conditions for all empirical circumstances. So Dirk can know the reference of the term 'water' insofar as his judgments are "guided by" the assumptions that constitute his "folk theory" of water.⁷ The Lockean explanation of the second point is equally straightforward. Interlocutors like Monima and Dirk know they share a common set of assumptions about the reference of 'water', according to the Lockean, precisely because there is a conventionally entrenched "folk theory" associated with the term. Thus, it will be mutually obvious to everyone that most speakers associate certain core assumptions with the term 'water'. The Lockean explanation of the third point builds on the previous two. The reason Monima can know that competent English speakers' use of the word 'water' is *de jure* co-referential with her own is that she knows they associate the same "folk theory" with the term and she implicitly understands that the "folk theory" associated with a term determines its reference in any empirical circumstance. So Monima can know that other competent English speakers' use of 'water' *must* be co-referential with her own use, no matter what the actual world is like.

To sum up: Lockean descriptivism claims that all competent speakers associate such-and-such reference conditions with the word 'water'. The idea that an individual speaker implicitly treats these reference conditions as decisive explains her judgments about hypothetical scenarios. The idea that these reference conditions are associated with a term through a tacit linguistic convention explains how different speakers can successfully communicate via

language. If this Lockean approach to language is right, then Jackson needn't be too worried if he cannot produce a precise and informative specification of reference conditions for 'water' that perfectly captures our best judgments about possible cases—for his theoretical apparatus assures him that speakers *must* implicitly rely on such reference conditions. A 2-D semantic matrix can then be pressed into service as a way of making this implicit knowledge explicit.

Our aim in this section has been to spell out the Lockean approach and show how it affords a substantial and attractive explanation of uncontroversial features about language. In the next two sections, we'll consider whether the Lockean explanation is the *best* explanation of these uncontroversial matters. We'll first consider the Lockean explanation of individual speakers' referential abilities, and then turn to the Lockean explanation of communication.

4. Judgements about cases: the *Meno* fallacy

Speakers can know what their own words would refer to in hypothetical situations. For instance, you can know that 'water', as you currently understand the term, would refer to XYZ if the actual world were like Punam's Twin Earth.

It may seem that Lockean descriptivism provides the *only* viable explanation of this Moorean fact. After all, it's not as if your verdict is simply a matter of brute stipulation: you could get the facts about the reference of your term wrong. So how do you recognize the correct answer when you do find it? Clearly you must rely on your current implicit understanding of water. Just as bounty hunters rely on a picture on a government handbill in order to catch their man, there must be some implicit understanding you rely on in order to recognize what counts as the reference of 'water' in different empirical scenarios.⁸ Otherwise, you'd simply be groping in the dark. So it seems that, in order to explain your ability to make explicit knowledgeable judgments about the reference of 'water' in hypothetical scenarios, we must conclude that you have implicit knowledge of reference conditions for the term.

This line of reasoning is tempting. It is, however, fallacious. Consider the following analogous argument offered by Plato in the *Meno*. Plato starts by drawing our attention to an indisputable fact about ordinary subjects' ability to answer mathematical questions. Merely by asking a series of simple questions,

⁷ The image of being *guided* by one's "folk theory" is from Jackson's explanation of the role of conceptual analysis in metaphysics, Jackson 1998a: 31.

⁸ The image is Jackson's; see his 1998a: 31.

Socrates leads an uneducated slave boy to recognize that the Pythagorean theorem is true. It's clear that the boy doesn't just accept the theorem on Socrates' authority—he genuinely understands that it *must* be correct. How are we to explain this ability to recognize a complex mathematical fact? Clearly, the boy's answers were constrained by his prior understanding of mathematics. Plato concludes that the boy must have *implicitly known* the correct answer all along: Socrates' questioning was just the catalyst that transformed this implicit knowledge into explicit knowledge (through the process of "recollection").⁹

Surely Meno's slave didn't need to know anything as complicated and topic-specific as the Pythagorean theorem in order to recognize its truth on the basis of Socrates' leading questions. All he needed was an implicit understanding of much simpler geometrical facts (such as a rough grasp of what it takes to be a triangle or a right angle) together with some very general reasoning and imaginative capacities. The boy's implicit understanding of the simple geometrical facts constrained his answers to Socrates' questions about more complex geometrical facts. Only by chaining these intermediate conclusions together was he able to see (for the very first time!) that the Pythagorean theorem must be true. Given the availability of this alternative account, the Platonic suggestion that the boy already implicitly knew the theorem all along is unsupported by the evidence. In general, the mere fact that a person can answer a particular kind of question doesn't prove that he is guided by implicit knowledge of a dedicated, topic-specific criterion for answering it. It may be more psychologically realistic to say that the subject relies on general, topic-neutral reasoning capacities to arrive at a genuinely new piece of knowledge.

Similarly, when the Lockean descriptivist says you have implicit knowledge of the reference conditions for 'water', he is making a substantive and potentially controversial psychological claim. Jackson's image of the bounty hunters provides graphic illustration of the kind of view he has in mind. Bounty hunters rely exclusively on the relatively restricted set of images and descriptions printed on the handbill in order to recognize the wanted man. In the same way, Jackson thinks you rely exclusively on a relatively restricted set of assumptions that

⁹ Although Socrates' audience in the dialogue seems to find the hypothesis of recollection quite convincing ("I feel, somehow, that I like what you are saying"), readers need to be wary of what is known as Platonic irony. Socrates immediately goes on to say: "And I, Meno, like what I am saying. Some things I have said of which I am not altogether confident. But that we shall be better and braver and less helpless if we think that we ought to enquire . . ." Of course, hardly anyone nowadays agrees with the hypothesis that a mathematician recollects truths that he has learned "when he was not a man". (Cf pp. 85–6 in Jowett's translation of the *Meno*.)

comprise your "folk theory" of water in order to recognize what it takes to be water in any given empirical situation. This limited set of assumptions will then determine your verdicts about the reference of your word 'water', since you take the word to refer to water. Thus Jackson assumes that it is only the "obvious and central" assumptions about water that guide your judgments about hypothetical cases (Jackson 1998a: 31).

Why must the assumptions be limited in this way? Why not allow that your *total state of mind* is relevant to determining the reference of your term 'water'? The answer is that Jackson's Lockean account of communication requires that most speakers associate the very same "folk theory" with the term 'water'. But no two speakers share *all* their assumptions about a given topic. So the Lockean story requires that there be a certain limited set of assumptions about water—the obvious and central ones—that all speakers share and that guide individual speakers' judgments about hypothetical cases. These assumptions serve as a kind of topic-specific template for recognizing what it takes to be water. So just as different bounty hunters all rely on the same handbill, members of the same linguistic community all rely on the same "folk theory" of water.

But is this really the best psychological explanation of how individual speakers figure out what the word 'water' refers to? Does a privileged subset of assumptions guide your verdicts about what counts as water no matter what the actual world is like? Not obviously so. It may be that the best psychological explanation of your verdicts doesn't afford a privileged role to any determinate and fixed subset of your assumptions.

Here's one consideration that might seem to support the Lockean explanation. When considering what would count as water if the actual world were like Putnam's Twin Earth, you may have an immediate and straightforward conviction about the correct answer: XYZ. This immediacy seems an indication that you're relying on an implicit criterion for identifying the reference of 'water'. So it can seem that you *must* have some fixed template that guides your verdicts about possible cases.

As one of us has argued elsewhere, however, this impression is due entirely to a poverty of examples (Schroeter 2003). Twin Earth is specifically designed to vindicate all of your current beliefs about water except for one. It's no surprise, then, that verdicts about the reference of 'water' in this sort of context are fairly straightforward. Things are different, however, when you start to consider contexts that diverge more sharply from your current assumptions about your actual environment. What should you say about the reference of 'water' if it turns out that you've been interacting with a number of different chemical substances in different contexts? Could 'water' turn out to

be ambiguous? Or could it refer to a disjunctive kind? What if there hadn't been any coherent chemical kind here at all? Could water have turned out to be an ecological kind? Or perhaps a culinary kind?¹⁰ How can you adjudicate among these many possible different interpretations? Clearly you don't have anything like an automatic template that churns out answers to these sorts of questions. Instead, you need to engage in detailed hermeneutical deliberation about which interpretation makes best sense of our total state of understanding in each particular empirical situation.

Hermeneutical reasoning about such cases does not seem to conform to the Lockean thesis that only a circumscribed set of "obvious and central" assumptions are ever relevant to determining the correct verdicts about what counts as water. Different assumptions about water will be crucial to determining your verdict about what counts as water, depending on what you assume about your actual context of use. In some contexts, your conviction that water is a fundamental explanatory kind will guide your verdicts about what water is; in other contexts this assumption will be irrelevant. In some contexts, your conviction that water is potable may be crucial; in other contexts it will count as false. And so on. Why suppose that the assumptions that strike you as obvious and central given your *current* opinions about your environment will continue to strike you as obvious and central when you accept a very different hypothesis about your environment? From the perspective of the deliberator, virtually any assumption could potentially affect the plausibility of a particular interpretation of your word in some context of use or other. There just doesn't seem to be a clear demarcation between assumptions that are potentially relevant to determining the nature of water in some hypothetical context or other and assumptions that are always irrelevant to this project. Given the strong intuitive plausibility of epistemological holism, moreover, it seems unreasonable to expect such a thing.¹¹

The psychological picture that naturally emerges from our experience as deliberators is simpler than the Lockean account. It should be granted by all parties that all-purpose reasoning skills are crucial to your ability to recognize the facts about what your words refer to in novel empirical contexts. According to the Lockean descriptivist, your deliberation about what counts as water is

¹⁰ It is chastening to recall that of the four elements postulated by the ancient physical theory—fire, air, water and earth—water was the only one that turned out to be a chemical natural kind.

¹¹ It's a testimony to the fundamental appeal of epistemological holism, that Bertrand Russell found the doctrine plausible even when he was a logical atomist, and was semantically and metaphysically diametrically opposed to holism (see for instance Russell 1959, p. 140). More recently, epistemological holism has become solidly entrenched in the philosophical landscape through the influence of Quine (e.g. Quine 1953).

always guided, in addition, by a particular circumscribed set of assumptions no matter what you suppose about your actual empirical environment. This fixed set of topic-specific assumptions, together with your hypothesis about the actual world, provides the input for your deliberation about what counts as water. Our suggestion is that this Lockean constraint on inputs is unrealistic and should simply be abandoned. In deliberating about what counts as water you try to make best sense of your *total* prior epistemic perspective with respect to water, in the light of your new hypotheses about the nature of your environment. We might call this *the holistic model* of your ability to recognize correct answers about the reference of your words. This holistic model provides a more economical psychological explanation of the phenomena, and one that better meshes with the highly plausible case for epistemological holism.

Now, we don't pretend to have demonstrated that the Lockean model is false. Perhaps further empirical inquiry will reveal that you do implicitly rely on a circumscribed set of assumptions about water that figure as fixed constraints on deliberation about its essential nature. This is an empirical possibility. Further inquiry might also reveal that Meno's slave boy really did know the Pythagorean theorem all along. Our point is just that this hypothesis is not *required* in order to explain your judgments about possible cases. And sound psychological methodology counsels against positing extra psychological states unless they are required in order to explain the data.

5. Linguistic communication: Classical or jazz?

A proponent of Lockean descriptivism may not be too worried about this point. Perhaps the primary attraction of the view is its ability to explain *linguistic communication*, rather than its ability to explain individuals' judgments about reference in controversial thought experiments. Indeed, Jackson himself has stressed that it is the voluntary and conventional aspects of communication that are "crucial to the plausibility of the description theory of reference" (Jackson 1998b: 204). In this section, we'll consider whether Lockean descriptivism does afford the best explanation of tacit linguistic conventions.

The Lockean explanation of communication can seem like plain common sense. Everyone agrees that there are tacit linguistic conventions that associate words with reference conditions. And everyone agrees that interlocutors who share the same language have some common ways of understanding their words. The Lockean explanation of these facts is straightforward: there are

tacit linguistic conventions that associate words with ways of understanding them—i.e. implicit assumptions about what conditions an object or property or event would have to meet in order for it to be the reference of a given expression. Jackson seems to find this style of explanation inevitable.

However, the Lockean account of communication faces serious difficulties. It's helpful to break the Lockean account down into a series of discrete claims.

1. There is a relatively limited set of assumptions about a putative feature of the world—a speaker's "folk theory"—that guides an individual speaker's judgments about the reference of her word 'water', regardless of what she assumes about her actual empirical environment.
2. Most members of a linguistic community associate the very same "folk theory" of a putative feature of their environment with the term 'water'.
3. This convergence in "folk theories" is mutually obvious to all members of the community.
4. The fact that most people converge in this way gives everyone a justifying and motivating reason to associate the standard "folk theory" with the term 'water'.

The conclusion of this line of reasoning is that the assumptions standardly associated with the term 'water' will become conventionally entrenched within the linguistic community: everyone will be motivated to accept those assumptions about the reference of 'water', insofar as others accept them. The upshot is that English speakers can know they're using the word 'water' to refer to the very same stuff because they know that their words are associated with the very same guiding assumptions about the nature of the putative reference.

We agree that the conclusion is well supported by the initial assumptions. The trouble is that there seem to be good reasons to doubt each of these assumptions.

We've already seen in the previous section why one might doubt the first step of the Lockean explanation of communication. The Lockean posits a hard and fast distinction between assumptions that can affect a speaker's verdicts about the nature of water and assumptions that cannot. From the point of view of psychological explanation, however, this seems to be an unnecessary epicycle in the mechanisms guiding a speaker's judgments. And from the point of view of normative epistemology, the account seems to conflict with the holistic nature of epistemological reasoning and about the extent of our fallibility about familiar things like water.

Now consider the second step of the Lockean explanation. Assuming that each individual speaker does rely on a fixed set of guiding assumptions in

deliberation about what it takes to be water, how plausible is it that most speakers will converge on precisely the same set of guiding assumptions? The uncontroversial Moorean fact is that for there to be successful communication on a particular occasion, the interlocutors must share *some* background assumptions about what it is a word like 'water' refers to. But this is a much weaker claim than the Lockean claim that there is determinate set of context-independent guiding assumptions about what it takes to be a particular kind of stuff that most speakers associate with the term 'water'. In particular, the Moorean fact is compatible with the idea that each speaker's "folk theory" overlaps to *some extent* with that of every other individual speaker, but there is no determinate list of assumptions that everyone is required to accept in order to qualify as competent.¹²

One way to test the Lockean suggestion is by consulting your intuitions about what's required for competence with the term 'water'. If the Lockean argument is correct, then all competent speakers must associate the very same "folk theory" of water with 'water'. But on the face of it, there doesn't seem to be any non-negotiable set of assumptions about water that competent English speakers are required to share. Take any widespread assumption about water—say, that it's potable, or that it fills the oceans—we can imagine someone who failed to share that assumption and yet who would strike us as linguistically competent with the term 'water'. Perhaps the speaker was raised in very unusual circumstances—in the Sahara, or on Mars. Or perhaps she has a weakness for eccentric empirical hypotheses about the non-potability of water. It is easy to imagine that we might in such cases have good reason to think that we were talking about the very same stuff, and that we'd be inclined to see such a person as ignorant or misguided rather than linguistically incompetent. *Prima facie*, then, the Lockean thesis that competent speakers must accept a determinate "folk theory" looks dubious.

This brings us to the third step of the Lockean explanation: given that most speakers associate the very same set of assumptions with 'water', their convergence on this set of assumptions will be *mutually obvious* to all speakers. Now, clearly we often do know roughly what another person is assuming about the stuff they call 'water'. The point we'd like to stress here is a relative one. Normally, facts about the reference of a speaker's word are a lot more obvious than facts about the speaker's conception of the reference. For instance, if a

¹² There's an important scope distinction here. The uncontroversial claim has the form:

$$(\forall x)(\forall y)(\exists z)(x \ \& \ y \ \text{are successful interlocutors and they share a set of assumptions } z)$$

whereas the controversial Lockean claim reverses the scope of the quantifiers:

$$(\exists z)(\forall x)(\forall y)(x \ \& \ y \ \text{are successful interlocutors and they share a set of assumptions } z).$$

foreign speaker points at a rabbit and says 'Gavagai!' in the right sort of context, it'll strike you as obvious that her word refers to rabbits. *Pace* Quine, we think you can even know this fact. However, you may have very little insight into the speaker's "folk theory" of rabbits. Does she think they're good to eat? Are they sacred totems? Is there any detailed rabbit lore that everyone hereabouts is privy to? You may have no idea. Similar observations apply to speakers getting initiated into their native languages. When a toddler's mum points to a furry creature and says 'Look Trev—a rabbit!' he'll immediately acquire a rough idea of what kind of thing her word 'rabbit' refers to. But it's much less likely that he'll have any reliable insight into her implicit assumptions about rabbits (e.g. 'they breed quickly, they're a biological species, they're related to hares, they're a pest in Australia, they're Easter totems, they are often eaten as stew ...'). On the Lockean account, it's the mutual obviousness of speakers' conception of the reference that is supposed to explain how it can be mutually obvious that they use words co-referentially. *Prima facie*, however, this seems to get things precisely backwards.

Finally, let's consider the fourth step of the Lockean explanation. Given that it's mutually obvious that most people accept a particular "folk theory" about the reference of 'water', everyone has a good reason to do likewise. Once again, we think there are grounds for scepticism. For the Lockean, a "folk theory" is a set of assumptions about a putative feature of the world that will guide a speaker's judgments about reference, regardless of what she assumes about her actual empirical environment. Accepting a standard "folk theory" for water, therefore, does not merely require you to accept that other speakers assume such-and-such about water. In addition, you must accept the standard assumptions yourself, using them to guide your judgments about what counts as water no matter what the actual world turns out to be like. But do you really have reason to accept such beliefs simply because other people accept them? On the face of it, it seems you could have sound epistemological reasons for rejecting standard assumptions about water. And it seems that if you did reject those assumptions, you would not thereby cease to be a competent English speaker. When Galileo insisted that the earth moves, he wasn't just changing the subject!

Each important claim in the Lockean account of communication seems to face difficulties in accommodating the commonsense point of view on language. We hope these observations have served to raise some doubts as to the adequacy of the Lockean approach. But we don't pretend these purely negative considerations are decisive reasons to reject the Lockean model. In the absence of any plausible alternative explanation of the data about communication, it's reasonable to hang on to the Lockean account. A leaking

raft, after all, is better than no raft at all. And perhaps it can be patched up. No one should abandon the Lockean model unless there is something more seaworthy in the neighbourhood. As it happens, we think there is an attractive alternative.

We'd like to propose a new model of communication. This model, we believe, provides a natural and unified explanation of the points we've raised as difficulties for Lockean descriptivism. The best way to appreciate the contrast between the two models is to focus on a central explanatory strategy of the Lockean account of communication.

In section 3, we highlighted two commonsense observations about communication that any acceptable theory should be able to explain.

1. *Substantive knowledge about the reference*: two interlocutors normally have substantially overlapping sets of assumptions about *what it is* their words represent and they are aware of this overlap.
2. *Knowledge of de jure co-reference*: two interlocutors normally know that their respective uses of a given term are co-referential and they take this co-reference to be *de jure* rather than merely *de facto*.

It would be surprising if these two kinds of linguistic knowledge were entirely independent. Lockean descriptivism provides a unified account, according to which interlocutors' knowledge that their use of a word is *de jure* co-referential is *based on* their knowledge that they associate the same set of assumptions with the word. Two interlocutors, say Monima and Dirk, may know that their respective uses of the term 'water' are *de jure* co-referential precisely because they know each of them associates the very same "folk theory" with the term. Since a person's "folk theory" constrains what counts as the reference of the term no matter what the actual world is like, the fact that Monima and Dirk share the very same "folk theory", means that there are no empirical circumstances in which the reference of Monima's word 'water' could diverge from the reference of Dirk's word—their words necessarily refer to the same stuff. It is the appreciation of these facts that explains why the two interlocutors take their words to be *de jure* co-referential. Thus, according to Lockean descriptivism, the first kind of knowledge is more basic: it both justifies and causally sustains the second kind of knowledge.

Our starting point is to reverse the direction of explanation posited by the Lockean account: it's your knowledge that other speakers' words are *de jure* co-referential with your own that is more basic than knowledge of others' substantive assumptions about the reference. When you enter into conversation with someone new, you can know that the other person's use of 'water' is co-referential with your own *before* you have any beliefs about the details of

their associated assumptions. The knowledge that your interlocutor uses 'water' co-referentially, together with your appreciation of the standardizing effects of communication and your assumptions about your interlocutor's idiosyncratic epistemic history, helps justify and causally sustain your knowledge of his substantive assumptions about water.¹³

But how, you may wonder, could one possibly know that another person's word was *de jure* co-referential with one's own without any information about her assumptions as to the nature of the reference? Wouldn't this just be a blind leap of faith? Surely knowledge that another person is referring to the same object, kind, or property as oneself must be based on *some* evidence! And surely the relevant evidence must ultimately concern the speaker's substantive understanding of the reference of her word.

We agree that knowledge of co-reference must be based on evidence. But we don't agree that you need evidence that the other speaker endorses some determinate list of assumptions about the reference of her term 'water'. It's enough if you have evidence that the other person *intends to use her words co-referentially with others in her linguistic community*.

An analogy might help explain the difference here. Whereas the members of a classical string quartet master a common score that serves to coordinate their performance, members of a jazz quartet are simply committed to building on whatever musical themes the other members try out. Each member trusts that the others will try to take his individual contributions on board and incorporate them into a coherent musical structure. Thus, the musicians' mutually obvious cooperative intentions keep the musical performance on track, developing interesting themes rather than degenerating into a welter of random notes. Because of these cooperative intentions, the musicians do not need to agree on any substantive plan in order to achieve a coordinated performance. We suggest that a similar structure of coordinated intentions can explain speakers' ability to coordinate on a joint representational practice with public language expressions like 'water'. We might call our account the *jazz model* of meaning—but perhaps it would be helpful to give it the more sober and informative title, *the cooperative intention model of de jure co-reference*. The key idea is that speakers' cooperative intentions to use a given word co-referentially

¹³ This is not to deny that you can have other, non-linguistic evidence of another person's conception of water. If you see someone lounging by the pool in an international hotel, with a book and a glass of water by her side, you can be justified in forming *some* ideas about what her conception of water is like without knowing that she even speaks English. Our point is that you don't *need* any precise understanding of the set of assumptions she takes to be crucial to determining what counts as water in order to know that her use of the word 'water' is co-referential with your own. In fact, you're unlikely to have access to anything so detailed, as our 'Gavagai' and 'rabbit' examples illustrate.

help constitute the *fact* that their uses of the word are co-referential. When speakers' words co-refer simply in virtue of such intentions, their co-reference is *de jure*—otherwise, it's *de facto*.

On this account, knowledge of *de jure* co-reference will be based on evidence that another speaker intends her use of words to be co-referential with that of others in her linguistic community. Such evidence is quite easy to come by. For instance, the mere fact that the speaker is addressing you with familiar words and grammatical structures is strong evidence that she intends to invoke common referential practices English speakers have built around those expressions. Speakers wear their co-referential intentions on their sleeves.

The obviousness of such cooperative referential intentions makes them well suited to figuring in tacit linguistic conventions. Everyone has an interest in making sure their use of a word is co-referential with that of others in the linguistic community, provided that others are ready to do likewise. And it's obvious that almost everyone does in fact have the relevant co-referential intentions. So, according to the standard Lewisian account of tacit conventions, there is a tacit linguistic convention to *make sure one's own use of the word 'water' refers to the same thing as others'*. This sort of bare co-referential intention, we're suggesting, exhausts the content of tacit linguistic conventions.

If you're sympathetic to the Lockean approach, this linguistic convention may strike you as inadequate. Unlike mere wishes, intentions require that the subject be able to envisage some realistic means of bringing about their fulfilment. And this is precisely what the Lockean account provides—realistic means for guaranteeing that different speakers co-refer. According to the Lockean descriptivist, speakers have a tacit convention to *treat such-and-such assumptions as determining the reference of 'water', come what may*. It's tempting to suppose that something along these lines is the *only* realistic means to guaranteeing co-reference: speakers must settle in advance on a determinate plan for how each word is to be understood.

However, jazz provides a viable alternative model of how speakers might achieve coordination. We don't need to settle on any determinate plan for exactly how a particular word is to be understood, as long as we're willing to try to make the best of everyone's contributions. Like jazz musicians, we can just improvise together.

To see how this might work, let's consider once again how individual speakers make judgments about what counts as the reference of their own word, 'water'. In the previous section, we suggested that individual speakers must rely on general hermeneutical reasoning to figure out what 'water' would refer to in surprising empirical contexts of use. What are the defining characteristics of water if it turns out that there is no single, underlying

explanatory kind that unites most of your paradigm examples? In responding to variations on this sort of case, it doesn't seem that you can restrict your attention to some determinate subset of assumptions about water. Instead, you rely on general hermeneutical skills to find an interpretation that makes best sense of your entire prior understanding of water.

Now, what happens when we include *other speakers* in the picture? How are you to make sure that your use of 'water' is co-referential with theirs no matter what the actual world turns out to be like? One way, of course, is the Lockean's: everyone can agree in advance on a determinate plan as to which features are relevant to deciding what counts as water. But there's another, easier way: you can simply take the assumptions *other speakers* associate with 'water' into account when revising your own beliefs about what counts as water. Of course, you may not know precisely what those assumptions are—you might have to wait for them to tell you. But the key idea is that you take the correctness of a given interpretation of 'water' to be answerable to *everyone's* assumptions about the reference, not just to your own assumptions. If other speakers are prepared to do likewise, then general hermeneutical skills will tend to mandate the same interpretation of 'water' for everyone, no matter what the actual world turns out to be like. The effect, then, of such cooperative hermeneutical intentions is to extend the basic unit of semantic interpretation from an individual's practice with a word to a whole community's practice with the word.¹⁴

On this approach to meaning, one must carefully distinguish two different explanatory roles: on the one hand, there are the facts that explain *de jure* co-reference, on the other hand there are the facts that explain why individuals' words refer to certain things. Our suggestion is that speakers' (undefeated) co-referential intentions play the first role, whereas holistic hermeneutical reasoning starting from the entire relevant set of assumptions plays the second role. In contrast, the Lockean descriptivist posits a single thing—the "folk theory" a speaker associates with a term—to play both roles.

There is plenty of work still to be done in fleshing out and motivating the cooperative intention approach to communication. Still, we think enough has been said to show that it represents a promising alternative to Lockean

¹⁴ We don't mean to suggest that having the requisite cooperative intention is a *failsafe guarantee* that your word is co-referential with others'. If the understanding an individual speaker associates with a word is extremely deviant, her co-referential intentions can be defeated. When Mis Malaprop declares that she finds the pope "exceedingly meretricious", she intends her word 'meretricious' to be co-referential with ours. However, she also assumes being meretricious is something like being meritorious—indeed, she hasn't an inkling of what the rest of us understand by 'meretricious'. Given such a radical divergence between the individual's understanding and that of the rest of the group, the individual speaker's own interests win out in determining the idiolect meaning of her word.

descriptivism. The model does as well as the Lockean model in explaining the Moorean facts about communication, while making much more modest claims about what is mutually obvious to all competent speakers. We think the weight of psychological and epistemological reasons militate favour of our proposal rather than the Lockean descriptivist approach. However, making that case in full is a task for another occasion.

Our goal in this section has simply been to cast doubt on Jackson's view that Lockean descriptivism constitutes the best—perhaps the only—strategy for explaining communication. We've argued that Lockean descriptivism is a substantive explanatory hypothesis that faces a number of important difficulties. In addition, we've argued that the Lockean approach is not the only viable contender. The moral we draw here will be a modest one: Lockean descriptivism is much less commonsensical than Jackson makes it out to be. Before jumping confidently onto Jackson's Lockean descriptivist bandwagon, you should insist on seeing a more careful articulation and defence of the view than Jackson has so far put forward.

6. Descriptivism redux?

In closing, we'd like to briefly consider a line of response one might be tempted to make on Jackson's behalf. It might seem that we've simply lumbered Jackson with an overly restrictive version of Lockean descriptivism. Perhaps on a more charitable understanding of the Lockean approach, there is no genuine conflict with our preferred explanation of communication.

We've assumed that Jackson's Lockean descriptivism requires that speakers have an implicit convention to the effect that such-and-such *topic-specific assumptions about water* are to guide verdicts about the reference of the term 'water' come what may. But perhaps this is too restrictive. Perhaps what's really crucial to Lockean descriptivism is just two ideas:

1. Speakers implicitly understand what their own linguistic conventions require of them.
2. This implicit understanding of linguistic conventions puts a speaker in a position to identify the reference of her word on the basis of ideal empirical information and ideal rational reflection.

If this is all that Jackson is claiming, then the cooperative intention model of communication does not represent a genuine alternative to his Lockean descriptivism. The first of the two claims just follows from the nature of tacit

linguistic conventions: there simply couldn't be a tacit convention in play unless participants had some implicit understanding of what the convention required of them. The second claim is similarly uncontroversial: it simply amounts to the idea that a speaker can, in principle, learn what her own convention-governed words refer to. Together, these two claims imply that competent speakers know something that puts them in a position to identify the reference of their words. However, the two claims leave it open exactly how speaker's implicit knowledge is structured.

How does the cooperative intention model fit this deflated characterization of Lockean descriptivism? According to the cooperative model, speakers have an implicit convention that requires that verdicts about the reference of 'water' must be guided by hermeneutical reasoning based on *all the assumptions that anyone in the community associates with the term 'water'* together with the empirical facts about their actual environment. Speakers implicitly know that this is what is required of them and this implicit understanding puts them in a position to identify the reference of their term on the basis of inquiry into their social and physical environment. The crucial point about this proposal is that it doesn't require the speaker to rely on any particular topic-specific assumptions about water that serve as necessary guides or templates for identifying the reference of 'water'. All that is required of a speaker is that she take *everyone's* topic-specific assumptions about the putative reference of 'water' into account when coming to a verdict about what water is.

Would it be more charitable to Jackson to count our cooperative intention model as just a particular version of Lockean descriptivism? We think not. There are two important respects in which this proposal fails to live up to Jackson's theoretical needs.

First, we find the purely formal and meta-linguistic character of the tacit conventions governing 'water' extremely hard to square with Jackson's insistence that speakers grasp an implicit "folk theory" of water. There is no *theory of water* that the competent speaker is required to accept on this account. Instead, there's a commitment to treat any theory (of whatever!) that's associated with the word 'water' by a member of the community as an input into hermeneutical reflection. This strikes us as an important departure from the spirit of Jackson's Lockean descriptivism. Jackson has repeatedly stressed that in order to achieve linguistic coordination, two speakers must share particular object-level assumptions about the stuff they're talking about—and it is these topic-specific assumptions that he takes to figure in tacit linguistic conventions.

Second, and more importantly, the cooperative intention model cannot underwrite the kind of a priori metaphysics that Jackson has powerfully advocated. In his book, Jackson argues that a priori metaphysics is underwritten

by the fact that certain "obvious and central" assumptions about the precise nature of, say, water or redness guide our deliberation about the nature of these things no matter what we assume about our actual environment. But why should these assumptions count as a priori? After all, any specific claim—like 'water is a natural kind' or 'the redness of objects helps causally explain our perceiving them as red'—seems empirically defeasible. Jackson's answer seems to hinge on a claim about what's required in order to use a word with a particular meaning: if you don't accept the relevant assumptions about water or redness as "obvious and central", you simply do not associate the same meaning with the words 'water' and 'red' as the rest of us. This claim about what it takes to use a word with a particular meaning grounds a kind of analytic truth: it's true *in virtue of the standard conventional meaning of 'water'* that water is the stuff that satisfies such-and-such assumptions about its nature, or comes suitably close to satisfying them. Since this "folk theory" of water is true (or close enough) in virtue of the meaning of the word 'water', it's reasonable to conclude that it enjoys a kind of a priori status.¹⁵

This defence of a priori metaphysics hinges on the idea that there is a single thing—the "folk theory"—a speaker associates with a word—that plays two distinct roles: it provides reference conditions for the speaker's word and it provides a standard for judging whether the speaker uses the word with the same meaning as others. The cooperative improvisation model, however, postulates two distinct things to play these two different roles: sameness of meaning is settled by (undefeated) co-referential intentions, whereas the actual reference is determined by the relevant set of substantive assumptions associated with a word. Since no particular assumption is required to use a word with the same meaning, there are no quasi-analytic truths of the sort that seem to be required for Jackson's defence of a priori metaphysics.

Our conclusion is that Jackson's philosophical program presupposes a classical, not a jazz model of meaning. A priori metaphysics and a priori conceptual analysis of the sort Jackson advocates require speakers to treat certain substantive assumptions about the putative reference of their words as true (or close enough to true) come what may. And it requires that we see acceptance of

¹⁵ In the joint paper "Conceptual Analysis and Reductive Explanation", Chalmers and Jackson (2001) advance a somewhat different account of the a priori status of conceptual analysis. They suggest that the conditional claim 'if the actual world is like Twin Earth, then water is XYZ' counts as a priori simply because all the relevant empirical information has been included in the antecedent of the conditional. On this approach, the a priori status of metaphysics would be secured by the simple expedient of making all metaphysical claims conditional ones. The same strategy could be used to secure the a priori status of physics or any other empirical discipline. Jackson's defense of conceptual analysis in his book, however, seems to focus on a stronger, *non-conditional* kind of a priori knowledge grounded in one's mastery of meanings. This is the line of thought we address in the text.

these substantive assumptions as settling whether a speaker is competent with the meaning of a word. This picture of language affords a powerful, unified explanation of many linguistic phenomena. However, it is not the *only* way of explaining these phenomena. Nor, we suspect, is it the *best* way—for the Lockean descriptivist model clashes with important psychological and epistemological commitments. These commitments provide the fundamental grounds for dissatisfaction with Jackson's defence of a priori metaphysics.

University of Melbourne
Monash University

References

- Block, N. and Stalnaker, R. 1999. 'Conceptual analysis, dualism, and the explanatory gap', *Philosophical Review* 108: 1–46.
- Byrne, A. and Pryor, J. 2004. 'Bad intensions'. In M. García-Carpintero and J. Macià (eds.), *The Two-Dimensionalist Framework: Foundations and Applications* (Oxford: Oxford University Press).
- Chalmers, D. 2002. 'On sense and intension'. In J. Tomberlin (ed.), *Philosophical Perspectives* 16: *Language and Mind* (Oxford: Blackwell).
- Chalmers, D. and Jackson, F. 2001. 'Conceptual analysis and reductive explanation'. *Philosophical Review* 110: 315–61.
- Jackson, F. 1994. 'Armchair metaphysics'. In M. Michaelis and J. O'Leary-Hawthorne (eds.), *Philosophy in Mind*, Philosophical Studies Series, vol. 60 (Boston: Kluwer).
- 1998a. *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Oxford University Press).
- 1998b. 'Reference and description revisited', *Philosophical Perspectives* 12: 201–18.
- 2003. 'From Water to H₂O: The relevance to a priori passage'. In H. Lillehammer and G. Rodriguez-Pereyra (eds.), *Real Metaphysics* (New York: Routledge).
- Kaplan, D. 1989. 'Demonstratives'. Reprinted J. Almog, J. Perry, and H. Wettstein (eds.), *Themes From Kaplan* (Oxford: Oxford University Press).
- Kripke, S. 1972. *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- Lewis, D. 1969. *Convention* (Cambridge, MA: Harvard University Press).
- Plato 1952. 'Meno'. In B. Jowett (trans.), *The Dialogues of Plato* (Chicago: University of Chicago Press).
- Punam, H. 1975. 'The meaning of "meaning"'. *Minnesota Studies in the Philosophy of Science* 7: 131–93.
- Quine, W. 1936. 'Truth by convention'. In O. H. Lee (ed.), *Philosophical Essays for A.N. Whitehead* (New York: Longman's, New York).
- 1953. 'Two dogmas of empiricism'. Reprinted in *From a Logical Point of View* (Cambridge, MA: Harvard University Press).

- Russell, B. 1959. *The Problems of Philosophy* (London: Oxford University Press).
- Schroeter, L. 2003. 'Gruesome diagonals', *Philosophers' Imprint* 3: <http://www.philosophersimprint.org/003.003/>.
- (forthcoming). 'The limits of conceptual analysis'. *Pacific Philosophical Quarterly*.
- Stalnaker, R., 'Assertion'. In P. Cole (ed.), *Syntax and Semantics* vol. ix (New York: Academic Press).