

## ABSTRACT

Title of Dissertation: TWO-DIMENSIONALISM AND THE PROBLEM OF  
INFORMATIVITY

Melissa Anne Ebbers, Doctor of Philosophy, 2013

Dissertation directed by: Professor Paul Pietroski  
Departments of Philosophy and Linguistics

I investigate two intuitions regarding the reference of proper names and natural kind terms that have received significant attention in the philosophy of language, and I discuss the role that they have played in modeling communicative exchanges using the two-dimensional framework in the views defended by Robert Stalnaker, Frank Jackson, and David Chalmers. I call the first intuition the *uniform extension intuition*: the intuition that proper names and natural kinds refer uniformly across worlds. This means that a proper name uniformly refers to the same individual—namely, the individual to which it refers at the actual world—across all worlds at which refers to anything at all, and a natural kind term uniformly refers to all and only samples of the same kind—namely, the same kind to which it refers at the actual world—across all worlds at which it picks out anything at all. I call the second intuition the *contingent extension intuition*: had the world turned out differently in the relevant respect(s), the extensions of the relevant expressions would have

differed accordingly. This means that, for example, had the kind that we've been calling water turned out to be, say, XYZ, 'water' would have referred to XYZ. An attraction of the two-dimensional framework is that it allows us to accommodate both of these intuitions.

I argue that using the two-dimensional framework to model communicative exchanges allows for what I call *the problem of informativity*: the problem of modeling the impact of an assertion on the state of a conversation. I argue that there is a class of cases for which Stalnaker's, Jackson's, and Chalmers's views do not accurately model the way that the assertion changes the state of the conversation. I offer a general framework of a solution to the problem as it arises on their accounts, and then give two interpretations of that framework: the first serves as a solution for Stalnaker's account, and the second serves as a solution for Jackson's and Chalmers's accounts. I conclude by discussing the consequences for each view of adopting the proposed solution.

**TWO-DIMENSIONALISM AND THE PROBLEM OF INFORMATIVITY**

by

Melissa Anne Ebbers

**Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
DOCTOR OF PHILOSOPHY  
2013**

Advisory Committee:

Professor Paul Pietroski, Chair  
Professor Peter Carruthers  
Professor Erin Eaker  
Professor John Horty  
Professor Allen Stairs

## **Acknowledgements:**

I would like to express my gratitude for the many wonderful individuals who have been instrumental in the success of this dissertation as the project evolved through its many incarnations before finally setting in its current form.

Foremost, I would like to express my appreciation for my advisor, Paul Pietroski. I have both very much enjoyed, and greatly benefitted from, our many conversations over the years. I have been very fortunate to have an advisor that both helped me to further develop the skills required for engaging in independent, self-directed research, and also inspired and taught me so much, and challenged me greatly along the way. I would also like to give special thanks to Erin Eaker for her very helpful, insightful conversations and feedback throughout the process. Thanks, also, to the rest of my committee—Peter Carruthers, John Horty, and Allen Stairs—for their feedback and support. Thanks to Georges Rey for initially piquing my interest in this topic in his Cognitive Science seminar, and for providing valuable feedback in the early stages of my research into two-dimensionalism. I would like to thank the Philosophy Department at the University of Maryland more generally for their support through my graduate studies.

I would also like to thank the Philosophy Departments at Rutgers and the Australian National University for the opportunity to be a Visiting Scholar during the 2007-2008 school year, and for both Fall 2010 and Fall 2011, respectively. Both experiences were incredibly fruitful, and played an important role in the ideas developed and defended in this work, or in work from which this project descends. Especially helpful were discussions with Jeff King, John Hawthorne, Jason Stanley (especially discussions on the material for his course on actualism), Jonathan Jenkins Ichikawa, David Chalmers, Wolfgang Schwarz, Paolo Santorio, with a very special thanks to Daniel Nolan, whose substantial feedback was invaluable.

I would also like to thank Graeme A. Forbes for both many valuable discussions in the later stages of the project, and for considerable assistance with the formatting of this document.

I would like to acknowledge the valuable feedback that I received from the audiences of my presentations at the Towards a Science of Consciousness Conference in Tucson, AZ (2008), the Essentialism conference at the University of Davis (2008), as invited Distinguished Alumna Speaker at Western Michigan University's Graduate Conference (2008), at the Australasian Association of Philosophy meeting (2010), and at talks presented to the Philosophy Department at the Australian National University (2010, 2011).

Lastly, I would like to express my deep gratitude to my wonderful family (especially my parents) and friends for all of their encouragement and support throughout the project.

## Contents

<b>Chapter One: Two-Dimensional Semantics and the Problem of Informativity</b>	<b>1</b>
<i>Section 1: Overview</i> .....	1
<i>Section 2: A Paradoxical Pair of Intuitions?</i> .....	5
<i>Section 3: Accounting for both Intuitions Using the Two-Dimensional Framework</i> .....	14
<i>Section 4: Using the two-dimensional framework to model conversations</i> .....	25
<b>Chapter Two: Stalnaker's Metasemantic Interpretation</b>	<b>42</b>
<i>Section 1: Stalnaker on the Force of Assertions on the Standard Interpretation</i> .....	42
<i>Section 2: Direct Reference Theory</i> .....	46
<i>Section 3: Which Worlds are in the Context Set?</i> .....	49
<i>Section 4: Diagonalization and the Necessary A Posteriori</i> .....	51
<i>Section 5: Some Mistakes Left Unexplained</i> .....	72
<b>Chapter Three: The Epistemic Interpretation of the Two-Dimensional Framework</b>	<b>95</b>
<i>Section 1: Jackson and Chalmers Put the Intuitions to Work</i> .....	97
<i>Section 2: Updating Hypotheses About the Way the World is in Light of Assertions</i> .....	126
<i>Section 3: First-order property</i> .....	131
<i>Section 4: Second-order property</i> .....	149
<i>Section 5: Picking Out All and Only Samples of the Same Kind</i> .....	156
<i>Section 6: Two Ways of Developing the Second-Order Property</i> .....	172
<i>Section 7: Bad Predictions</i> .....	174
<i>Section 8: Diagnosis of the Problem</i> .....	177
<b>Chapter Four: Diagnostics and Background for My Three-Dimensional Proposal</b>	<b>179</b>
<i>Section 1: An Introductory Sketch of the Proposal</i> .....	179
<i>Section 2: On Knowing How an Individual or Kind is Presented</i> .....	186
<i>Section 3: A Solution Employing a Three-Dimensional Framework</i> .....	195
<i>Section 4: Constructing the Three-Dimensional Matrix</i> .....	197
<i>Section 5: Constructing the Three-Dimensional Approximating Matrix</i> .....	222
<i>Section 6: Evaluating the General Three-Dimensional Reinterpretation as Proposed as an Extension to the Two-Dimensional Approaches</i> .....	237
<b>Chapter Five: Three-Dimensional Approach as Extension to Stalnaker's Account</b>	<b>243</b>
<i>Section 1: Pragmatic Reinterpretation</i> .....	243
<i>Section 2: Pragmatic Reinterpretation and the Three-Dimensional Matrix</i> .....	245

<i>Section 3: Why Adopt the Extension?</i> .....	260
<b>Chapter Six: Proposed Extension to Jackson’s and Chalmers’s Accounts</b>	<b>263</b>
<i>Section 1: The Proposed Extension as Pragmatic Reinterpretation</i> .....	263
<i>Section 2: A Proposed Extension to Jackson’s Account</i> .....	264
<i>Section 3: A Proposed Extension to Chalmers’s Account</i> .....	280
<i>Section 4: The Extended Accounts and the A Priori</i> .....	298
<i>Section 5: Is Conceivability a Good Guide to Possibility?</i> .....	315
<i>Section 6: Concluding Diagnostic Remarks and Looking Forward</i> .....	317
<b>References:</b>	<b>324</b>

## Table of Figures

<b>Chapter One: Two-Dimensional Semantics and the Problem of Informativity</b>	<b>1</b>
<i>Figure 1</i>	16
<i>Figure 2</i>	20
<i>Figure 3</i>	21
<i>Figure 4</i>	22
<i>Figure 5</i>	23
<i>Figure 6</i>	32
<i>Figure 7</i>	38
<b>Chapter Two: Stalnaker’s Metasemantic Interpretation</b>	<b>42</b>
<i>Figure 1</i>	56
<i>Figure 2</i>	59
<i>Figure 3</i>	63
<i>Figure 4</i>	79
<i>Figure 5</i>	81
<b>Chapter Three: The Epistemic Interpretation of the Two-Dimensional Framework</b>	<b>95</b>
<i>Figure 1</i>	129
<b>Chapter Four: Diagnostics and Background for My Three-Dimensional Proposal</b>	<b>179</b>
<i>Figure 1</i>	222
<i>Figure 2</i>	224
<i>Figure 3</i>	228
<i>Figure 4</i>	229

## Chapter One: Two-Dimensional Semantics and the Problem of Informativity

### *Section 1: Overview*

In this project, I investigate two intuitions about the reference of proper names and natural kind terms that have received wide-spread attention especially in the last thirty years, and the role they play in modeling communicative exchanges within a two-dimensional framework. I will first introduce the two intuitions, and then discuss the role that they play in modeling communicative exchanges within a two-dimensional framework.

The first intuition is that proper names and natural kind terms always refer to the same individual or all and only samples<sup>1</sup> of the same natural kind, respectively, across all worlds at which they pick out anything at all. The second intuition is that, had our world turned out differently in the relevant respect(s), even though the reference of the expression was fixed by using the same reference-fixing procedure as at the actual world, the proper names and natural kind terms would have referred to something other than that to which they actually refer.

As I discuss in the next section, on the face of it, this pair of intuitions appears paradoxical. How can it be, for example, that ‘water’ both always picks out all and only H<sub>2</sub>O across all worlds at which it picks out anything at all, *and* that, had another world been actual—for example, one at which the predominant clear odorless liquid is

---

<sup>1</sup> I will use ‘samples’ as a way of referring to both members of a kind for count nouns, e.g., each tiger counts as a sample of the kind tiger, as well as any amount of a kind for mass nouns, e.g., a puddle of water counts as a sample of the kind water.



superficially indistinguishable, but chemically distinct, from  $H_2O$ , call it XYZ—‘water’ would pick out something other than  $H_2O$ , namely, XYZ? Consider such a world at which XYZ is the predominant clear odorless liquid: is the extension of ‘water’ at that world  $H_2O$  or XYZ? Prima facie, it looks like these two intuitions suggest different answers to this question, with the former suggesting that the answer is  $H_2O$ , and the latter, that the answer is XYZ.

One attractive way of accommodating both intuitions is by using the two-dimensional framework. The two-dimensional framework offers the resources to capture these intuitions in such a way that we can see that they are not, in fact, paradoxical. In this chapter, I will discuss these two intuitions in more detail, introduce the general two-dimensional framework, and then illustrate how that framework can accommodate these two intuitions.

In chapters two and three, we will see how these intuitions are accommodated within specific interpretations of the two-dimensional framework. In chapter two, we will look at how these two intuitions are accommodated within Robert Stalnaker’s Metasemantic interpretation of the two-dimensional framework. In chapter three, we will look at how they are accommodated within the Epistemic interpretations of the two-dimensional framework as developed by Frank Jackson and David Chalmers.

But, on the views developed by Stalnaker, Jackson, and Chalmers, the two-dimensional framework is not merely employed to *accommodate* these two intuitions. These

intuitions are also themselves employed within the two-dimensional framework on these accounts to do further work within these views. We will see that the views developed by Stalnaker, Jackson, and Chalmers all take these two intuitions to play a role in modeling communicative exchanges, though their views differ with respect to precisely what work these intuitions do.

I raise a challenge that these intuitions cannot be used in this way to model communicative exchanges in all cases, and I illustrate this by drawing attention to a class of assertions that are such that both Stalnaker's metasemantic interpretation, and Jackson's and Chalmers's epistemic interpretations, are unable to explain how these assertions affect the state of the conversation. I argue that these problematic cases cast doubt on, if not undercut the motivation for, one of the following three claims: the claim that we *have* precisely these intuitions in the first place; the claim that they can be used for the purposes of modeling the way in which informative assertions change the state of a conversation; or the claim that they can be used *in the ways that Stalnaker, Jackson, or Chalmers employ them* for the purposes of modeling the way in which informative assertions change the state of the conversation in these problematic cases. I argue that it is the third claim that is challenged by the problem that I raise for their accounts, and I offer an extension to their accounts as a minimal modification that allows them to accommodate these problematic cases.

In chapter four, I present the general framework for what I call my three-dimensional solution to the problem that I raise for their accounts. My solution is proposed as an

*extension* to their accounts, rather than a completely distinct, competing approach. In chapter four, I present the three-dimensional framework in a general way—that is, without giving the framework an interpretation.

Just as with the two-dimensional framework, which can be given the metasemantic interpretation or the epistemic interpretation, the three-dimensional framework is in need of an interpretation. Depending on the nature of the interpretation, this interpretation is what makes it an extension of Stalnaker's view, or of Jackson's and Chalmers's views. In chapters five and six, I offer two interpretations of the three-dimensional framework that serve as extensions to Stalnaker's view, and to Jackson's and Chalmers's views, respectively. These proposed extensions make use of one of the two intuitions in a slightly different way that allows for Stalnaker's, Jackson's, and Chalmers's views to explain how assertions from the class that proved problematic for their views change the state of conversations. I conclude chapters five and six by discussing the consequences of adopting this extension to their respective views. In particular, we will see that adopting the proposed extension to Chalmers's view has somewhat deflationary consequences for some projects of significant interest to him. On the one hand, it offers *some* support to his claims about the a priority of some sentences having as constituents proper names or natural kind terms, and to the claim that conceivability of some such sentences is a good guide to their possibility. On the other hand, I argue that, on the extended view, we are not in a position to determine *how* good of a guide conceivability is to possibility.

## *Section 2: A Paradoxical Pair of Intuitions?*

We have two intuitions about proper names and natural kind terms that, on the face of it, appear paradoxical. On the one hand, we have the intuition that when we use a proper name or natural kind term, we intend to pick out the same individual, or the same kind, respectively, across worlds. On the other hand, we have the intuition that, had the world turned out differently, proper names and natural kind terms might have picked out a different individual or natural kind.

In this section, we will look at these two intuitions in more detail. One way to capture these intuitions is to think of them in terms of intuitions about the modal profile of expressions. The modal profile of an expression is the extension of an expression across worlds. We can capture this notion of the modal profile of these expressions by thinking of these expressions as *intensions*: as functions from worlds to extensions, such that, given a world, the expression returns the extension for the expression at that world. As we will see in section 2.1 and 2.2, these intuitions seem, at least *prima facie*, to offer inconsistent views about the modal profile of these expressions.

### *Section 2.1: The Uniform Extension Intuition*

Let us call the first intuition *the uniform extension intuition*. This is the intuition that, in using a proper name, we intend to pick out the same individual across worlds even when the individual doesn't have the properties that we associate with the individual's name, and, in using a natural kind term, we intend to pick out the same natural kind across

worlds, including those samples of that kind that do not have, at the time the expression is used, the property that we associate with that natural kind term.

For example, if one has the uniform extension intuition regarding proper names, one has the intuition that the modal profile of ‘Aristotle’ picks out the man we call Aristotle at all worlds at which it picks out anything at all.<sup>2</sup> This includes picking out that man at worlds at which he fails to satisfy any of the qualitative (e.g., having the sort of superficial appearance that we’ve come to expect from seeing a statue of him) and functional (e.g., being the last great philosopher of Antiquity, being the most famous student of Plato, or being the teacher of Alexander, etc.) characteristics that we associate with ‘Aristotle’.

This also means that ‘Aristotle’ fails to pick out anyone else, even if another individual has the qualitative and/or functional properties that we associate with ‘Aristotle’. Consider another world very much like our own, except that the man we call Aristotle is given a different name, grows up to be a politician instead of a philosopher, and that, due to a tragic wrestling accident, his appearance at that world differs from what we’ve come to associate with him. However, at that world, there is another person who the individuals at that world call Aristotle, and who looks just like we expect Aristotle to look, and grows up to do all the things that we associate with the name ‘Aristotle’: he is Plato’s most famous student, he is the teacher of Alexander, and, as things later turn out, he is the last great philosopher of Antiquity. According to the uniform extension intuition, our name, ‘Aristotle’, picks out the man *we* call Aristotle, and not the man *they* call Aristotle, even

---

<sup>2</sup> The ideas in this discussion of ‘Aristotle’, including the ‘Aristotle was fond of dogs’ example, are owed to Kripke’s discussion on pages 6-12 in *Naming and Necessity*. Please note that I have modified some of the details of the example, though these modifications do not affect the point the example is illustrating.

though the former fails to have the qualitative or functional characteristics that we associate with the name, and the latter has all those qualities.

Similarly, if one has the uniform extension intuition regarding the modal profile of 'water', one has the intuition that 'water' is such that it picks out the stuff we call water—H<sub>2</sub>O—at all worlds at which it picks out anything at all. This includes picking out H<sub>2</sub>O even under circumstances in which those samples of H<sub>2</sub>O fail to have the qualitative and functional characteristics that we associate with 'water'. For example, suppose that, under very high pressure at a very low temperature, H<sub>2</sub>O exhibits an appearance very much unlike the characteristics that we associate with it—even H<sub>2</sub>O under those conditions is in the extension of 'water'.

In addition, no other substance, even if it does have the qualitative and functional characteristics that we associate with 'water', is in the extension of 'water'. Consider a world in which XYZ has all the functional characteristics that we associate with 'water': it falls as rain; comes out of taps; forms rivers, lakes, oceans, and streams; is used to bathe; and so on. Nonetheless, according to the uniform extension intuition, XYZ is not in the extension of 'water' at that world. Only the H<sub>2</sub>O at that world, if there is any, is in the extension of 'water' at that world. There remains the question of whether 'water' is an empty term at worlds at which there is no H<sub>2</sub>O, but we won't worry about settling this matter here since it won't make a difference to our immediate discussion.

One way to see that we have the uniform extension intuition about the modal profile of proper names and natural kind terms is to consider our intuitions about the truth-conditions of sentences in which such expressions are a constituent. For example, consider Saul Kripke's example from *Naming and Necessity* of the sentence 'Aristotle was fond of dogs'.<sup>3</sup> Under what conditions is that true? It seems that we have a pretty strong intuition that it is true just in case *the individual that we call Aristotle*—and not merely anyone who happens to share some features that we associate with that individual—was fond of dogs. In other words, this strongly suggests that we do not take the reference of 'Aristotle' to be mediated by any qualitative or functional properties that we associate with him, such as the property of being the last great philosopher of Antiquity, or with any description that we associate with him that describes such properties, such as 'the last great philosopher of Antiquity'.

We can see this if we consider that, for example, though we might strongly associate with 'Aristotle' the property of being the last great philosopher of Antiquity, or with the description 'the last great philosopher of Antiquity', we would not say that 'Aristotle was fond of dogs' is true at a world just in case the individual who happened to be the last great philosopher of Antiquity, whoever he or she turned out to be, was fond of dogs. Rather, we would take that sentence to be true just in case *the man that we call Aristotle*—whatever he might be called at that world, and whatever properties he might have at that world—was fond of dogs.

---

<sup>3</sup>See the second edition of *Naming and Necessity* (second edition), p. 12-13.

To see this, consider, as we did earlier, a world at which the man that we refer to by using ‘Aristotle’ went into a different profession, and at which another individual was referred to by the name ‘Aristotle’ who had all of the qualitative and functional characteristics that we associate with ‘Aristotle’, but otherwise this world is exactly the same as our own, such that, at that world, the individual that they refer to using ‘Aristotle’ has the property of being the last great philosopher of Antiquity. Suppose that, at that world, the man we refer to using ‘Aristotle’ was not fond of dogs, but the man that they refer to using ‘Aristotle’ was fond of dogs. If the truth-conditions are such that the sentence is true if and only if the man we refer to using ‘Aristotle’ was fond of dogs, then that sentence would be false at that world.

Contrast this with the modal profile of ‘Aristotle’ if the reference of the name were mediated by the property or the description that we associate with the name. If the reference of ‘Aristotle’ is mediated by this property or description that we associate with the name, such that ‘Aristotle’ picks out at each world the individual that has that property at that world, then since the individual that has that property at that world—the man that they call Aristotle—was fond of dogs, then the sentence would be true.

A moment’s reflection reveals the intuition that the sentence is *not* true at that world. When we say that Aristotle is fond of dogs, we’re saying of that man—the one we call Aristotle—is fond of dogs. If he isn’t fond of dogs then the sentence is false. It doesn’t matter that another person that superficially resembles him, and has functional characteristics that we associate with ‘Aristotle’, is fond of dogs. We simply aren’t



talking about him in making such a statement. Thus, this strongly suggests that we have the intuition that ‘Aristotle’ refers to the same individual across all worlds at which it refers to anything at all.

Similarly, consider the sentence, ‘there is water in the oceans’. If, given that ‘water’ refers to  $H_2O$  in the actual world, ‘water’ refers to  $H_2O$  at all worlds at which it refers to anything at all, then its contribution to the truth-conditions of that sentence must be such that the sentence is true if and only if there is  $H_2O$  in the oceans. In contrast, if the reference of ‘water’ were mediated by, say, some qualitative or functional property that we associate with ‘water’, such as the property of being the predominant clear odorless liquid, then the contribution of ‘water’ to the truth-conditions would be such that the sentence is true just in case whatever has the property of being the predominant clear odorless liquid (at the world under consideration) is in the oceans. Would the sentence of our language, ‘there is water in the oceans’, be true as evaluated at such a world? If we have the intuition that we use ‘water’ to refer to all and only samples of the same kind, namely,  $H_2O$ —and it seems very plausible that we do have this intuition—then we will have the accompanying intuition that the sentence, evaluated at that world, is false.

Another way to motivate the claim that we have such an intuition is to consider the truth-conditions for the sentence, ‘water has three states: gas, liquid, and solid’, which one might find in an introductory chemistry text. It seems intuitive that what is relevant to the truth-conditions of this sentence is whether that substance to which ‘water’ refers— $H_2O$ —has three such states. To see this, suppose that XYZ also has three such states.

Evaluated at an XYZ world, are the truth-conditions of the sentence such that it is true just in case XYZ has three such states? Our intuitions strongly suggest that these are not the conditions under which the sentence is true at that world: that XYZ has three such states is irrelevant to the truth-conditions of the sentence, since the sentence is making a claim about that substance that *we* call water, and that substance is H<sub>2</sub>O.

Similarly, if we have the uniform extension intuition, we have the intuition that, in using ‘tiger’, we refer to the same kind—tigers—at all worlds at which we refer to anything at all.<sup>4</sup> We do not, by contrast, refer at each world to that which has some property that we associate with ‘tiger’, such as the property of being a large black-and-orange striped four-legged animal with whiskers. For example, according to the uniform extension intuition, at a world at which there are animals that are superficially indistinguishable from our tigers, yet are actually members of the reptile family, our word, ‘tiger’ would not refer to them.

Thus, if we have the uniform extension intuition regarding proper names, we take proper names to refer to the same individual across worlds, and if we have the uniform extension intuition about natural kind terms, we take natural kind terms to refer to samples (as in the case of mass nouns, such as in the case of ‘water’) or members (as in the case of count nouns, such as in the case of ‘tigers’).

### *Section 2.2: The Contingent Extension Intuition*

---

<sup>4</sup> This example is owed to Kripke. See his discussion of this example on page 122 of *Naming and Necessity*. See also Putnam (1975) for similar challenges for descriptivism.

Let us call the second intuition *the contingent extension intuition*.<sup>5</sup> This is the intuition that the extensions of our terms depend in some way on how the world turns out, and since the world might have turned out differently in the relevant respects, the extensions of our terms might have turned out differently, too.

For example, this general intuition seems to underlie more specific intuitions such as even though this is a world at which H<sub>2</sub>O is the predominant clear odorless liquid that constitutes the rivers, lakes, and oceans; comes out of taps; is used to bathe; quenches thirst; and so on, had an XYZ world been actual instead, ‘water’ would have referred to XYZ rather than H<sub>2</sub>O. Thus, it seems that we also have an intuition about the modal profile of ‘water’ that is such that had the world turned out differently, the extension of our expressions would have been different, too.

There is an issue about whether to think of this notion of considering hypotheses about how the extensions of our terms might turn out on various hypotheses about which world

---

<sup>5</sup> Kripke discusses the notion of an epistemic counterpart of oneself having a superficially indistinguishable experience from the one has in, say, observing a desk. Suppose that one walks into a room and sees an unfamiliar desk, and that one thinks to oneself that it might be that the desk is made of wood, or it might be that the desk is made of ice cleverly formed (and, presumably, colored) to look like wood. There is an epistemic sense of turns out, along the lines of *contrary to what we currently think*, on which the table could turn out to be made of ice. The thought is that an epistemic counterpart of yourself—an individual in a superficially indistinguishable situation from your own—could be in a situation in which the desk *is* made of ice. Since the two of you are, prior to further investigation of the desk, in the same epistemic state with respect to the desk, such that you aren’t in a position to know in which of these two situations you find yourself, this is a situation in which, *for all you know*, the desk could turn out to be made of wood, or of ice. This epistemic notion of ‘turns out’ is importantly different from the metaphysical sense of ‘turns out’. Given that the table is, in fact, made from wood, it could not turn out that that very desk could have been made of anything other than that very wood. (See page 141 of *Naming and Necessity*.) It is the epistemic sense of ‘turns out’ that is relevant here for capturing the contingent extension intuition.

is actual, as a subjunctive or as an indicative.<sup>6</sup> The subjunctive suggests a more metaphysical reading of the intuition: it could have been that another world was actual, and if another world was actual, then the extensions of our expressions might have differed. We can think of this as saying that, though, in fact, this world is actual, it might have been that another world was actual, and if another world was actual, the extensions of our terms would have differed accordingly. For example, for some XYZ world,  $v$ , it might have been that  $v$  was actual, and if  $v$  were actual, then the extension of ‘water’ would have been XYZ.

The indicative suggests a more epistemic reading: for all one knows, the world could turn out to be the way that  $v$  is, and if the actual world is the way  $v$  is, then the extension of our natural kind term, ‘water’, is XYZ, but, for all one knows, the world could also turn out to be the way some H<sub>2</sub>O world,  $w$ , is, and if the actual world is the way that  $w$  is, then the extension of ‘water’ is H<sub>2</sub>O. On the epistemic understanding, we either genuinely do not know how the world turns out in the relevant respect, or we bracket that knowledge for the purposes of entertaining alternative hypotheses about the extension of our expressions.

Perhaps these two ways of thinking about it are equivalent, but since this is a matter of contention, we will continue to regard them as two ways of thinking about it, and leave it as an open question whether they are equivalent. Regardless of the relationship between the two, it is the epistemic reading—the indicative version—that is of interest to us in our

---

<sup>6</sup> For a discussion of this by David Chalmers, see his “The Tyranny of the Subjunctive.” Stalnaker discusses this in “Assertion Revisited” (though he doesn’t explicitly talk about the subjunctive vs. the indicative).

project. As we will see later in this chapter, and again in chapters two and three, it is on the epistemic understanding of the contingent extension intuition that it is used to model the state of the conversation, and the way that assertions change the state of the conversation.

### *Section 3: Accounting for both Intuitions Using the Two-Dimensional Framework*

On the face of it, when considered together, these two intuitions, indeed, seem paradoxical. To see this, consider again the question about the extension of ‘water’ at an XYZ world. According to the first intuition, the extension of ‘water’ at the XYZ world is the same as it is at the actual world and all other worlds: ‘water’ picks out H<sub>2</sub>O at all worlds at which it picks out anything at all, and so XYZ would not be in the extension of ‘water’ at that world. According to the second intuition, had that XYZ world been actual, ‘water’ would have referred to XYZ, not H<sub>2</sub>O, and so XYZ would have been in the extension of ‘water’. How can we make sense of these two independently plausible but at least prima facie inconsistent intuitions?

#### *Section 3.1: The Two-Dimensional Matrix for Proper Names and Natural Kind Terms*

One way of attempting to account for both of these intuitions is by making use of the two-dimensional framework.<sup>7</sup> In this section, we will introduce the general two-

---

<sup>7</sup> Indeed, Garcia-Carpintero and Macia note in their introduction to their anthology on Two-Dimensional Semantics that Kripke himself lays out a blueprint for two-dimensional semantics in *Naming and Necessity* (see p.2 of their Introduction). For other examples of using the two-dimensional framework to accommodate these two intuitions about proper names and natural kind terms, see Gareth Evans’s “Reference and Contingency”, Davies, M. and Humberstone, I.L. “Two Notions of Necessity”, in addition to the work of Robert Stalnaker, Frank Jackson, and David Chalmers that we will discuss in this project.

dimensional framework that accommodates these two intuitions, and then, in subsequent chapters, we will discuss different interpretations of this general framework.

We will begin by seeing how to account for the uniform extension intuition. We can represent the uniform extension intuition by using a matrix that has one row for the actual world, and one column for each possible world under consideration, including a column for the actual world. We understand the world in the row—that is, the actual world—as the world at which the reference of the expression is fixed. We then ask, of the world in each column, given what the expression picks out in the actual world, what does the expression pick out at the world in that column (provided that it picks out anything at all)? The cell in the matrix that is the intersection of the row marked with the actual world and the column of the world at which we are considering the extension of the expression is marked with the individual or kind which is in the extension of the expression at that world.

For example, consider ‘water’. At the actual world, ‘water’ refers to  $H_2O$ . Suppose that the worlds under consideration in the columns are the actual world (an  $H_2O$  world), and an XYZ world. To account for the uniform extension intuition, at each world under consideration, ‘water’ picks out  $H_2O$ . As we can see in the matrix represented in figure 1, below, given that the actual world is an  $H_2O$  world (as represented by the  $H_2O$  world row), then, at both the  $H_2O$  world and the XYZ world (represented by the intersection of the  $H_2O$  column and the  $H_2O$  row, and the XYZ column and  $H_2O$  row, respectively),

‘water’ refers to H<sub>2</sub>O. This captures our intuition that ‘water’ refers to the same kind across all worlds at which it refers to anything at all.

Figure 1

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>
<b>H<sub>2</sub>O</b>	H <sub>2</sub> O	H <sub>2</sub> O

Next, to account for the contingent extension intuition, we consider that the extension of our expressions depend in part on the way the world turns out, on the epistemic understanding of ‘turns out’. That is, we can consider various hypotheses about the extensions of our terms based on various hypotheses about how the actual world turns out, on the epistemic understanding of ‘turns out’.

For example, consider the epistemic state of chemists who are researching the underlying chemical constitution of water circa 1800. Suppose that their competing hypotheses are that water is H<sub>2</sub>O, and that water is XYZ—these are two ways that the world could turn out, for all they know. These scientists can entertain the hypothesis that our world turns out to be such that the predominant clear odorless liquid turns out to be H<sub>2</sub>O, in which case ‘water’ refers to H<sub>2</sub>O. They can also entertain the hypothesis that the predominant clear odorless liquid turns out to be XYZ, in which case ‘water’ refers to XYZ.

More needs to be said about *which* worlds are candidates for being considered on the hypothesis that they could have turned out to be actual. I take it that the uniform extension intuition is not the more general intuition that for any expression, that

expression could have been assigned a different extension, such that, for example, ‘water’ could have been assigned as the name for the number that we call seven. Rather it is a more specific intuition regarding how the extensions of proper names and natural kind terms might have been different had *a different world that is superficially like our own in at least some relevant respects* been actual, **and** that the facts regarding how the reference of the expression is fixed are the same, or are suitably similar, in a way to be qualified in chapters two and three. For present purposes, we will only concern ourselves with holding fixed the facts about how the reference of a given proper name or natural kind term was fixed, and then, in chapters two and three, we will consider an account that allows for the facts about how the reference of a proper name or natural kind term is fixed, on varying hypotheses about the way the world turns out, to differ from the facts of how the reference of the expressions of our language actually were fixed. The important point at this time is that our qualification has constrained the space of worlds that can be considered as actual to those that are (a) superficially indistinguishable from our world in the relevant respects, and (b) the facts about how the reference of the proper names and natural kind terms are held fixed for the purposes of evaluating how the extensions of our terms might turn out on varying hypothesis about how the world might turn out.

There are two general approaches to thinking about how the reference of an expression is fixed that we will consider in this project. The first is that the reference of a token of an expression is mediated in some way, perhaps by virtue of a description or property that we associate with that expression, such that a given token of the expression refers to



whatever satisfies the relevant description or has the relevant property at the time the token is produced.

As a first pass at capturing this view, let us suppose, without further qualification, that the description that is associated<sup>8</sup> with a proper name or natural kind term mediates the reference of that expression. For example, suppose that ‘water’ is associated with the property of being the predominant clear odorless liquid. Suppose that a speaker makes an assertion that has as a constituent expression ‘water’. On this view of the property mediating the reference of the expression, that token of ‘water’ refers to H<sub>2</sub>O by virtue of H<sub>2</sub>O having that property, i.e., by virtue of H<sub>2</sub>O being the predominant clear odorless liquid. In this case, the property of being the predominant clear odorless liquid is mediating the reference of ‘water’, such that ‘water’ refers to whatever has that property.

However, this cannot be the whole story about reference on such an account, if such an account is intended to also accommodate the uniform extension intuition. This is because, without further qualification, this allows for the modal profile of ‘water’ to be such that it picks out at each world whatever has that property at that world. For example, at a world at which the predominant clear odorless liquid is XYZ—call such a world an XYZ world—XYZ has the property, and so, for all that we’ve said about this manner of fixing the reference, ‘water’ would pick out XYZ rather than H<sub>2</sub>O at that world, thus failing to be in accordance with the uniform extension intuition. In order to

---

<sup>8</sup> Here, I’m being deliberately vague regarding *what party* is doing the associating. Is it the property that an *individual* associates with an expression, or the property that a *community of language users* associates with an expression? The former will tie the process of reference-fixing more closely to the epistemic state of the individual, and the latter will tie the process of reference-fixing more closely to the conventions of the language. We will discuss this in more detail in chapter three.

accommodate this intuition, we need to modify the account such that the modal profile picks out the same individual or kind across worlds.

One way of doing this is to modify the description or property such that it is actually-involving. For example, rather than associating the property *the predominant clear odorless liquid* with ‘water’, we associate the property *the **actual** predominant clear odorless liquid*. These two properties have different modal profiles: the former picks out at each world whatever has the property at the world under consideration, while the latter picks out at each world that which has the property at the actual world. Associating *the actual predominant clear odorless liquid* with ‘water’ makes it the case that, at each world, ‘water’ picks out whatever kind has the property of being the predominant clear odorless liquid at the actual world, and thereby picks out H<sub>2</sub>O at all worlds.

Similarly, one can do this for a view on which descriptions mediate the reference of expressions, too. Instead of associating the description ‘the predominant clear odorless liquid’ with ‘water’, if the description associated with ‘water’ is ‘the *actual* predominant clear odorless liquid’, then ‘water’ will pick out H<sub>2</sub>O at all worlds at which it picks out anything at all

We can represent these hypotheses about how the reference of expressions depends on the way the world turns out within the same matrix. We allow the worlds on the left side to represent each world that we will considering on the hypothesis that that world is

actual. Each row represents the considering of that world on the hypothesis that it turns out to be actual, as in figure 2:

Figure 2

<b>H<sub>2</sub>O</b>		
<b>XYZ</b>		

As in figure 1, we then add worlds to the top of each column. The world in each column stands for being the world at which the expression is evaluated, on the hypothesis that the world in the row is actual. That is, the worlds for each column are counterfactual worlds, given that the world in the row is the actual world. For example, consider the H<sub>2</sub>O row in figure 2. At the intersection of the H<sub>2</sub>O world and the XYZ column, the idea is that we are evaluating whether what is said at the H<sub>2</sub>O world (on the hypothesis that it is the actual world) is true at the counterfactual XYZ world.

We can then fill in the other cells of our matrix. We will begin by filling in the cell that is the intersection of the H<sub>2</sub>O row and H<sub>2</sub>O column. As before, the world in the column under consideration—here, the H<sub>2</sub>O world—is the world at which we are asking about the extension of the expression, ‘water’, *given that the world in that row is actual*. Since the cell is in the H<sub>2</sub>O row, we are considering the extension of ‘water’ at the H<sub>2</sub>O world, on the hypothesis that the H<sub>2</sub>O world is actual, and so the value of the cell is the answer to the question: if an H<sub>2</sub>O world turns out to be actual, then what is the extension of

‘water’ at that H<sub>2</sub>O world? The answer is H<sub>2</sub>O, and so we fill in the cell accordingly. We can then repeat the same process for the intersection of the XYZ world row and XYZ world column. The result of filling in both of these cells can be seen in figure 3.

Figure 3

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>
<b>H<sub>2</sub>O</b>	H <sub>2</sub> O	
<b>XYZ</b>		XYZ

To fill in the remaining cells in our two-dimensional matrix for ‘water’, we will consider the extension of ‘water’ at the XYZ world, given that the H<sub>2</sub>O world is actual, and we will consider the extension of ‘water’ at the H<sub>2</sub>O world, given that the XYZ world is actual. Given our commitment to the uniform extension intuition, if our world turns out to be an H<sub>2</sub>O world, and so ‘water’ refers to H<sub>2</sub>O at our world, ‘water’ will refer to H<sub>2</sub>O at all worlds at which it refers to anything at all. This means that, at an XYZ world, ‘water’ picks out H<sub>2</sub>O, and so we can mark the intersection of the H<sub>2</sub>O row and the XYZ column with H<sub>2</sub>O. Let us call two-dimensional matrices of this kind *standard matrices*.<sup>9</sup>

Similarly, if our world turns out to be an XYZ world, then ‘water’ refers to XYZ, not only at the XYZ world, but also at the H<sub>2</sub>O world. This allows us to fill in the XYZ row such that, at each of these worlds, ‘water’ refers to XYZ. This results in the completed matrix that we can see in figure 4.

---

<sup>9</sup> I am giving them a name here because I will introduce a different kind of two-dimensional matrix in chapter four.

Figure 4

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>
<b>H<sub>2</sub>O</b>	H <sub>2</sub> O	H <sub>2</sub> O
<b>XYZ</b>	XYZ	XYZ

Now we are in a position to see how the matrix captures both intuitions. We can see that it captures the contingent extension intuition: the extension of ‘water’ depends on the way the actual world turns out, such that if the actual world is an H<sub>2</sub>O world, then ‘water’ refers to H<sub>2</sub>O, and if the actual world is an XYZ world, ‘water’ refers to XYZ, and so on. We can also see that it captures the uniform extension intuition because, whatever kind ‘water’ *does* refer to, it refers to the same kind across worlds.

Furthermore, this allows us to see a distinction between two different ways of evaluating the extension of ‘water’ at a given world: we can evaluate the extension at a world considered as counterfactual, or we can consider the extension at a world considered as counterfactual. We consider a world,  $\alpha$ , as counterfactual when we hold fixed our world as the one that is actual, such that the extension of ‘water’ is dependent on the way this world turns out, and then we ask what the extension is at counterfactual world  $\alpha$ . We consider  $\alpha$  as actual when we consider that  $\alpha$  is the way our world turns out—that is, we consider that our world is  $\alpha$ —and then we consider what is the extension of ‘water’ given that our world is  $\alpha$ .

This distinction between counteractual and counterfactual evaluation allows us to answer the question about the extension of ‘water’ at an XYZ world to which, initially, it looked as though the contingent extension and uniform extension intuitions gave inconsistent answers. Considered as counterfactual, the extension of ‘water’ at the XYZ world is H<sub>2</sub>O. Considered as actual, the extension of ‘water’ at the XYZ world is XYZ. This distinction resolves the tension by showing that the question needed further qualification—whether the XYZ world is considered as actual or counterfactual—such that there are two questions rather than one, and we see that the answers to the two questions aren’t inconsistent.

*Section 3.2: The Two-Dimensional Matrix for Assertions*

Just as we can evaluate the extensions of proper names and natural kind terms either counteractually or counterfactually, we can also evaluate the truth-value of sentences counteractually or counterfactually. Consider the sentence, ‘there is water in the oceans’, evaluated using an H<sub>2</sub>O world and an XYZ world, as we can see in figure 5.

Figure 5

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>
<b>H<sub>2</sub>O</b>	<b>T</b>	<b>F</b>
<b>XYZ</b>	<b>F</b>	<b>T</b>

First, let's look at the truth-value for the cell that is at the intersection of the H<sub>2</sub>O row and the H<sub>2</sub>O column. Since we are considering the H<sub>2</sub>O as actual, we are determining what is said by an assertion of 'there is water in the oceans' at the H<sub>2</sub>O world. Since, as we've discussed, if the H<sub>2</sub>O world is actual then 'water' refers to H<sub>2</sub>O, and so the sentence is true just in case there is H<sub>2</sub>O in the oceans. Since we are in the H<sub>2</sub>O column, we are treating the H<sub>2</sub>O world as the circumstance in which we evaluate the assertion—is what the assertion says true at the H<sub>2</sub>O world? Since, as asserted at the H<sub>2</sub>O world, what is said is true just in case there is H<sub>2</sub>O in the oceans, and the H<sub>2</sub>O world is such that there is H<sub>2</sub>O in the oceans, then what is said at the H<sub>2</sub>O world is true at the H<sub>2</sub>O world, and so the cell that is at the intersection of the H<sub>2</sub>O world row and the H<sub>2</sub>O world column is marked with a T for 'true'.

Moving one cell to the right, we can ask whether what was said at the H<sub>2</sub>O world—that there is H<sub>2</sub>O in the oceans—is true at the XYZ world. Since the XYZ world is such that there is XYZ in the oceans, rather than H<sub>2</sub>O, then what is said at the H<sub>2</sub>O world is false at the XYZ world. For this reason, the intersection of the H<sub>2</sub>O row and the XYZ column is marked with an F for 'false'.

Moving one cell down, we can then ask whether what is said at the XYZ world is true at the XYZ world. Since what is said at the XYZ world is that XYZ is in the oceans, then what is said at the XYZ world is true at the XYZ world just in case there is XYZ in the oceans. Since the XYZ world is such that there is XYZ in the oceans, then what is said at the XYZ world is true at the XYZ world, and so that cell is marked with a T.

Filling in the last cell, at the intersection of the XYZ row and the H<sub>2</sub>O column, we ask whether what is said at the XYZ world is true at the H<sub>2</sub>O world. Since what is said at the XYZ world is, as we've said, that there is XYZ in the oceans, and the H<sub>2</sub>O world is such that there is H<sub>2</sub>O, not XYZ, in the oceans, then what is said at the XYZ world is false at the H<sub>2</sub>O world. For this reason, that cell is marked with an F.

In the next section, we will discuss how the matrix for an assertion is used to model the force of that assertion.

#### ***Section 4: Using the two-dimensional framework to model conversations***

For Stalnaker, Jackson, and Chalmers, the two-dimensional framework is not merely a means of *accommodating* these two intuitions: the two intuitions are employed for the purposes of modeling communicative exchanges. In this section, we will look at features that are shared by the approaches to modeling communicative exchanges within the views defended by Stalnaker, Jackson, and Chalmers. In chapter two, we will look at the details of Stalnaker's approach, and in chapter three we will look at the details of Jackson's and Chalmers's approaches.

##### ***Section 4.1: What Kind of Communicative Exchanges?***

Because there are many kinds of speech acts, and we have conversations for many reasons, not all of which are relevant to present purposes, we need to be clear about what



kind of conversation is of interest to our project. In this project, we will focus on the speech act of making an assertion that is intended to be taken literally (and is such that the audience of the assertion can reasonably be expected to understand that the assertion is intended to be taken literally), within a conversation in which the common goal of the interlocutors is to narrow down the space of ways the world could turn out to be, for all they know, in some relevant respect.

An intuitive way to think about participating this kind of conversation is that the interlocutors are engaged in a collaborative effort to make progress on figuring out how the world might be in the relevant respect. This proceeds by both ruling out some possibilities that are still ways the world might turn out, for all they know, and by adding possibilities for ways the world might be that aren't presently under consideration. At a given point in such a conversation, there are still some facts about the world that at least some of the interlocutors do not know—which we will represent as ways this world could turn out to be, for all they know—and so the purpose of the conversation is to rule out at least some of these possibilities by making assertions that exclude from further consideration those worlds that aren't the way this world is being said to be. Furthermore, sometimes a new possibility is introduced that wasn't previously under consideration but, if the interlocutors accept the assertion that says that the world is that way, that possibility is subsequently added to the ways the world might turn out to be, for all they know.

In such a conversation, each participant presupposes that their fellow interlocutors know some things about the way the world is—that is, each participant presupposes some common knowledge about the world among the interlocutors—and also presupposes that there are some things that their fellow interlocutors do not know about the way the world is that are relevant to the topic of the conversation.

For example, in a conversation between bakers about how to make the best sourdough bread, these interlocutors—the bakers—presuppose some common knowledge relevant to the topic of conversation, such as presupposing that the other bakers know what flour is, and that it is needed to bake bread, and that the bread is baked in an oven, etc. Each baker also presupposes that there are some things that their fellow interlocutors may not know that is relevant to the conversation. Suppose that they are discussing how to make the best crust. Each baker can presuppose that his fellow interlocutors aren't sure which of the methods under discussion will result in the best crust, and so each baker understands that it would be appropriate to make assertions that give reason to think that one method is better than the others. If one of the bakers has had what he takes to be very good success with brushing the loaf lightly with olive oil before baking, resulting in a firm, golden crust, and that this seemed to work better than brushing the loaf with butter, or not coating the loaf with anything, then it would be appropriate for him to say so.

Furthermore, each baker understands that it would be inappropriate to make an assertion that either says the world is the way that the world is already presupposed to be, or to

make an assertion that excludes a possibility that isn't under consideration because it is commonly understood not to be the case. Since it is presupposed that all methods under consideration involve baking the bread in the oven, it would be inappropriate in this context for one of the bakers to suggest that the best method involves baking the bread in the oven. Similarly, it would be inappropriate for a baker to suggest trying a recipe without flour, since all the methods under consideration involve using a recipe that requires flour.

We can model the state of the conversation as the set of worlds that are all consistent with what can reasonably be presupposed by the interlocutors to be common knowledge among those engaged in the conversation, and that the worlds in this set are such that, for each hypothesis that the interlocutors can reasonably presuppose to be under consideration in the conversation regarding the way the world might be, there is at least one world that is the way that hypothesis represents the world to be.

Let us, following Stalnaker, call this set of worlds *the context set*.<sup>10</sup> For any assertion, let us say that there is a *prior context set*, or the set of worlds that captures the state of the conversation prior to the assertion, and that there is a *posterior context set*, or the set of worlds that captures the state of the conversation as it has been affected by the assertion.

In making an assertion in such a conversation, the speaker is proposing that the context set is modified by doing at least one of two things. One thing she might be doing is proposing to her fellow interlocutors that any worlds under consideration that aren't the

---

<sup>10</sup> See *Assertion*.

way her assertion says the world is should be ruled out as no longer being candidates for ways this world might turn out. Another thing she might be doing is proposing to her fellow interlocutors that there is a further hypothesis that should be under consideration, namely that the world is, or might be, the way that her assertions says the world is, that should be added to the context set (to the exclusion of other hypotheses, if her claim is that the world *is*, rather than *might be*, the way the assertion says the world is). This serves as a proposal to add to the context set those worlds that are both consistent with the presuppositions *and* are the way her assertion says the world is. Both of these assertions serve as proposals to change the context set: the former does so by proposing the exclusion of at least one world from the prior context set, and the latter does so proposing the inclusion of at least one world not already in the context set. Let us say that the *force* of an assertion is the way the assertion affects the context set, which, as we've said, might be to rule out worlds, or to add worlds to the context set.

For some examples to illustrate how an assertion changes the state of the conversation in these ways, consider the following scenario. Suppose that a fifth-grade class is engaged in an exercise in Australian geography, and that one of the interlocutors—call him Bruce—as the teacher of the fifth-grade class, is more knowledgeable about Australia than his students. Suppose that the current task is to determine which Australian city is the capital. The students believe that the major cities in Australia are Sydney, Perth, Brisbane, Melbourne, and Adelaide; and they know where these cities are located within Australia; and they know that the capital of Australia is a major city, but they aren't sure

which major city it is. Furthermore, suppose that the students are in a position to presuppose that their classmates share these beliefs.

We model the initial state of this conversation in which they are trying to pare down the space of ways the world could turn out, for all they believe, with respect to the capital of Australia, using a set of worlds that is as follows. All worlds in the set are such that the major cities of Australia are Sydney, Adelaide, Melbourne, Brisbane, and Perth—these worlds are all the same in that respect, and each of these cities is all located in the same (correct) location at each world, since the interlocutors know where these cities are located. The respect in which the worlds differ is that there is at least one world in the set that is such that the capital is Sydney, there is at least one world in the set that is such that Melbourne is the capital, and so on.

Of course, unlike his students, Bruce knows which city is the capital, but he wants to help the students to figure out the capital for themselves, rather than telling them directly. As an example of an assertion that serves as a proposal to rule out worlds, Bruce offers them a hint to help them to figure out the capital of Australia by asserting, ‘the capital of Australia is not located in Western Australia or South Australia.’ In making this assertion, he is proposing that his fellow interlocutors—in this case, his students—rule out the worlds from the context set at which the capital of Australia is in Western Australia or South Australia, namely the world(s) at which Perth is the capital, and the world(s) at which Adelaide is the capital, respectively. Given that his students know which major cities are located in Western Australia and South Australia—Perth and

Adelaide—and they have reason to think that Bruce knows what he is talking about (suppose that they know that he was born and raised in Australia, in addition to the authority that comes with being the teacher), they accept his proposal and thereby rule out those worlds.

For an example of adding a hypothesis that was not already under consideration, suppose that, through further conversation with his students, it becomes apparent to Bruce that the class isn't aware of the fact that Canberra is a major city. If Bruce asserts, 'Canberra is a major Australia city', this serves as a proposal to include the hypothesis that Canberra is the capital of Australia, since they all know that the capital of Australia is a major Australian city. This assertion also serves as a proposal to rule out any worlds at which Canberra is not among the major Australian cities. Recall that we said that the set of worlds initially was such that, at each world, the major Australian cities are Sydney, Brisbane, Perth, Melbourne, and Adelaide. These worlds will be excluded because they aren't the way that Bruce says the world is (since they don't have Canberra as among the major Australian cities), and added to the context set are worlds at which the major Australian cities are Sydney, Brisbane, Perth, Melbourne, Adelaide, and Canberra, among which at least one world is such that Canberra is the capital of Australia, though none at which either Perth or Adelaide are the capital, since those possibilities have already been excluded.

In cases in which the interlocutors know what the expressions refer to, and so know what proposition was expressed, the worlds that are ruled out are those that are false on *the*

*horizontal*—that is, false in the row at which the world considered as actual is the actual world (i.e., our world), as shown in figure 6, below. For example, consider again the assertion, ‘the capital of Australia is not located in Western Australia or South Australia’. We can represent this in a matrix, as in figure 6.

Figure 6

	Canberra	Sydney	Melbourne	Perth	Adelaide	Brisbane
Canberra	<b>T</b>	<b>T</b>	<b>T</b>	<b>F</b>	<b>F</b>	<b>T</b>

Each world at the top of the matrix is a world at which one of the cities under consideration is the capital: the first world is a world at which Canberra is the capital, the second is a world at which Sydney is the capital, the third is a world at which Melbourne is the capital, and so on. The world in the row is the actual world, and so Canberra is the capital at that world. Consider the intersection of the Canberra world row and the Canberra world column: is what is said at the Canberra world true at the Canberra world? Yes, since at the Canberra world, the capital is Canberra, and Canberra is not in either Western Australia or South Australia. Moving to the intersection of the Canberra world row and the Sydney world row, what is said at the Canberra world is true at the Sydney world? Yes, because at the Sydney world, Sydney is the capital of Australia, and Sydney is not in Western Australia or South Australia.<sup>11</sup> Consider the intersection of the Canberra world row and the Perth world column: is what is said at the Canberra world true at the Perth world? No, since at that world, Perth is the capital, and Perth is in

<sup>11</sup> Note that ‘the capital of Australia’ is a non-rigid description, and so does not pick out Canberra at all worlds. Rather it picks out, at each world, the city that is the capital of Australia at that world, provided that there is one, and is otherwise empty.

Western Australia, so at that world, what is said at the Canberra world is false at the Perth world, and so the Perth world is ruled out by the assertion because it isn't the way the assertion says the world is. The Adelaide world is ruled out, too, because Adelaide is in South Australia.

#### *Section 4.1: The Norms of Assertion*

One reason why being able to model the state of conversations, including how they are affected by assertions, is valuable is that, together with norms of assertion, it explains whether a given assertion is (or would be) appropriate in a particular context. That is to say, there are certain norms of assertion that govern whether we ought to make a given assertion in a particular context, and whether we ought to do so is determined in part by the state of the conversation in that context.

The important norms that govern what we ought, and ought not, to say in such conversations include the Gricean Maxim of Quality<sup>12</sup> that we ought to say what is true (or, at least, what we have good reason to believe to be true), and the Gricean Maxim of Quantity that we ought to say things that are informative, though only as informative as is appropriate to the conversation. Let us call the first the *truth* norm, and the second the *informative* norm. These are the two most important norms for our purposes in this project. There are other relevant norms, too, such as the Gricean Maxim of Relation, which requires interlocutors to say only things that are relevant to the conversation, and the Gricean Maxim of Manner, which requires the interlocutors to say things in a clear, unambiguous way, such that their fellow interlocutors can understand what is said. The

---

<sup>12</sup> All of these norms can be found in Grice, "Logic and Conversation".



widespread acceptance of these norms is a precondition for successfully having these kinds of conversations.

The truth norm requires that we only make assertions that we have good reason<sup>13</sup> to believe to be true. We can see that the acceptance of this norm is an important precondition for having such conversations because the central aim of such conversations is for the interlocutors to make progress on learning about the way the world is. If the assertions are not such that they aid in that project—that is, if they do not serve to tell the interlocutors how the world is, or how it is not—then such a conversation cannot be had.

Thus, in such a conversation, the interlocutors must both adopt this norm, and assume that their fellow interlocutors have adopted it, as well. This means that the interlocutors are in a position to take each assertion as a proposal to rule out those worlds that aren't the way the assertion says the world is, or to include those worlds that are the way the world is (but were not previously under consideration), provided that they have sufficient reason to believe that the speaker knows what she is talking about. That is to say, even though the interlocutors may reasonably assume that the speaker, like the other participants in the conversation, has adopted the truth norm, it doesn't follow from this

---

<sup>13</sup> *How* good the reason(s) must be may vary by context, especially how important it is that your assertion is true. If I am trying to find my wallet so that I can show you the unflattering photo on my university ID card, and you have a faint memory that I'd left it on the kitchen counter by the refrigerator, that may count as a good enough reason to assert, 'it is on the kitchen counter by the refrigerator'. This is because the stakes aren't terribly high if you aren't right about that. At most, you risk me being slightly annoyed at going on a fruitless errand to the kitchen, and missing out on the opportunity to ridicule me about my photo. However, if the stakes are very high—say, that the house is on fire, and I need to take my wallet with me because it is my only personal identification, then your faint memory may not be a sufficiently good reason. Finding the wallet is time-sensitive, and it is very important that I have my personal identification, so it is very important that I do not spend that valuable time looking around in places unless there is very good reason to think it might be there.

alone that the interlocutors ought to thereby trust what the speaker says. The interlocutors still need reason to think that what the speaker believes to be true, and, for that reason, asserts, is true. That is, the assertion itself serves as evidence that the *speaker* believes that the world is, or is not, the way the assertion says the world is, but, in order for the interlocutors to accept the assertion, and update the relevant possibilities accordingly, the interlocutors must believe that the speaker knows what she is talking about. If the interlocutors have reason to think that the speaker is sufficiently knowledgeable about the topic, they will accept the assertion and, if not, they won't accept it.

The informative norm requires that what one says must rule out some possibilities, introduce some possibilities not under consideration, or both. This norm particularly highlights the importance of being able to model the state of the conversation because modeling the prior context set allows us to capture the possibilities under consideration in the state of the conversation immediately preceding the assertion, thus allowing us to determine whether a given assertion would be informative. Being aware of which possibilities are under consideration allows the speaker to be aware that she should not say something that is true in all of the worlds because that would not be informative. It also allows her to be aware that if she says something that is true in some but not all of those worlds, or that is true in none of those worlds, her assertion will be informative, and so she will have acted in accordance with the norm.

The interlocutors' acceptance of these two norms for the purposes of having these kinds of conversations, together with our ability to (at least tacitly) have a grasp of the state of the conversation, as well as being able to anticipate how making various assertions would affect that conversation, are preconditions for them to be able to engage in such conversations. The theorist's ability to model such exchanges allows her to explain why a given assertion is informative, as well as why it is accepted by the interlocutors, and, thereby, appropriate.<sup>14</sup> These points will be important in motivating why we should be concerned to provide a solution to the problem that I raise for Stalnaker in chapter two, and for Jackson and Chalmers in chapter three. We will return to this point at that time.

#### *Section 4.2: Introducing the Problem of Informativity*

As we've been discussing, we make assertions to inform others about the way the world is, and we update our beliefs about the way the world is in light of assertions made by others. This kind of exchange is so familiar, and appears to be so straightforward, that it hardly seems problematic. Yet, when we try to model the impact of some assertions on the state of a conversation, we encounter some difficulties.

---

<sup>14</sup> I say here that it would be appropriate on the grounds that it accords with both the truth norm and the informative norm. However, it seems that, in determining whether an assertion is appropriate, a norm of relevance may also be applicable: one ought to only make assertions that are relevant to the conversation. Because we are modeling conversations using a particularly coarse-grained means—worlds—it is difficult to capture how to go about following the norm of relevance by looking at set of worlds in the prior context set. However, if the aim of the conversation is to get closer to determining the way the world is *in some respect*, the interlocutors will judge assertions (even informative ones) as inappropriate if they aren't in the service of this more specific task. For example, if the interlocutors are trying to figure out which major Australian city is the capital, and one participant makes an assertion that is informative (because true at some but not all worlds in the prior context set) but irrelevant (because concerned with, say, the price of tea in China), the interlocutors will not find it appropriate. Furthermore, if the interlocutors do not follow the norm of relevance more generally, it may well lead the conversation to devolve into a series of unrelated assertions.

First, we face difficulties in explaining why some assertions are informative at all. We encounter difficulty in explaining why some assertions are informative because our commitment to the uniform extension intuition makes it the case that, using the two-dimensional framework to model the force of assertions, we would expect that some assertions—most notably the necessary a posteriori ones—would not be informative, but these assertions clearly are informative.

The problem with necessary a posteriori assertions is that, on the one hand, they are necessary, meaning that they are true at all worlds. If they are true at all worlds, this means that there is no world at which they are false. However, these assertions are also informative, and if the fact that an assertion is informative is to be explained by virtue of that assertion ruling out at least one world that isn't the way the assertion says the world is, then this is puzzling, since there *isn't* any world that isn't the way the assertion says the world is. Thus, there is a problem in being able to explain how such an assertion could be informative.

For example, suppose that someone asserts, 'water is H<sub>2</sub>O', as said in a conversation in which the interlocutors did not know this fact. The uniform extension intuition suggests that the modal profile of 'water' is such that it picks out the same extension at all worlds at which it picks out anything at all, and so this assertion is necessarily true. Since it is necessarily true, there are no worlds that *aren't* the way the assertion says the world is, and since assertions rule out those worlds that aren't the way the assertion says the world is, we would expect this assertion to not rule out any worlds, and so to not be informative.

Yet, to those who do not yet know this fact, the assertion is informative, which suggests that the assertion *does* rule out at least one world. How can we explain that this assertion is informative?

We can explain the informativeness of the assertion by appealing to the two-dimensional matrix in much the same way as we did in sections 3 and 4, above. The interlocutors can all presuppose that the others know that there is a predominant clear odorless liquid (that has the sorts of functional characteristics that we associate with ‘water’, and they can all presuppose that the others are uncertain about water’s underlying chemical composition. For this reason, all worlds in the prior context set will have a predominant clear odorless liquid with the functional characteristics that they associate with ‘water’. However, since they are uncertain about the underlying constitution of water, the worlds will differ with respect to the chemical constitution of the predominant clear odorless liquid at that world. Consider two such worlds, an H<sub>2</sub>O world and an XYZ world, as featured in the two-dimensional matrix in figure 7.

Figure 7

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>
<b>H<sub>2</sub>O</b>	<b>T</b>	<b>T</b>
<b>XYZ</b>	<b>F</b>	<b>F</b>

We first consider consider the H<sub>2</sub>O world as actual, and evaluate what is said at that world. Given that the H<sub>2</sub>O world is actual, ‘water’ picks out H<sub>2</sub>O, and so what is said is

that  $H_2O$  is  $H_2O$ . Next, we evaluate whether what is said at the  $H_2O$  world is true at that  $H_2O$  world. Since what is said is true at that world, the intersection of the  $H_2O$  row and  $H_2O$  column is marked with a T. Next, we consider whether what is said at the  $H_2O$  world is true at the XYZ world. Since it is necessarily true that  $H_2O$  is  $H_2O$  then it is true at the XYZ world, and so we mark the intersection of the  $H_2O$  row and XYZ column with a T. Next, we consider the XYZ world as actual, and evaluate what is said by the assertion. If the XYZ world is actual, ‘water’ picks out XYZ, and so the assertion says that XYZ is  $H_2O$ . Since XYZ isn’t  $H_2O$ , then what is said at the XYZ world is false at the XYZ world. Furthermore, since it is necessarily false that XYZ is  $H_2O$ , then what is said at the XYZ world is false at the  $H_2O$  world, too.

To explain the informativity of the assertion, we look at the diagonal that is constituted by the intersection of each world row with that world’s column. The truth-value in each cell on the diagonal represents whether what is said, at that world on the hypothesis that that world is actual, is true at that world. Those worlds at which what is said at that world, on the hypothesis that that world is actual, is false are ruled out. A useful way to think about this is that, even though the interlocutors don’t know which world in the context set is actual, they do believe that they’ve narrowed down the ways the world could turn out in the relevant respects to those ways represented by the worlds in the context set. We can model the way in which they rule out worlds in such a situation by evaluating whether what is said at each world, on the hypothesis that that world is actual, is true at that world, for each world in the context set. Since they know that the assertion serves as a proposal to rule out those worlds at which what is said is false, then they rule

out those worlds at which what is said at that world, on the hypothesis that the world is actual, is false at that world.

This problem with necessary a posteriori sentences is an instance of a more general problem. More generally, we face difficulties in explaining the way some assertions change the state of conversation, and in modeling *how* these assertions change the state of the conversation in this way. These are two distinct but related problems. The first has to do with the difference between the state of the conversation just prior to, and the state of the conversation immediately following, the assertion, and the second problem has to do with modeling the mechanism by virtue of which the assertion affects this change on the context. We can see that these are distinct issues because two different views could predict the same change in the state of the conversation but theorize that it happens by virtue of different mechanisms. These issues are what I call *the problem of informativity*: the challenge of modeling the way in which an assertion affects the state of a conversation, and giving an adequate account of the mechanism by which it does so.

Let us say that the *informativity* of an assertion is the way the assertion changes the state of the conversation from the state immediately preceding the assertion to the state immediately following the assertion. Continuing with our example, had Bruce made his first assertion in the conversation, ‘the capital of Australia is Canberra’, rather than ‘The capital of Australia isn’t in Western Australia or South Australia’, both assertions would have been informative. However, they would have differed in informativity: the first assertion both introduces the Canberra possibility *and* rules out the Sydney, Melbourne,

Brisbane, and Perth possibilities, while the second only rules out the Perth and Adelaide possibility.

So far, this all seems very natural and unproblematic. However, as we will see in chapters two and three, we encounter difficulties in explaining the informativity of some assertions of sentences containing proper names or natural kind terms.



## **Chapter Two: Stalnaker's Metasemantic Interpretation**

In this chapter, we will look at the way that Stalnaker uses the two intuitions in modeling how assertions change the state of a conversation. In particular, we will look at the way that he uses the uniform extension intuition in modeling the force of necessary a posteriori assertions. Next, I will discuss the problem of informativity<sup>15</sup> as it arises on his view, and argue that the resources of his account are inadequate to address this problem.

### ***Section 1: Stalnaker on the Force of Assertions on the Standard Interpretation***

As we discussed in the preceding chapter, Stalnaker tells us that assertions are proposals to change the context set by ruling out those worlds that aren't the way the assertion says the world is, or to add worlds to the context set that are the way the assertion says the world is, if there aren't such worlds already in the prior context set. If an assertion is accepted, then those worlds in the prior context set that aren't the way the assertion says the world is are ruled out in light of that assertion, or those worlds that are the way the assertion says the world is introduced to the context set (if they aren't already included), yielding the posterior context set.

From this point forward, when I discuss Stalnaker's view on the force of an assertion, I mean *the force of an assertion that is accepted*. Assertions that aren't accepted might well not change the context set, or might have a different effect on it than is predicted in the case in which that assertion is accepted. For example, if one interlocutor makes an

---

<sup>15</sup> See section 4.2 of chapter one, in which I introduce and explain what I call the problem of informativity.

outrageous claim that the other interlocutors take to suggest that she isn't as well-informed on the subject matter as they might have originally thought, not only might the other interlocutors not update the context set in the way she proposes in making her assertion, but they might also reintroduce hypotheses that her earlier assertions had ruled out. At any rate, we will not concern ourselves in this project with the way in which assertions that aren't accepted affect the context set.

Also, we will only focus on the way that assertions change the context set in light of their content (or, as we will discuss, in at least some cases, *what the interlocutors take to be their content*). That is, we will only be concerned with the changes to the context set in light of the way the assertion says the world is, or in light of the way the interlocutors take the assertion to say the world is. So, for example, suppose that we are talking about which Australian city is the capital, and you assert 'the capital of Australia is Canberra'. Suppose that your assertion *somehow*, oddly enough, causes the other interlocutors and I to rule out *other*, irrelevant, hypotheses about the way the world might be, such as the hypothesis that the rain in Spain falls mainly on the plains. That such a case is rather unlikely, or that I haven't given a story about how such a case might come to pass, isn't important here—the point of the example is merely to draw attention to which way in which assertions change the context set that is of interest to us, rather than any old way they might happen to change the context set.

For our purposes, the *force* of an assertion is the impact of the assertion on the prior context set, yielding the posterior context set. About his account of force, in his introduction to *Context and Content*, Stalnaker writes:

I suggested that an assertion should be understood as a proposal to change the context by adding the content to the information presupposed. This is an account of the *force* of an assertion [...] [T]he assertion rule takes the prior context set to a posterior context set, which is the intersection of the prior context set with that content (10-11).

We are to model the force of an assertion, according to Stalnaker, as the intersection of the set of worlds in the context set just prior to the assertion and the set of worlds that is the way the assertion says the world is. As we saw in chapter one, one way to think about what this means, such that it accommodates the possibility of *adding* worlds to the context set, rather than merely ruling out worlds, is as follows.<sup>16</sup> The process of updating the context set occurs in two stages. The first stage is that, immediately following the assertion, *the content of the assertion*—the set of worlds that is the way that the assertion says the world is—is added to the context set. The second stage is that all worlds that aren't in the intersection of the prior context set (after the content of the assertion has been added) and the content of the assertion are ruled out by the assertion. Thus, all worlds in the posterior context set are the way the assertion says the world is.

It is important that we accommodate the phenomena that at least some assertions involve adding worlds to the context set, rather than merely ruling out worlds from the context set. Sometimes, a possibility is introduced in conversation that wasn't previously among

---

<sup>16</sup> Thanks to Wolfgang Schwarz and Daniel Nolan for helpful discussion.

the presuppositions of the conversation, and in some of these cases, it doesn't involve also ruling out hypotheses currently under consideration. For example, recall our example from chapter one in which the fifth-grade class is trying to work out which major Australian city is the capital. When Bruce asserts, 'Canberra is a major Australian city', this introduces a new possibility and so changes the presuppositions of the conversation. Before Bruce's assertion, the presuppositions in the conversation were that ways that the Australian capital could turn out were that it is Brisbane, that it is Melbourne, that it is Sydney, that it is Adelaide, and that it is Perth, and so there is at least one world in the prior context set at which the capital is Brisbane, and at least one at which the capital is Melbourne, and at least one at which the capital is Sydney, and at least one at which the capital is Adelaide, and at least one at which it is Perth, and there are no worlds in the prior context set at which the capital is any city other than these, or that Australia doesn't have a capital, or that Australia fails to exist, etc. Bruce's assertion serves as a proposal to add to the context set a way the world might turn out to be that was not under consideration prior to his assertion: the world might turn out such that Canberra is the capital. This changes the presuppositions for the conversation, such that the way that the students may all reasonably presuppose are ways that the capital of Australia might turn out, for all their fellow classmates believe, are that the capital is Brisbane, or that it is Melbourne, or that it is Sydney, or that it is Adelaide, or that it is Perth, or that it is Canberra, and that these are the only ways that the capital of Australia could turn out (and so Australia does not fail to have a capital, and Australia exists). Furthermore, since at none of the worlds in the prior context set was Canberra a major city, then first, introduced to the prior context set will be worlds at which Canberra,

Sydney, Brisbane, Melbourne, Perth, and Adelaide are all major cities, at least one of which is such that Canberra is the capital, and another at which Sydney is the capital, and so on, such that each city is the capital at one or more worlds in the prior context set, and then all the worlds that failed to have Canberra as a major city are ruled out from the prior context set.

### ***Section 2: Direct Reference Theory***

As we noted in chapter one, we can model the content of an assertion using the set of worlds that are the way that assertion says the world is. Now that we are investigating the particular details of Stalnaker's view, in order to know which worlds to use to model the content of the sentence on his view, we need to know what semantic theory to apply in order to know what the sentence says according to that semantics.<sup>17</sup> As before, we will continue to restrict our focus to proper names and natural kind terms, and, in keeping with Stalnaker,<sup>18</sup> we will adopt direct reference theory as giving the account of the semantics of those expressions. Direct reference theory (DR) says that expressions refer *directly* (i.e., in an unmediated way) and *rigidly* (i.e., picks out the same extension at any world at which it picks out anything at all) to their referents. When we use the natural kind term 'water', we refer directly to H<sub>2</sub>O, regardless of the superficial characteristics of

---

<sup>17</sup> To see this, contrast direct reference with a simple descriptivism that says the reference of the expression is mediated by description. Consider what each theory says about the semantics for 'water'. Direct reference says that 'water' refers rigidly to H<sub>2</sub>O. Descriptivism says that 'water' refers to whatever satisfies, say, 'the predominant clear odorless liquid'. These differences result in a different modal profile for 'water', and different truth-conditions for sentences in which 'water' is a constituent expression.

<sup>18</sup> For example, see "Assertion Revisited", p. 295.

H<sub>2</sub>O: 'water' refers to H<sub>2</sub>O even when it is not clear, odorless, or a liquid, and it doesn't refer to anything else (e.g., a sample of XYZ<sup>19</sup>) even if it *is* a clear odorless liquid.

For example, suppose that you are hosting some friends at your home, and you offer your guests something to drink, gesturing at an opaque pitcher. When one of your guests asks what's on offer, you assert, 'there is water in the pitcher'. Your assertion, 'there is water in the pitcher', is modeled using the set of worlds in which there is H<sub>2</sub>O in the pitcher.

Here, we can see that direct reference theory captures the uniform extension intuition: proper names and natural kind terms uniformly pick out the same individual or kind, respectively. Furthermore, since, according to DR, proper names and natural kind terms refer directly, this means that, whatever role (if any) is played by the descriptions we associate with these expressions, it isn't a matter of mediating the reference. On Stalnaker's view, these descriptions play a *reference-fixing role* (if they play any role at all). The idea is that, at the time the reference of the expression was fixed, the individual who introduced the expression used the reference-fixing description to uniquely determine the extension of the expression.

For example, suppose that the reference of 'water' was fixed by the description 'the predominant clear odorless liquid'.<sup>20</sup> Since H<sub>2</sub>O is what satisfied the description (or, at least the relevant sample of water, presumably nearby the location of the reference-fixing event at the time the reference was fixed), the reference of 'water' was fixed to H<sub>2</sub>O. In general, we needn't assume that the reference of all such expressions was fixed by

---

<sup>19</sup> Suppose, as before, that XYZ is superficially indiscernible from H<sub>2</sub>O, though they differ in chemical constitution.

<sup>20</sup> I am bracketing worries about uniquely identifying descriptions here. The worry is that if the would-be reference-fixing description doesn't uniquely refer then reference-fixing will be unsuccessful.

description, but for ease of discussion in what follows, we will treat these expressions as though they were fixed by description.<sup>21</sup>

Once the reference of the expression has been fixed, subsequent successful use of ‘water’ to refer to H<sub>2</sub>O depends on the speaker of the token of ‘water’ standing in an appropriate causal chain to the reference-fixing event. Generally, this causal chain is such that the links are the points of transmission of the expression from one individual to another. Suppose that individual A acquired the expression ‘water’ from individual B, and B acquired the expression from individual C, and C acquired it from D, who introduced the term by fixing the reference of ‘water’. The causal chain that ensures that A’s use of a particular token of ‘water’ refers to H<sub>2</sub>O is the chain by which she came to have the expression, which ultimately ends with the person who fixed the reference of the expression: it is because she acquired it from B, who acquired it from C, who acquired it from D, who fixed the reference of the term.

The metasemantic facts for an expression are those facts that explain why the extension assigned to the expression by an interpretation is the right one. For example, the metasemantic facts for ‘water’ explain why it is that the extension of ‘water’ is H<sub>2</sub>O. If the reference of ‘water’ was fixed using the description ‘the predominant clear odorless liquid’, then both that description, and the fact that H<sub>2</sub>O satisfied the description for the purposes of reference-fixing, are among the metasemantic facts for ‘water’. The metasemantic facts for why a particular token of ‘water’ refers to H<sub>2</sub>O include the reference-fixing story as well as an appropriate causal chain between the token and the

---

<sup>21</sup> The worry that I raise for Stalnaker later in this chapter can be modified to accommodate other ways of fixing the reference, such as using demonstratives (e.g., fixing the reference of ‘water’ pointing at a sample of water, saying “call *that* ‘water’”).

reference-fixing event. These metasemantic facts play an important role for Stalnaker in capturing the contingent extension intuition, and for the purposes of modeling the force of assertions, as I discuss in section two of this chapter.

### *Section 3: Which Worlds are in the Context Set?*

There remains a tension regarding which worlds are in the prior context set for an assertion. Given DR, proper names and natural kind terms refer uniformly to the same individual and kind, respectively. This means that, given that the reference of ‘water’ was fixed to H<sub>2</sub>O, all subsequent tokens of ‘water’ that stand in the right relation to the initial reference-fixing event<sup>22</sup> refer uniformly to H<sub>2</sub>O whether a given group of interlocutors knows it or not. On the other hand, the context set is intended to represent the relevant ways the world could turn out to be, for all (at least some of) the interlocutors know. For example, suppose that you are hosting a party, and you offer your guests a drink, gesturing at an opaque pitcher. You’d already told your guests earlier that the drinks you have to offer are water, chardonnay, and lemonade. Suppose that your guests don’t know whether water is H<sub>2</sub>O, or some other substance that is superficially indistinguishable from H<sub>2</sub>O—call it XYZ. If the context set is to include all of the relevant ways that the world could turn out, for all the interlocutors know, and if, in a discussion about water, for all the interlocutors know, water could turn out to be H<sub>2</sub>O or XYZ, then it seems that both H<sub>2</sub>O and XYZ worlds belong in the prior context set.

---

<sup>22</sup> For example, the appropriate relation might be a causal-historical one, such that a given token of ‘water’ refers rigidly to H<sub>2</sub>O just in case the reference-fixing event for ‘water’ fixed the reference to H<sub>2</sub>O, and that token of ‘water’ is used by an individual is connected to a causal-historical chain of individuals who acquired his or her use of ‘water’ ultimately from the individual who fixed the referent of ‘water’. See my discussion of this in chapter one for more details.



Suppose that one of your guests asks you what is in the pitcher, and you assert, ‘there is water in the pitcher’. If there are both H<sub>2</sub>O worlds and XYZ worlds in the prior context set, now we are faced with a problem. If the force of assertions is the intersection of the *content* of the assertion (i.e., the set of worlds at which there is H<sub>2</sub>O in the pitcher) with the prior context set, and if there are XYZ worlds in the prior context set, then your assertion, ‘There is water in the pitcher’, rules out not only the lemonade world and the chardonnay world, but also any worlds at which there is XYZ in the pitcher—call them XYZ worlds. You might worry that this pares down the context set too far: even in light of your assertion, it looks like, for all your guests know, the contents of the pitcher could turn out to be H<sub>2</sub>O or XYZ.

The problem is this. On the one hand, if the context set is intended to capture the relevant ways the world could turn out to be, for all the interlocutors know, and for all the interlocutors know, water could turn out to be XYZ, and, for all they know, there could be water in the pitcher, then it looks like worlds at which there is XYZ in the pitcher belong in the prior context set. On the other hand, if such a world is in the prior context set, it will be ruled out by the assertion, even though, at least intuitively, the interlocutors are not in a position to rule out the XYZ worlds in light of your assertion. In light of this, should the prior context set include worlds at which there is XYZ in the pitcher?

Since the XYZ worlds would be ruled out in light of such an assertion, it would be useful for the XYZ worlds to not be in the prior context set, provided that Stalnaker can give a reason for excluding them. Fortunately, his commitment to externalism about content

provides some means for excluding them. By externalism about content, I mean that content is determined by factors beyond our mental state—it is called externalism because these other factors are external to our heads, namely factors regarding what is in our environment, such that we come in causal contact with it. Our thoughts about water are thoughts about H<sub>2</sub>O, and not about XYZ, because we stand in the right sort of relation to the former and not the latter, namely that there is H<sub>2</sub>O, but not XYZ, in our environment. This means that in entertaining the possibility that there is water in the pitcher, we are entertaining the possibility that there is H<sub>2</sub>O in the pitcher, whether we know that we are entertaining the latter or not. We are not also entertaining the possibility that there is XYZ in the pitcher because we do not stand in the right relation to XYZ to have such thoughts, and so that there is XYZ in the pitcher is not among the possible thoughts we (or, more specific to our example, your guests) can entertain. Thus, for *most* assertions about water, there will be H<sub>2</sub>O worlds, but not XYZ worlds, in the prior context set.

#### ***Section 4: Diagonalization and the Necessary A Posteriori***

I say “for most assertions” because there are some cases, such as assertions of necessary a posteriori sentences, in which this would leave the force of the assertion unexplained. Stalnaker notes this, for example, in “Assertion Revisited”, in which he writes:

The problem is this: with some statements, there is a tension between global intuitions about the information that the statement conveys, as represented by the possibilities that the statement seems to exclude, and

what semantic theories that are otherwise well motivated say about the truth-conditions of the statement (294).

The problem is that, according to DR, such sentences are necessary, and so if an assertion of such a sentence serves as a proposal to rule out worlds that aren't the way the assertion says the world is, and there are no worlds that aren't the way the assertion says the world is (i.e., worlds at which the sentence is false), then we cannot explain the informativity of that assertion in terms of ruling out at least one world that isn't the way the assertion says the world is. However, such sentences *are* informative, and Stalnaker wants to explain the informativity of an assertion by the assertion being a proposal to change the state of the conversation, either by ruling out at least one world that isn't the way the assertion says the world is, or introducing at least one world that serves as introducing a new hypothesis about the way the world is that wasn't previously under consideration in the conversation, such that the context set is altered in this way. But if all worlds are in the prior context set are the way the assertion says the world is (because all worlds are the way the assertion says the world is), then there aren't any worlds that aren't the way the assertion says the world is such that they would be ruled out by the assertion. Similarly, since all worlds are the way the assertion says the world is, adding any world to the context set will not serve as adding a new hypothesis about the way the world is to the conversation because all of the worlds will already be this way. In light of this, how do we explain the informativeness of the sentence?

To see the problem, consider an assertion of 'water is H<sub>2</sub>O'. For one who doesn't know this already, this assertion is informative, seeming to rule out other competing hypotheses

regarding how water could turn out—this suggests that at least one world was ruled out by the assertion. Yet, given that, according to DR, ‘water’ refers uniformly to H<sub>2</sub>O, there are no worlds at which the sentence asserted is false, and so no worlds are ruled out by it.

To better understand why this is a problem, we will first revisit a more mundane case—your assertion of ‘there is water in the pitcher’—to see how such an assertion is evaluated both on the standard interpretation, and then on the kind of reinterpretation proposed by Stalnaker. We will then see how this reinterpretation can account for the informativeness of assertions that are necessary but also a posteriori. Recall that, for Stalnaker, a proposition is a set of worlds, and that, on the standard semantic interpretation, the proposition expressed by an assertion is the set of worlds that are the way the assertion says the world is on the standard interpretation. For example, on the standard interpretation, given that the reference of ‘water’ is fixed to H<sub>2</sub>O, the proposition expressed by asserting ‘there is water in the pitcher’ is the set of worlds at which there is H<sub>2</sub>O in the pitcher.

That this is the proposition expressed by this assertion is determined by the metasemantic facts for the constituent expressions: the facts about why those expressions have the extensions that they have. As noted earlier, on Stalnaker’s view, among the metasemantic facts for a token of a proper name or natural kind term are the facts about how the reference was fixed, and the particular causal chain between the original reference-fixing event and that token—these facts determine why a given token of that proper name or natural kind term has that extension. Given that proper names and natural

kind terms refer uniformly to the same individual or kind, respectively, across all worlds at which they pick out anything at all, this strongly suggests that, also among the metasemantic facts for such expressions is that they refer uniformly to the same individual or kind across all worlds at which they pick out anything at all—call this the *unifying feature*. Furthermore, that proper names and natural kind terms pick out the same kind uniformly across all worlds at which they pick out anything at all captures our uniform extension intuition.

For proper names and natural kind terms, there are many acceptable causal chains that could have linked a given token of such an expression to the reference-fixing event, and variation among these does not change the fact that that token has the extension that it has, and has that extension rigidly. For example, there could have been many acceptable causal chains that would just as well have linked my token of ‘water’ to the reference-fixing event. The variation in chains could, for example, be produced by me, or other members of the chain, acquiring the use of ‘water’ from a different individual. What is important here is that there is *some* causal chain or other connecting my token of ‘water’ to the reference-fixing event. For this reason, we will not concern ourselves with the specific details of the particular causal chain between the token of the expression as uttered by the speaker in the context and the reference-fixing event for the expression. Instead, we will only focus here on the first link of the causal chain: the reference-fixing event, and that there is some appropriate causal-historical link between the token of interest to us and the reference-fixing event for that expression. As I will explain shortly, we will treat the reference-fixer together with the unifying feature as the metasemantic

facts relevant to the reinterpretation. We will, for ease of exposition, suppose that the reference-fixer is a reference-fixing description.<sup>23</sup>

Keeping with our example, we will focus on the metasemantic facts for ‘water’. As we discussed earlier, according to DR, ‘water’ refers to H<sub>2</sub>O because the expression was introduced by virtue of fixing the reference of ‘water’ to H<sub>2</sub>O—for example, perhaps by using a description that the relevant sample of H<sub>2</sub>O satisfied at the time of the reference fixing, or because H<sub>2</sub>O was the thing demonstrated at the time the reference was fixed—and the particular token of ‘water’ that is a constituent of the assertion, ‘there is water in the pitcher’, stands in the right causal relation to the reference-fixing event, such that it picks out H<sub>2</sub>O.

Call the context in which the assertion was made *the context of utterance*, and call the circumstance in which we evaluate whether what was said by the assertion made in the context of utterance is true *the circumstance of evaluation*. The circumstance of evaluation may be the context of utterance, though it need not be. We can take what was said in the context of utterance and consider whether it is true under some other circumstances. Specific to our example, we can evaluate whether what was said in that context—namely, that there is H<sub>2</sub>O in the pitcher—is true at the context at which it is asserted, or we could evaluate whether what was said in that context is true at some other

---

<sup>23</sup> The worries that I raise for Stalnaker arise regardless of whether the reference is fixed by description, by a property such that whatever has the property at the time of reference-fixing is in the extension of the expression, by using demonstratives (‘call *that* water’), and other ways that the reference can be fixed. I have chosen to use a reference-fixing description for our purposes merely for the ease of exegesis.

circumstance, such when you offer your guests a drink, using that same pitcher, at some other party.

To represent this, we can use a matrix in which the worlds on the vertical column are the contexts of utterance, and the worlds on the horizontal column are the circumstance of evaluation. Since our example is the assertion, ‘there is water in the pitcher’, asserted in a context in which there is H<sub>2</sub>O in the pitcher, we have as the context of utterance an H<sub>2</sub>O world with H<sub>2</sub>O in the pitcher. For this reason, on the left-hand column, we see an H<sub>2</sub>O world that represents the context in which the assertion is made.

Figure 1

	<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>XYZ world with XYZ in pitcher</b>
<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>T</b>	<b>F</b>

In the top row, we see both an H<sub>2</sub>O world with H<sub>2</sub>O in the pitcher, and an XYZ world with XYZ in the pitcher. These two serve as two different circumstances of evaluation. At each of these circumstances, we ask whether what is said in the context of the assertion is true at the particular circumstance of evaluation under consideration.

Starting with taking the H<sub>2</sub>O world as the circumstance of evaluation, we ask whether what is said at the context of the assertion—that is, at the H<sub>2</sub>O world with H<sub>2</sub>O in the pitcher—is true at that circumstance of evaluation. Since what is said, namely, that there is H<sub>2</sub>O in the pitcher, is true at that circumstance, then what is said at the context in which the assertion is made is true at that circumstance of evaluation, and so the cell of the

matrix that is the intersection of the H<sub>2</sub>O world column and H<sub>2</sub>O world row is marked with a T. Next, we consider the XYZ world as the circumstance of evaluation. What is said at the context in which the assertion is made is false when evaluated at the XYZ circumstance because what is said is that there is H<sub>2</sub>O in the pitcher, and there isn't H<sub>2</sub>O in the pitcher at the XYZ circumstance—instead, there is XYZ in the pitcher. For this reason, the cell that is the intersection of the H<sub>2</sub>O world and the XYZ column is marked with an F.

We can also reinterpret the assertion by considering what proposition would be expressed in a *different* context. This captures the contingent extension intuition: we can consider hypotheses about how the extensions of our expressions would have differed, had the world been different in some respect(s). Suppose that this world had turned out to be an XYZ world—which we will understand to mean that we adopt the supposition that the actual world is an XYZ world—and so water had turned out to be XYZ. At such a world, if the reference-fixing event for 'water' made use of the same reference-fixer—let us suppose that the reference-fixer was the description, 'the predominant clear odorless liquid'—then, for the purposes of this reinterpretation, we use these metasemantic facts to reinterpret 'water' such that its extension is reassigned as XYZ.

Note that this is a *reinterpretation*, and so this is not to say that at an XYZ world, our expression, 'water', picks out XYZ. Rather, according to DR, on its standard semantic interpretation, 'water' uniformly picks out the same kind—H<sub>2</sub>O—across all worlds at which it picks out anything at all. This reinterpretation is not a feature of the *semantics*



of water, and is merely being used here for the purposes of modeling the force of assertions in cases in which the standard explanation cannot account for the informativeness of the assertion.

The idea behind this reinterpretation is that the interlocutors may not know how the world turns out in some relevant respects. In our example, your guests know some things about water, and they don't know other things about it. They know that it is a clear odorless liquid, that it comes out of taps, that it is commonly offered as a beverage, etc. They don't know its underlying chemical constitution. The reinterpretation allows us to capture this about their epistemic state with respect to water. By allowing the worlds in the vertical column to represent ways this world could still turn out to be, for all the interlocutors know, we can engage in standard interpretation for the actual world's row (though the interlocutors won't know which row this is) *and* reinterpretation for the rows for the other worlds in the context set, such that we use the same metasemantic facts about how the reference is fixed at our world to reassign the extension of the expression (for the purposes of reinterpretation only!) at each world in the vertical column.

For the purposes of reinterpretation, let's treat the XYZ world at which there is XYZ in the pitcher as the context of utterance. Reinterpreting 'water' such that XYZ is in its extension, the assertion, 'there is water in the pitcher', reinterpreted as uttered in that context, expresses the proposition that there is XYZ in the pitcher. We can represent both this context of utterance and the preceding context of utterance in the same matrix by adding another row in which the XYZ world at which there is XYZ in the pitcher is

the context of utterance. We can then evaluate whether what is said in that context is true at each of the two circumstances of evaluation.

Figure 2

	<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>XYZ world with XYZ in pitcher</b>
<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>T</b>	<b>F</b>
<b>XYZ world with XYZ in the pitcher</b>	<b>F</b>	<b>T</b>

Evaluated at the H<sub>2</sub>O world, what is said (on reinterpretation) at the XYZ context of utterance is false, and so the intersection of the XYZ world row and the H<sub>2</sub>O column is marked with an F. Evaluated at the XYZ world, what is said (on reinterpretation) at the XYZ context of utterance is true, and so the intersection of the XYZ row and XYZ column is marked with a T.

We can now see an interesting feature of the matrix: it has a *diagonal* that consists in the intersection of the H<sub>2</sub>O column and row, and the intersection of the XYZ column and row. This is why we, following Stalnaker, call the reinterpretation *diagonalization*. Notice that, for this assertion, both of these cells on the diagonal are marked with a T, telling us that, at each of these contexts, what is said at each context is true when evaluated at that context. Stalnaker, Jackson, and Chalmers all acknowledge this diagonal, but their views differ in what they take to be the significance of it. Here, we will focus on what Stalnaker takes to be the significance of the diagonal, and we will discuss what Jackson and Chalmers take to be the significance of it in the next chapter.

Given Stalnaker's externalism about content, which has, as a consequence that, as we discussed earlier, since we can entertain thoughts about H<sub>2</sub>O and not thoughts about XYZ, there will be H<sub>2</sub>O worlds but not XYZ worlds in the context set unless we have some other means of appealing to them for the purposes of reinterpretation. This means that in order to appeal, for the purposes of reinterpretation, to a diagonal on which the extension of at least one of the proper names or natural kind terms that is a constituent of the assertion differs from its extension on the standard interpretation, we, as theorists who are modeling the force of assertion, need to add to the prior context set worlds that are qualitatively similar to our own world, but that differ with respect to the relevant facts. For example, in order to appeal to the diagonal that we discussed above, we would need to add XYZ worlds to the prior context set. We, as theorists engaged in modeling the force of assertion must add these worlds to the prior context set for the purposes of reinterpretation if they are not already in the prior context set.

Note that we are not, as theorists, attributing to the interlocutors thoughts *about* XYZ (e.g., the thought that 'water is XYZ') because, given the commitment to externalism, the interlocutors aren't in causal contact with XYZ, and so cannot have such thoughts. That is to say, it is important to note that this reinterpretation is consistent with his externalism about content: it is the *theorist*, in modeling the communicative exchange, that makes use of these worlds, and so this is not a rejection of the earlier claim that the interlocutors do not stand in the right relation to XYZ to have thoughts about XYZ. In other words, we aren't attributing such thoughts to the interlocutors, but rather, as theorists, we are using

them to represent alternative possibilities for ways that water could turn out, for all the interlocutors know.

Since, according to Stalnaker, we appeal to reinterpretation only when the standard interpretation won't explain the informativeness of an assertion, and the standard interpretation for assertions about water will only require H<sub>2</sub>O worlds, then it is very likely that there won't already be XYZ worlds in the prior context set. There would only already be XYZ worlds in the prior context set if we theorists, in modeling the conversation, needed to introduce them previously to explain the informativity of an assertion that could not be explained by appealing only to standard interpretation, and, had they been introduced previously, it is likely for the purposes of being in the prior context set of an assertion, such that they were ruled out by the assertion, and so would not be in the context set anymore. We add XYZ worlds to the prior context set for the assertion so that this serves as a world at which what is said, on reinterpretation, differs from what is said at the actual context. As we saw earlier, on reinterpretation at the XYZ world, the assertion, 'there is water in the pitcher', said that there is XYZ in the pitcher, and this differs from what is said by the assertion on its standard interpretation, namely that there is H<sub>2</sub>O in the pitcher.

In the mundane example of 'there is water in the pitcher', if this assertion is informative, it is because the interlocutors weren't sure about the contents of the pitcher (perhaps it is water, lemonade, or chardonnay), and so we need only appeal to standard interpretation in this case. However, for the purpose of accounting for necessarily true assertions that

are informative, we need to appeal to reinterpretation, which, as we noted earlier, Stalnaker calls *diagonalization*: a reinterpretation of an assertion on which what is expressed is not the set of worlds at which what is said on standard interpretation is true, but rather that the assertion expresses the diagonal proposition—the set of worlds at which what is said on reinterpretation at that world, is true at that world. Explaining this in “Assertion Revisited”, he writes:

On this kind of account, diagonal propositions [...] are derivative from the standard semantics, as it is in the actual world, and in the relevant alternative possible worlds. My project began from the assumption that the standard semantics [...] was essentially right—the project was to reconcile it with the facts that statements that were necessary, according to that semantics, could be used to communicate contingent information. Diagonalization was *reinterpretation*—interpretation that was parasitic on the standard interpretation, which was assumed to give the right result for what was expressed and communicated in the normal case. If the standard semantics for names did not give the right result in the normal case—that is, if it were not right to say that normally the beliefs one expresses and the information that one conveys when one uses proper names are singular propositions about the individual named—then the semantic account would not be defensible (297).

Diagonalization is parasitic on standard interpretation: just as the metasemantic facts for an expression explain why it is that the expression has the extension that it has, we use the metasemantic facts for that expression to determine the reassigned extension of that

expression at each world in the context set, such that, had that world been actual, the extension of the expression would have been as reassigned at that world (e.g., in the present case, as XYZ). This reinterpretation procedure allows us to model the force of such assertions by ruling out from the prior context set each world at which the assertion, as reinterpreted at that world, is not true at that world, thus giving us the posterior context set as just those worlds that were in the prior context set at which what was said, on reinterpretation at that world, is true at that world.

Figure 3

	<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>XYZ world with XYZ in pitcher</b>
<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>T</b>	<b>T</b>
<b>XYZ world with XYZ in the pitcher</b>	<b>F</b>	<b>F</b>

Let's look at how diagonalization will help to account for the informativity of 'water is H<sub>2</sub>O'. Recall that, according to DR, that assertion is necessary, and so there are no worlds that aren't the way it says the world is, and yet it is informative, which suggests that it rules out at least one world as not being the way it says the world is. Suppose that the interlocutors in the conversation in which this assertion is made do not already know that water is H<sub>2</sub>O—for all they know, it might be H<sub>2</sub>O, or XYZ, or some atomic substance that fails to admit of further decomposition, or something else. Introducing XYZ worlds into the prior context set for the assertion, 'water is H<sub>2</sub>O', allows us to explain why this assertion is informative.

The assertion, on its standard interpretation, expresses the proposition that  $H_2O$  is  $H_2O$ . This proposition is necessarily true, and so we can see that it is true as evaluated at the XYZ world. However, if we engage in reinterpretation, in which the metasemantic facts for ‘water’ are used to reassign the extension of ‘water’ at the XYZ world as XYZ, then the proposition expressed at the XYZ world is that XYZ is  $H_2O$ , and that is necessarily false. Since what is said on reinterpretation at the XYZ world is false at that XYZ world (i.e., on the diagonal), that world is ruled out by the assertion. Thus, reinterpretation allows for at least one world to be ruled out by the assertion, accounting for the fact that the assertion is informative.

Stalnaker thinks that the important work that can be done by the diagonal proposition—the set of world at which what is said at that context is true at that context, which is all the world that are marked as true on the diagonal—is that it can be used to model the force of assertions in the cases in which modeling the force of the assertion using the standard interpretation would not allow us to explain the informativeness of the assertion. Rather than modeling the force of the assertion by first adding the *content* of the assertion—the set of worlds that are the way the assertion says the world is, on the standard interpretation of the assertion—to the context set, and then ruling out from the context set any worlds that aren’t the way the assertion says the world is, we model the force of the assertion by first adding any worlds that we need to add to the prior context set (in our present example, adding at least one XYZ world), and then ruling out any worlds at which the assertion isn’t true on the diagonal. Looking at the diagonal, we can see that the assertion is false on the diagonal for the XYZ world, and so the XYZ world is

ruled out by the assertion. This, according to Stalnaker, explains the informativity of the assertion.

There are at least two ways of thinking about diagonalization, depending on which way one understands diagonalization to be parasitic on normal interpretation. One way is to understand this to mean that the very same metasemantic facts that explain why our words have the extensions they do in the actual world are used for reinterpretation. Let us call this the *fixed metasemantic* interpretation of diagonalization because we hold fixed the actual relevant metasemantic facts—in this case, the reference-fixing description—as we reinterpret across worlds. Another way is to interpret diagonalization to require that at each world in the context set, we only hold fixed very general facts about the nature of metasemantic facts, such as that there must be some manner in which the reference of the expression is fixed (and perhaps further qualifications regarding how reference can be fixed), and facts about what makes subsequent tokens refer to the same extension as the token that introduced the expression, etc., but it is the metasemantic facts at *that* world—which might well differ from our own—that determines the extension of the expression at that world. Let us call this the *metalinguistic* interpretation. We will discuss these two ways of understanding diagonalization.

#### *Section 4.1: The Fixed-Metasemantics Approach to Diagonalization*

On the fixed metasemantic approach, it is the actual metasemantic facts about the expression that are used for reinterpretation. This means that the actual reference-fixer—that is, whatever was used to fix the reference of the expression at the actual world—is



used to reinterpret the sentence at each world. On this approach, reinterpretation proceeds in the way discussed above for our 'water is H<sub>2</sub>O' example. If the reference-fixer for 'water' was 'the predominant clear odorless liquid', then we use this description to reinterpret sentences in which 'water' is a constituent by allowing the extension of 'water' to be reinterpreted at a given world as whatever satisfies that description at that world. Since, at the XYZ world, it is XYZ that satisfies that description, then on reinterpretation, 'water' refers to XYZ, and so the assertion, 'water is H<sub>2</sub>O', reinterpreted at that world, is false at that world.

There are substantial worries for this approach that arise from the possibility of reference-fixing descriptions or properties that appeal to not commonly known features of that which is in the extension of the expression. One problem is that we encounter a kind of epistemic problem on which the theorist, in attempting to use diagonalization to model the force of an assertion of, say, some necessary a posteriori sentence, won't know precisely what was the reference-fixer for a proper name or natural kind term that is a constituent expression of that assertion, such that she cannot use it for the purposes of modeling the force of such assertions. That is, if the idea is that we use the metasemantic facts as they are in our world to reinterpret the constituent expressions at all of the worlds in the prior context set, then if the theorist doesn't *know* what served as the reference-fixer, she won't be able to use that same reference-fixer for the purposes of reinterpretation in order to use it to explain the informativity of the assertion. Though there is a fact of the matter regarding the diagonal (given some set of worlds as the prior context set, and determinate facts about the metasemantics of the constituent

expressions), the theorist isn't in a position to use that information in giving an account of the force of the assertion.

Note that my point here is not that the theorist cannot offer an explanation at all about why the assertion is informative, but only that she cannot use the reference-fixer for the purposes of reinterpretation for explaining the informativity of the assertion. She might well use something else, such as a description of the manifest properties of the referent. Perhaps such a description *is* the reference-fixer for the expression. The point is that, in at least some cases, the theorist isn't in a position to know what the reference-fixer was, and since she doesn't know it, she cannot use it in predicting the force of the assertion. This is unattractive, given that, on this understanding of diagonalization, the procedure that the theorist is to follow is to use the reference-fixer for the purposes of reinterpretation to explain the informativity of such assertions.

The second worry is that, depending on the reference-fixer for an expression, diagonalization may make bad predictions about the force of the assertion. To see this, consider the following example. Suppose that the reference of 'water' was fixed by the description 'whatever liquid pools in St. Nectan's Kieve'. St. Nectan's Kieve is a small pool at the base of a waterfall in Cornwall, England and is perhaps best known as the place where, in the legend of King Arthur, the Knights of the Round Table were baptized just before embarking on their quest for the Holy Grail. It is important to note for the purposes of our example that St. Nectan's Kieve is a real place, and not merely just a fictitious part of the King Arthur story. Suppose that I mistakenly think that St. Nectan's

Kieve is merely a mythical pool—that it is part of the King Arthur story, which, let's suppose, I think is purely myth—and nothing more. If I think this of St. Nectan's Kieve, then all worlds that are still ways this world could turn out, for all I know, won't contain St. Nectan's Kieve. Furthermore, let's suppose that you share this belief that St. Nectan's Kieve isn't a real body of water (and that we both know that the other believes this), and that we are the only two people in the conversation, such that none of the worlds in the context set contain St. Nectan's Kieve.<sup>24</sup>

Suppose that you assert, 'water is H<sub>2</sub>O', and that I find this assertion informative. As we know, Stalnaker prescribes that we use diagonalization to explain the informativeness of this assertion. Here's the problem. If we are to use the reference-fixing description for 'water' (i.e., 'whatever liquid pools in St. Nectan's Kieve'), then on reinterpretation at each world in the context set, 'water' will fail to be assigned to anything because each world in the context set fails to have St. Nectan's Kieve. I take it that this means that the sentence will be false on the diagonal for each world in the context set, since 'water' will fail to have extension on reinterpretation at each of these worlds. But this gets the wrong result as an attempt to model the force of the assertion, as it seems that, intuitively, the assertion rules out some but not all worlds in the context set, but this diagonalization predicts that all such worlds are ruled out.<sup>25</sup>

---

<sup>24</sup> Note that this means that the actual world is not in the context set, since there really is a St. Nectan's Kieve in the actual world, and no worlds in the context set contain St. Nectan's Kieve.

<sup>25</sup> It's worth noting that even though the worries I raise here are explicitly directed at using reference-fixing description, a similar case can be made for other reference-fixers. Suppose that the reference of 'water' was fixed via demonstrating at a sample of water, saying 'call that 'water'.' Since we don't know the conditions (such as what a counterpart of the situation would have to be like in order to reinterpret 'water' at another world in the context set, such as what was demonstrated at in the reference-fixing event, or if I happen to not believe in the existence of whatever was the thing demonstrated), the same epistemic and bad

One might object that it is very likely that, in general, the reference-fixing description for a proper name or natural kind term described some observable or functional features of that which is in the extension of the expression. For example, one might argue that it is likely that the reference-fixing description for ‘water’ described manifest properties of water (e.g., that is a clear odorless liquid), and perhaps facts about where it can be found (e.g., falling as rain, forming lakes, rivers, and oceans) and what one can do with it (e.g., drink it, use it to hydrate one’s crops, wash one’s dog with it, etc.). My point here is not that, for any given reference-fixer, that reference-fixer did not make use of the sorts of observable and functional features that we associate with that which is in the extension of the expression. My point is rather than nothing *guarantees* that, in each case, the reference-fixing description describes these features, and that, when it does not, this allows for the possibility of poor predictions of the force of assertions.

I take these to be serious worries for the fixed metasemantic approach, making it an unattractive approach to explaining the informativity of assertions that are necessarily true. Since these worries do not arise on the metalinguistic approach, we will adopt this as our understanding of diagonalization in what follows.

#### *Section 4.2: The Metalinguistic Approach to Diagonalization*

Let us understand the metalinguistic approach to prescribe the following procedure for diagonalization. At each world in the context set, we engage in reinterpretation at that

---

prediction worries can arise (e.g., suppose the person who introduced the term was pointing at the water in St. Nectan’s Kieve.)

world, not by using the metasemantic facts for the constituent expressions of the sentence as they are in our world, but as they are in that world. At each world in the context set, we reinterpret the sentence, allowing ‘water’ to pick out whatever satisfies the reference-fixing description for ‘water’ as it is at that world.<sup>26</sup> As on the fixed metasemantic account, if what is said on reinterpretation at a world is false at that world, the world is ruled out of the context set.

One attractive way of understanding this approach is to allow the worlds in the context set to be such that the metasemantic facts for an expression at a given world coincides with what the interlocutors know about that which is in the extension of the expression.<sup>27</sup>

The idea is that, in a conversation, the interlocutors will know, or at least believe, some things about the nature of that which they are discussing, and not others. For example, the interlocutors might know that water is the predominant clear odorless liquid, but not that it is H<sub>2</sub>O. We use the interlocutor’s common knowledge—or at least common beliefs, since knowledge is factive<sup>28</sup>—about water to determine the metasemantics for ‘water’. Since the interlocutors believe that water is the predominant clear odorless liquid, then all of the worlds in the context set are such that the reference-fixing description for ‘water’ is the description ‘the predominant clear odorless liquid’.

---

<sup>26</sup> This is the simplest sort of case. If the interlocutors are uncertain about some features of that which they are discussing, this uncertainty will be represented by variation in the metasemantic facts, such that they differ across worlds to represent the varying hypotheses about the way the thing they are talking about could turn out to be, for all they know. This more sophisticated version is unnecessary for present purposes, and so we won’t discuss it here.

<sup>27</sup> Thanks to Wolfgang Schwarz for helpful discussion here.

<sup>28</sup> We will get better predictions of the force of an assertion if we use their common beliefs, regardless of whether the beliefs are true. We will discuss this further when it becomes particularly important later in this chapter.

Applying this to our ‘water is H<sub>2</sub>O’ example, after adding at least one XYZ world to the prior context set, we reinterpret ‘water is H<sub>2</sub>O’ at each world in the prior context set using the description that captures their common beliefs about water, i.e., ‘the predominant clear odorless liquid’. Since what is said on reinterpretation at the H<sub>2</sub>O worlds is true at those H<sub>2</sub>O worlds, those worlds aren’t ruled out, and since what is said on reinterpretation at each XYZ world(s) is false at that world, each XYZ world is ruled out. That at least one XYZ world is ruled out explains the informativity of the assertion.

The most important difference between the fixed metasemantic approach and the metalinguistic approach to diagonalization is that the metalinguistic approach is *guaranteed* to be closely tied to the epistemic state of the interlocutors, while there is no such guarantee on the fixed metasemantic approach. On the metalinguistic approach, we are using what the interlocutors believe about water to determine the reference-fixing description for ‘water’ at each world in the context set, as opposed to, on the fixed metasemantic approach, using the actual reference-fixing description for diagonalization, whether or not it is a description that the interlocutors associate with the expression. The fixed-metasemantic approach, as I noted above, can lead to problematic predictions about the informativity of some assertions.

However, as we will see in the next section, even though the metalinguistic approach isn’t subject to the worries that I’ve raised for the fixed metasemantic approach, it is not without its problems. It makes bad predictions about the state of the conversation after some assertions.

### *Section 5: Some Mistakes Left Unexplained*

Even if we adopt the metalinguistic approach to diagonalization, Stalnaker's view leaves unexplained some kinds of mistakes that, at least intuitively, we seem able to make in updating regarding what we take the world to be like, in light of assertions. Suppose that Bruce has taken his fifth grade class on a field trip to Alaska, and they are standing in a field on a cold morning, looking at a basin from a great enough distance that they cannot see its contents. One of his students wonders aloud about what the contents of the basin might be. Suppose that his students have seen water in both its liquid and solid forms, but they do not know that they are of the same kind in different states—instead, they are firmly of the opinion that water and ice are of two different kinds. Imagine the situation in whatever way you like such that these students do not know that ice is water in its frozen state. Perhaps they are from a warm climate that happens to not have freezers available for home use, and so their only exposure to ice is in their science class, and it is removed from the classroom before it begins to melt.

Suppose that his students do know that ice is found in cold places, and that water is pretty much found anywhere, and that perhaps they suspect that the basin might also contain gravel or mud, and so among the ways the contents of the basin could turn out, for all they know, is that it contains mud, gravel, liquid water, frozen water (though they certainly wouldn't describe the case this way). Given that they don't know that water is a solid at cold temperatures, they will expect water to exhibit the appearance familiar to them: as a clear odorless *liquid* (i.e., in its liquid state). This means that the prior context

set contains worlds at which there is ice in the basin (among which is the actual world), and worlds at which there is a clear odorless liquid<sup>29</sup> in the basin, worlds at which there is mud in the basin, and worlds at which there is gravel in the basin.

Bruce does not know that his students have this mistaken belief about ice since he was unaware that his students weren't paying attention when he told them that ice is water in its frozen state. Unlike his students, he knows that there is water in the basin (though he doesn't know whether it is in its solid or liquid state currently because he doesn't know whether it is above or below the freezing point), and so he replies to the query by asserting, 'there is water in the basin'. In light of his assertion, the students rule out the possibility that there is ice in the basin.

It certainly seems correct to say that his students *mistakenly* rule out the actual world, and all other worlds at which there is ice in the basin, since what he says is true at those worlds: what he said is true at the actual world, and at other H<sub>2</sub>O worlds at which there is ice in the basin, and yet they took his assertion to be a proposal to rule out such worlds, and they accepted his proposal, thereby ruling out those worlds. How can we explain this mistake using the resources of Stalnaker's theory? What we are looking for is a way of explaining, using the resources of Stalnaker's account, why his students took his assertion to be a proposal to rule out the actual world and any other H<sub>2</sub>O worlds with ice in the basin. We are interested in this explanation because if the audience of an assertion takes that assertion to be a proposal to rule out certain hypotheses about the way the world is—

---

<sup>29</sup> Note that this cannot be a world at which there is water in the basin because water is in its frozen state at the ambient temperature in this example.



namely, those hypotheses that the world is not as the assertion says the world is—then in order to explain why the students took his assertion as a proposal to rule out the actual world and any other H<sub>2</sub>O worlds that have ice in the basin, we need to find a way of explaining how his students could have come to take his assertion as a proposal to rule out those worlds. If such an explanation cannot be given then this case is left unaccounted for on Stalnaker's view.

Let's first hypothesize that the force of Bruce's assertion is the intersection of the prior context set with the content of the sentence asserted. The sentence is true at the worlds in the context set at which there is ice in the basin (including the actual world) by virtue of the fact that 'water' rigidly refers to H<sub>2</sub>O, and ice is H<sub>2</sub>O in its frozen state. This predicts that they wouldn't take his assertion as a proposal to rule out the actual world and other H<sub>2</sub>O worlds with ice in the basin, and so this does not allow us to explain the mistake.

From this, we can see that we cannot explain their mistake by appealing to the standard interpretation of his assertion, since what is said on the standard interpretation of his assertion is true at the actual world and other H<sub>2</sub>O worlds with ice in the basin. Since we cannot explain it on the standard interpretation of his assertion, we will need to use other resources of Stalnaker's view to explain it, namely by appealing to diagonalization.

However, we are only to appeal to diagonalization in cases in which there is a prima facie violation of a conversational maxim, and so we must consider whether it is appropriate to appeal to diagonalization in this case. Following Grice, Stalnaker thinks that following

conversational maxims—such as not asserting things that are necessarily true or necessarily false (since neither rules out any possibilities on its standard interpretation), or to assert things that one’s addressees understand—is required for the purposes of having a conversation, and so they are to be taken as common ground among the interlocutors in a given conversation. In “Assertion Revisited”, he writes:

The proposal made in “Assertion” was that in special cases, where there was a prima facie violation of certain conversational rules, utterances should be reinterpreted to express the diagonal proposition, rather than the proposition expressed according to the standard semantic rules. [...] The presumption that such rules are being followed constrains what is said. It is presumed that hearers will try to find a way of understanding what is said that conforms to the maxims, and speakers may exploit this presumption by saying things that would be manifest violations of conversational rules if they were interpreted in a standard way, and so will require reinterpretation. In the case of our problematic statements, the relevant maxim is that the speakers presume that their addressees understand what they are saying. [...] Reinterpretation by taking the diagonal proposition to be the one the speaker intends to communicate brings the statement into conformity with the rule, and seems to give the intuitively correct response.

Whether we may appeal to diagonalization here depends on whether Bruce’s assertion violates a conversational maxim. It seems that the relevant maxim here is that speakers say things that their audience will understand, since it is by virtue of his students not

knowing that ice is water in its solid state that led them to take his assertion as a proposal to rule out the actual world and other H<sub>2</sub>O worlds with ice in the basin.

However, from Bruce's perspective, he has no reason to think that what he said violated this conversational maxim. He has no reason to think that what he said would not be understood by his students: he stated in class one day that ice is water in its frozen state, and he was unaware that his students weren't paying attention when he explained this to them. From the students' perspective, they have no reason to think that what he said violated this conversational maxim, either. They take themselves to be competent with 'water' and 'ice', and to be correct in taking them to be members of distinct kinds. So, none of the participants in the conversation have reason to think that his assertion violated this maxim. That means that *neither Bruce nor his students* would have reason to think that his assertion required reinterpretation.

However, we, as theorists, see that, unbeknownst to both Bruce and his students, his assertion *did* violate the conversational maxim to only say things that one's audience understands. Let us understand this to entitle us to use diagonalization to explain why his students took his assertion as a proposal to rule out the actual world and other H<sub>2</sub>O worlds with ice in the basin.

First, let us try to explain it using the fixed metasemantic approach to diagonalization. On this approach, we reinterpret Bruce's assertion at each world in the prior context set, allowing the extension of 'water' to be reassigned to whatever satisfies

the reference-fixer at that world. Suppose, first, that the reference of 'water' was fixed by demonstration, and so the reference of 'water' is reinterpreted as whatever happened to be the thing demonstrated at the time of reference-fixing.

Without any constraints imposed on what might be demonstrated, we could end up with the extension of 'water' being quite different in nature from its extension at the actual world. For example, perhaps where there was a bucket of water in the circumstances in which the reference was fixed, such that the contents of the bucket was what was demonstrated. However, suppose that there is a world in the context set at which there is a shoe where the bucket was in our world, such that at this other world, it is the shoe that is demonstrated.

Luckily, we can avoid this problem because all the interlocutors know that, however water ultimately turns out, at least 'water' refers to a clear odorless liquid that is quite prevalent. This means that all worlds in the context set will be consistent with their knowledge of this, and so all worlds in the context set are such that 'water' refers to a prevalent clear odorless liquid. This means that there won't be any such "shoe" world in the context set.

Regardless of whether the reference of 'water' was fixed by demonstration or description, at any H<sub>2</sub>O world in the context set, using the metasemantics as they are in our world, if H<sub>2</sub>O is in the extension of 'water', then ice is in the extension of 'water'. Given that H<sub>2</sub>O was the substance demonstrated during the reference-fixing event, or satisfied the

reference-fixing description for 'water' at the time of the reference-fixing, and given that 'water' is a natural kind term, such that included in the metasemantic facts for 'water' is that it refers uniformly to all and only samples or members of that kind, then using these same metasemantic facts for 'water' to reassign the extension of 'water' at other worlds in the context set will have the consequence that if H<sub>2</sub>O satisfies the reference-fixer at a world, then ice is in the extension of 'water'. This is because whatever is in the extension of 'water' on the reinterpretation, 'water' picks out all members or samples of that kind uniformly. At worlds at which, on reinterpretation, 'water' picks out H<sub>2</sub>O, then 'water' uniformly picks out all and only H<sub>2</sub>O, including ice. If, on reinterpretation at a world, 'water' is such that ice is in its extension, then if there is ice in the basin at that world, Bruce's assertion, on reinterpretation at that world, is true at that world. Thus, if we reinterpret Bruce's assertion as expressing the diagonal, as determined by fixed metasemantic diagonalization, then what is said on reinterpretation at a world at which the reference of 'water' is fixed to H<sub>2</sub>O, and that is such that there is ice in the basin, is true at that world, and so this doesn't explain why the students ruled out such worlds. If this is the proposition that Bruce expressed, then his assertion is a proposal to rule out from the prior context set any worlds that *aren't* among the set of worlds that is the proposition expressed.

Keeping with the fixed metasemantic reinterpretation approach, another strategy is to say that some world in the context set contains what we'll call 'pseudo-ice'—a substance that is superficially indiscernible from ice, though microphysically distinct from it—in the basin, and it is *that* world that is ruled out by fixed metasemantic diagonalization. This

strategy is intended to capture the way that the students think the world is: they think that liquid water and ice are distinct kinds, and so the pseudo-ice world is one at which there is a substance that is superficially indistinguishable from liquid water—call it pseudo-water—and at which there is pseudo-ice, and that the two are of different kinds. The idea is that, on reinterpretation at such a world, pseudo-ice is not in the extension of ‘water’ because pseudo-water is in the extension of ‘water’ on reinterpretation at that world, and pseudo-ice is of a different kind than pseudo-water. Suppose that we reinterpret his assertion at a world at which there is both pseudo-ice and pseudo-water, that there is pseudo-ice in the basin, and that, on reinterpretation at that world, ‘water’ picks out pseudo-water. If we take it that his assertion expresses the diagonal proposition as determined by fixed metasemantic diagonalization, and there is such a pseudo-ice world in the prior context set, then at such a world, the sentence, as reinterpreted at that world, is false at that world, such that his assertion is false on the diagonal at that world, which

Figure 4

	<b>H<sub>2</sub>O world with H<sub>2</sub>O in pitcher</b>	<b>XYZ world with XYZ in pitcher</b>
<b>H<sub>2</sub>O world with ice in basin</b>	<b>T</b>	<b>F</b>
<b>Pseudo-Ice world</b>	<b>F</b>	<b>F</b>

explains how they could have taken his assertion to be a proposal to rule out that world.

However, recall that the students’ mistake was in taking his assertion as a proposal to rule out those H<sub>2</sub>O worlds at which there is *ice* in the basin, in light of Bruce’s assertion. That

pseudo-ice worlds are ruled out doesn't explain why the students also ruled out the H<sub>2</sub>O worlds at which there is ice in the basin, including the actual world. Even though 'there is water in the basin' is false on fixed metasemantic diagonalization at the pseudo-ice worlds, it is true on fixed metasemantic diagonalization at the H<sub>2</sub>O worlds at which there is ice in the basin, including the actual world. So, this doesn't explain the students' mistake, either. Furthermore, if the students did rule out the pseudo-ice world, they wouldn't be making a mistake since his assertion, on reinterpretation, is false at the pseudo-ice world.

Similarly, it might be that among the worlds in the context set is a world at which XYZ is the predominant clear odorless liquid, but there is a very small amount of H<sub>2</sub>O, too—in fact, suppose that the only sample of H<sub>2</sub>O in that world is in the basin. Suppose that, on reinterpretation at that world, 'water' picks out XYZ. This would be a world at which there is ice in the basin, and at which 'there is water in the basin' is false because, reinterpreted at this world, 'water' picks out XYZ, and there is no XYZ in the basin.

My response to this example is very similar to my response to the pseudo-ice case. If we take his assertion to express the diagonal proposition, such that it serves as a proposal to rule out all worlds in the prior context set at which his assertion is false on the diagonal, then it is true that among the worlds that his assertion proposes to be ruled out is such a world. However, that his assertion serves as a proposal to rule out the XYZ world with frozen H<sub>2</sub>O in the basin doesn't explain why it also led his students to take it to be a proposal to rule out *our* world and other worlds like it, namely, the H<sub>2</sub>O worlds with ice

in the basin. As with the pseudo-ice case, that some other worlds were ruled out doesn't explain why the students ruled out *this* world and other H<sub>2</sub>O worlds at which there is ice in the basin. Also, as with the pseudo-ice case, if we interpret his assertion as expressing the diagonal, then the students would not be making a mistake in taking his assertion to be a proposal to rule out the XYZ world with ice in the basin, as his assertion is false on the diagonal at that world, as we can see in figure 5.

Figure 5

	<b>H<sub>2</sub>O world with ice</b>	<b>XYZ world with ice in basin</b>
<b>H<sub>2</sub>O world with ice in basin</b>	<b>T</b>	<b>T</b>
<b>XYZ world with ice in basin</b>	<b>F</b>	<b>F</b>

Nor is it attractive to attempt to account for the ruling out of the actual world and other H<sub>2</sub>O worlds with frozen H<sub>2</sub>O in the basin by employing the metalinguistic approach to diagonalization. Even though this gives us the resources to offer an explanation of why the students ruled out the world, it doesn't look like a *plausible* explanation of what is going on in this case. In order to offer an explanation of this sort, we'd need to reinterpret Bruce's assertion at a world that differs from our own with respect to the metasemantic facts but not with respect to the way the world is (i.e., an H<sub>2</sub>O world in which there is ice in the basin) because, remember, we're trying to explain why it is that the students ruled out the possibility in which there is ice in the basin. Successfully accounting for this on the metalinguistic approach requires a world at which the metasemantics for 'water' are such that, somehow, ice is excluded from the extension of



‘water’. If there is an H<sub>2</sub>O world with ice in the basin at which the metasemantic facts for ‘water’ are such that ice is excluded from the extension of ‘water’, then we will have an account of why his students took his assertion as a proposal to rule out such a world.

Recall that his students believe that water is the most common clear odorless liquid, and they believe that water is always in its liquid state. This tells us that the worlds that belong in the prior context set are such that the metasemantic facts for ‘water’, for the purposes of reinterpretation, include that its reference-fixing description is ‘the predominant clear odorless liquid’.

If the interlocutors take ‘water’ to be a natural kind term that refers uniformly to whatever kind to which its reference is fixed, then it looks like it isn’t possible for there to be a world at which ‘water’ picks out liquid water but not ice. The important point to note is that so long as the description that we use as the reference-fixing description—as determined by the description that captures the qualities that the interlocutors (in this case, the students) associate with water—picks out H<sub>2</sub>O at the H<sub>2</sub>O worlds in the context set, then given that natural kind terms refer uniformly to all samples of the same kind, ‘water’ will pick out ice, and so ‘there is water in the basin’ will be true at worlds at those H<sub>2</sub>O worlds which there is ice in the basin.

Furthermore, as on the fixed metasemantic approach to diagonalization, it won’t help on the metalinguistic approach to diagonalization to offer an explanation of why the students took his assertion as a proposal to rule out worlds at which there is a predominant clear

odorless liquid that is superficially indistinguishable, but microphysically distinct, from liquid H<sub>2</sub>O, and that there is ice in the basin, and that ice and this liquid are of distinct kinds. For example, as before, suppose that we are evaluating Bruce's assertion at a world at which XYZ is the predominant clear odorless liquid, and though there is H<sub>2</sub>O, it is incredibly rare—in fact, let us keep with our earlier stipulation that the only H<sub>2</sub>O in that world is frozen in the basin. Reevaluated at this XYZ world using the metalinguistic approach, 'water' picks out XYZ, and there is frozen H<sub>2</sub>O, not XYZ, in the basin. As such, Bruce's assertion, as said at this XYZ world, is false at this XYZ world. This explains why his students took his assertion as a proposal that this XYZ world be ruled out. The question that remains is whether explaining why this world was ruled out explains Bruce's students' mistake.

Unfortunately, it does *not* explain the students' mistake. Remember that the students' mistake is in ruling out *the actual world and other H<sub>2</sub>O worlds with ice in the basin*. As in the pseudo-ice case, above, that these XYZ worlds are ruled out doesn't explain why they also ruled out the world and other H<sub>2</sub>O worlds like it. Furthermore, as before, the students wouldn't be making a mistake in ruling out those XYZ worlds—the sentence, as reinterpreted at that world, is false. However, the sentence, as reinterpreted at our world (and at the H<sub>2</sub>O worlds like it), is true at each of those worlds, and so explaining why those XYZ worlds were ruled out doesn't help to explain why the H<sub>2</sub>O worlds were ruled out.

We might, for the purposes of this reinterpretation, allow for the metasemantic facts for ‘water’ to be such that it is not a natural kind term in at least some worlds in the context set. This allows us to consider the hypothesis that the students don’t take ‘water’ to be a natural kind term, and instead take it to refer to the predominant clear odorless liquid in its liquid state, but not in any other state (though they wouldn’t describe the case this way since they do not know that it has other states). On this approach, we hypothesize that the appropriate reference-fixing description for ‘water’ is one such that it ensures that ‘water’ only picks out, say, the predominant clear odorless liquid *in its liquid state*, and, since we are not treating it as a natural kind term, ‘water’ won’t rigidly pick out all samples of H<sub>2</sub>O, but only those in its liquid form. Given this metasemantics for ‘water’, at an H<sub>2</sub>O world at which there is ice in the basin, ‘there is water in the basin’, reinterpreted at that world is false at that world, and so that world is ruled out.

Nonetheless, there is a test that shows that this is an implausible explanation for why the students ruled out the possibility that there is ice in the basin. Suppose that some of the students responded to Bruce’s assertion by saying something that made it clear to Bruce that they’ve taken his assertion to exclude the possibility of ice in the basin. Consider what the students’ reaction would plausibly be if Bruce subsequently informed them that ice *is* water in its frozen state (and suppose that they were paying attention this time). It seems quite plausible that they would no longer take his earlier assertion as a proposal to exclude the actual world and other H<sub>2</sub>O worlds like it. If they take ‘water’ to be a natural kind term, and they learn that ice *is* water, then they will see that they’ve made a mistake in taking his assertion to be a proposal to rule out such worlds. This suggests that their

mistake was not in failing to treat ‘water’ as a natural kind term, but rather in taking ice to not be in the extension of ‘water’—that is, they took ice to be of a different kind—and so this casts doubt on the plausibility of using the metalinguistic approach in this way to explain why they ruled out the ice possibility in light of Bruce’s initial assertion.

One might argue that there is some sort of interesting pragmatic feature of the conversation that explains why the ice possibility was ruled out. If Bruce thought that there was *ice* in the basin—that is, that the water in the basin was frozen—then he would have said so. If saying that there is ice in the basin is a more specific, and, so, stronger, claim than that there is water in the basin, then if he intended for the students to come to believe that there is ice in the basin, he would have made the stronger claim.

There are two considerations here. First, remember that Bruce doesn’t know whether the current temperature is above or below the freezing point, and so might not have been confident about whether, at the current ambient temperature, the water in the basin was a liquid or a solid. More importantly, regardless of Bruce’s beliefs about whether the water in the basin is in its liquid or solid state, the students aren’t in a position to make use of any such pragmatic move. Since they believe that ice and water are two distinct kinds then they won’t take ‘there is ice in the basin’ to be a more specific claim than ‘there is water in the basin’.

These cases allow us to see the importance of not just modeling the force of an assertion as ruling out at least *some* world or other from, or introducing some world or other to, the

context set. This is too weak of a requirement. Stalnaker's view can explain *that* Bruce's students find his assertion informative by appealing to the pseudo-ice world or the XYZ world with very little H<sub>2</sub>O, using these worlds in modeling the prior context set, and then arguing that they are ruled out by diagonalization. However, merely ruling these worlds out does not explain the students' mistake in taking his assertion to be a proposal to rule out the H<sub>2</sub>O worlds with ice in the basin. If the actual world and other H<sub>2</sub>O worlds with ice in the basin are in the prior context set, we cannot, using the resources of Stalnaker's view, explain why the students took his assertion as a proposal to rule out such worlds.

Yet, as we noted earlier, if we consider the students' reaction to being told that ice is water in its frozen state, we can see that they did take his assertion as a proposal to rule out such worlds.<sup>30</sup> If Bruce subsequently asserted that ice is water in its frozen state, the students would very likely see that they had made a mistake and the latter assertion would serve as a proposal to reintroduce those worlds to the context set. Furthermore, it is because the ice worlds were ruled out that it is appropriate for Bruce to subsequently assert that ice is frozen water. If his preceding assertion had not ruled those worlds out, we would not be able to explain why his subsequent assertion was informative. This is because if the H<sub>2</sub>O worlds with ice in the basin had not been ruled out, her subsequent assertion that ice is water in its frozen state would not serve as a proposal to add a hypothesis—that there is ice in the basin—to the context set that wasn't presently under

---

<sup>30</sup> Furthermore, consider how the students would react to looking into the basin after Bruce's assertion. They would be surprised to find a clear solid in the basin (and might suspect that either Bruce was mistaken, or that he was trying to deceive them). This reaction suggests that they did take the ice possibility to have been ruled out by his assertion.

consideration, and so would fail to explain why his second assertion was informative to the students.

This is why it is crucial to concern ourselves with modeling the informativity of an assertion, rather than merely being able to explain that the assertion is informative because we can explain that the interlocutors took it as a proposal to rule out some world or other. Explaining the informativity of assertions by arguing that the interlocutors merely took it as a proposal to rule out some world or other leaves open the possibility that we will be unable to explain the informativity of subsequent assertions, such as why a subsequent assertion made by Bruce, telling his students that ice is water in its frozen state, would be informative

Furthermore, it's important to note that this worry isn't specific to 'water'—we could generate similar cases for proper names or other natural kind terms. This kind of case can arise in cases in which that which they are talking about is presented in an uncharacteristic or unfamiliar way. For example, suppose that, instead of discussing the contents of the basin, Bruce is telling the students about some statues. One of the students ask, "Which is your favorite statue?", and Bruce replies, "I'll give you a hint—my favorite statue is made of copper". His students, who don't know that copper turns green when oxidized, take his assertion as a proposal to rule out the possibility that the Statue of Liberty is his favorite statue. We could, of course, generate a great many more such cases.

There still remain two potential kinds of responses. The first is that what is going on in these kinds of cases is tacit quantifier domain restriction: the quantifier in his assertion is restricted to ranging over the domain of liquid water, rather than all water. The second is that perhaps the H<sub>2</sub>O worlds with ice in the basin weren't *in* the prior context set. If these worlds weren't in the prior context set, then the students couldn't have taken his assertion as a proposal to rule out such worlds from the context set, and so there is no problem to be explained.

Starting with the first of these two responses, perhaps what is going on in this case is tactic quantifier domain restriction. Often, when we use a quantifier—such as *none*, *some*, *many*, or *all*—we intend for that quantifier to be restricted in such a way that it only ranges over some domain that is a subclass of the domain over which it normally ranges. For example, suppose that you and I are loading up my car with things that we are planning to bring with us to a party. When you tell me, in that context, 'All the water is in the car', you mean to be telling me that *all the water that we are taking to the party*, rather than all the water simpliciter, is in the car. That is, using an unrestricted quantifier, 'all the water' would refer to all the water in the domain of the world. In using quantifier restriction, 'all the water' is restricted to a subclass of the members of the domain over which the unrestricted quantifier ranges—in this case, the domain is restricted to the water that we are taking to the party.

Similarly, perhaps there is tacit quantifier restriction in the case of Bruce's assertion. The existential quantifier, expressed by 'there is', is restricted to ranging over only the

domain of *water in its liquid state*, rather than the domain of all of the water in the world. In saying, ‘there is water in the basin’, the domain over which the existential quantifier ranges is restricted to water in its liquid state, as opposed to the domain of all water over which it would have ranged, had the quantifier not been restricted.

There are at least three ways that this kind of response could be developed. The first is that this tacit quantifier domain restriction is part of the semantics of the sentence, such that the token of ‘water’ in his assertion is only referring to water in its liquid state. In such a case, we need only appeal to the standard interpretation, on which what he said is false at the actual world and at other H<sub>2</sub>O worlds at which there is ice in the basin. This allows for an explanation of why his assertion rules out those worlds.

The worry for building the quantifier domain restriction into the semantics is in giving a semantic rule for the existential quantifier such that it picks out some contextually relevant domain of water that is a subclass of the domain of all water, *and* determining some feature of the context such that it makes it the case that the quantifier would be restricted in this way. Regarding the first consideration, there isn’t some salient subclass of water in that context. First, note that there isn’t any such salient subclass of water as far as Bruce is concerned. Recall that Bruce doesn’t know whether it is above or below the freezing point, and so does not know whether the water in the basin is presently in its liquid or solid state. Thus, he won’t intend to use any such quantifier restriction. Furthermore, there is no salient subclass of water for the students. This is because they aren’t in a position to take liquid water to be a subclass of the class of all water—



remember, they think that liquid water *is* all the water, rather than some subclass of the whole class of water. Thus, there is no such salient subclass for them, either.

Still, let us bracket worries about the former consideration—I am more concerned about the latter. There isn't anything about the context that motivates restricting the quantifier to ranging over the domain of liquid water. While it is true that the students don't know that water has solid and gaseous states, and so they expect water to only be presented as a clear odorless liquid, Bruce does not know this, and so he has no reason to restrict the quantifier in this way. Furthermore, it was not the case that the preceding discussion had been in any way focused on water in its liquid state. This suggests that whether or not there is some semantic rule for the existential quantifier that determines that it restricts when there is a contextually relevant domain that is a subclass of the domain over which it would otherwise range, there isn't a feature of the context such that it triggers such a restriction.

The second and third ways of developing the quantifier domain restriction strategy is to take the restriction to be part of the pragmatics, with the second approach developed as arising from the speaker's intension to restrict the quantifier in this way, and the third as arising from the audience taking the speaker's assertion as a proposal that is expressed by making use of the restriction. This means that, in the former case, because Bruce intends to use 'there is water' in such a way that the quantifier ranges only over the domain of liquid water, then his assertion is such that the quantifier does range only over that

domain. In the latter case, the students take his assertion to be such that the quantifier is restricted in this way (whether he, in fact, intended it or not).

The challenge is in motivating these strategies. Consider the case in which you tell me, ‘All the water is in the car’. You intend to use ‘all the water’ to refer to the contextually salient water, namely the domain of water that we mean to take to the party. Taking ‘all the water’ to quantify over the water that we intend to take to the party, and supposing that all that water is in the car, then what you’ve said is true.

But Bruce’s assertion isn’t analogous to this case. As we’ve already noted, he doesn’t intend for ‘water’ to only pick out water in its liquid state. Recall that he knows that the ambient temperature, at the time of his assertion, is around the freezing point, but he doesn’t know whether it is above or below it, such that he is unsure whether the water in the basin is in its frozen state or liquid state. Since he knows that there is water in the basin, but not whether it is a liquid or solid, he didn’t intend to restrict the quantifier such that it ranges over the domain of water in its liquid state.

The second way of developing this approach is to say that even though Bruce didn’t intend to be using tacit quantifier domain restriction, her audience—that is, her students—took her to be doing such a thing, and so, for this reason, took his assertion to rule out the actual world and other H<sub>2</sub>O worlds at which there is ice in the basin. That is to say, they took his assertion as a proposal to rule out from the context set all worlds at which there is a clear odorless liquid in the basin.

In explaining that his students took his use of ‘there is water in the basin’ to pick out only liquid water, it explains why they took what he said to be false at the actual world and H<sub>2</sub>O worlds like it, and so why they took such worlds to be ruled out by his assertion. For this reason, it is an attractive response, provided that it can be motivated. However, I think we have already seen a good reason to think that this can’t be what is going on in this case. While it may very well be right that his students took his token of ‘water’ to refer only to liquid water, it isn’t right to attribute to his students the belief that he was tacitly employing quantifier restriction in this way. Since his students believe that water only has a liquid state, they would not take using ‘some water’ to pick out only a clear odorless liquid to be any sort of restriction: that just *is* what they take ‘water’ to pick out. So, for this reason, they can’t be taking his assertion to involving tacit quantifier restriction.

There is another way that Stalnaker could attempt to avoid the problem: he could argue that the actual world, and other H<sub>2</sub>O worlds like it, aren’t in the prior context set. He could motivate this by arguing that these worlds aren’t consistent with the ways the interlocutors think the world might turn out. At these worlds, the stuff in the basin and the predominant clear odorless liquid are, unbeknownst to the interlocutors, of the same kind—the interlocutors believe that they are not of the same kind. This is a reason to think that such worlds aren’t in the prior context set.

However, this otherwise attractive move comes at a steep cost to Stalnaker's view. For now, not only must we modify the process of diagonalization to be more closely tied to the epistemic state of the interlocutors than the resources of his view allow, we must also change the mechanism by virtue of which assertions introduce new hypotheses to the context set. Here's why: as stated in the passage quoted on the first page of this chapter, an assertion is a proposal to *add the content* of the assertion to the context set. Since the assertion is true at the actual world and the other H<sub>2</sub>O worlds like it, these worlds are among the members of the set of worlds that are the content of the assertion, and so even if these worlds weren't previously in the context set, they would be added, as the content of his assertion, before all the worlds that aren't the way the assertion says the world is are ruled out. Thus, this approach to avoiding the problem requires that Stalnaker reject his claim that assertions are a proposal to add the content of the assertion to the context set, replacing it with a claim such as the assertion is a proposal to add *what the interlocutors take to be the content of the assertion* to the prior context set, which would avoid adding the actual world, and other H<sub>2</sub>O worlds like it, to the prior context set in light of his assertion.

I think that leaving this kind of case unexplained is a significant consideration against Stalnaker's view. Furthermore, avoiding the problem requires giving up at least one, and possibly both, of the steps of the mechanism for updating the context set in light of assertions that he defends in favor of a mechanism much more closely tied to the epistemic state of the interlocutors. As we will discuss in section 7 of chapter five, my

diagnosis of the problem is that the uniform extension intuition, on the understanding of the intuition that Stalnaker employs, cannot do the work that he wants it to do.

### Chapter Three: The Epistemic Interpretation of the Two-Dimensional Framework

In this chapter, we will focus on Jackson's and Chalmers's epistemic interpretations of the two-dimensional framework. We will discuss how Jackson and Chalmers employ the two intuitions in their views. I will argue that, as we saw with Stalnaker's view in chapter two, there are some problematic assertions, the informativeness of which their views cannot accommodate.

As I noted in chapter one, Jackson and Chalmers use the two intuitions to do different work on their views than Stalnaker uses the intuitions to do for him. One difference in the ways in which the intuitions are employed is that Jackson and Chalmers use the contingent extension intuition in a much broader range of cases than Stalnaker does. That is because, unlike Stalnaker, they take the property or description that is used to determine the extension of a proper name or natural kind term on the diagonal as part of the semantics of that expression, and so the evaluation of the statement on the diagonal is part of the semantic interpretation, rather than a feature of reinterpretation.

This is why their interpretations of the two-dimensional framework are categorized as two-dimensional *semantics*.<sup>31</sup> For example, Jackson writes:<sup>32</sup>

---

<sup>31</sup> For example, Stalnaker categorizes them as such in "Assertion Revisited", p. 300 (see especially footnote 5). Jackson identifies as such, especially by making claims such as the quote provided here from "Why We Need A-Intensions". In taking this intension to capture the cognitive significance, Chalmers also seems to take it to be part of the meaning, though he does sometimes say that he is not concerned with what gets called part of the *semantics*.

<sup>32</sup> "Why We Need A-Intensions",

[T]he resilient, known associations are *semantic* properties of the words in the sense that the knowledge in question is part of what's involved in understanding a language that contains the words. People who learn a language like English are able to use it to report on, learn about and discuss a vast range of possible concatenations of properties. What we learn *inter alia* when we learn English is that the word "square" in English is a word to use when we want to discuss the existence of something with a certain property, debate whether or not that property is ever found together with being a diamond bigger than the Ritz, or tell someone what their new desk is like, and so on. [...] But it is plausible that, as a rule, adjectives and common nouns do have resilient, known associations with properties.

Whereas Stalnaker uses these properties or descriptions only whenever the standard interpretation, as understood on his view, would fail to explain why the assertion was informative, Jackson and Chalmers use it in all cases to explain how the audience of an assertion can use the assertion to rule out hypotheses under consideration about the way the world might turn out, or to consider new hypotheses in addition to those already under consideration about the way the world turns out. This amounts to a significant difference between these two interpretations of the two-dimensional framework: while Stalnaker only appeals to the contingent extension intuition for the purposes of pragmatic reinterpretation, Jackson and Chalmers appeal to it as part of the standard interpretation—they take it to capture one kind of semantic content.

### ***Section 1: Jackson and Chalmers Put the Intuitions to Work***

Jackson and Chalmers use the two intuitions to motivate the idea that we can engage in two kinds of evaluations of proper names and natural kind terms, and of assertions in which these expressions are constituents. These involve evaluating these expressions using the same space of worlds in two different ways: evaluating the expressions at a world considered as counterfactual, and evaluating the expressions at a world considered as actual.

The uniform extension intuition motivates the idea that we can, and (at least tacitly) do, engage in counterfactual evaluation of proper names and natural kind terms, and sentences in which they are a constituent. We can evaluate proper names and natural kind terms counterfactually by determining their extension across worlds such that, at each counterfactual world, the proper name or natural kind term in question uniformly picks out the same individual or kind, respectively, as it does at our world, (i.e., at the actual world). Relatedly, we can evaluate counterfactually sentences in which proper names or natural kind terms are a constituent by evaluating what is said at the actual world, holding that fixed, and evaluating whether *that* is true at a world considered as counterfactual. This counterfactual evaluation captures the uniform extension intuition that we have about proper names and natural kind terms: these expressions refer uniformly to the same individual or kind across all counterfactual worlds as they refer to at the actual world.



For example, we can evaluate the extension of ‘water’ counterfactually. Given that the actual world is an H<sub>2</sub>O world (i.e., a world at which that which has the property that we associate with ‘water’ is H<sub>2</sub>O), such that ‘water’ picks out H<sub>2</sub>O uniformly (i.e., picks out all and only H<sub>2</sub>O), then at each counterfactual world, ‘water’ uniformly picks out H<sub>2</sub>O. Similarly, we can evaluate counterfactually an assertion, ‘water is the watery stuff’ by first determining what is said by that assertion, as said at our world: given that our world is an H<sub>2</sub>O world, and so the reference of ‘water’ is fixed to H<sub>2</sub>O, what is said is that there is H<sub>2</sub>O in the oceans. We can then evaluate whether what is said at our world is true at a counterfactual world. Consider, for example, an XYZ world (i.e., a world at which that which has the property that we associate with ‘water’ is XYZ) in which there is no H<sub>2</sub>O: what is said is false at that counterfactual world because there isn’t H<sub>2</sub>O in the oceans.

While it is true that this—that there is H<sub>2</sub>O in the oceans—is what is said by that assertion at our world, it is also true that we were not always in a position to know that it was precisely *this* that was said, because weren’t in a position to know how our world turns out in the relevant respect, namely that water is H<sub>2</sub>O. Furthermore, our epistemic state with respect to what is said by a given assertion in a great many such cases is similarly impoverished: because we don’t know how the world turns out in the relevant respect(s), we don’t know precisely what is said. If we don’t know precisely what is said, then we don’t know precisely which worlds aren’t the way the assertion says the world is, and so we don’t know precisely which worlds are excluded by the assertion.

According to Jackson and Chalmers, we, as theorists, can model how we go about ruling out possibilities (i.e., worlds) in light of an assertion, given our impoverished epistemic state, by using what we will call, following Yablo,<sup>33</sup> counterfactual evaluation. About counterfactual evaluation, Yablo writes:

Recall what we do to judge metaphysical necessity. We ask of various worlds  $w$  whether  $S$  (*our*  $S$ , *natch*) is true of  $w$ . The Kripkean tells us that to judge conceptual necessity, we need to ask, not whether  $S$  is true *of*  $w$ , but whether it is true (as spoken) *at*  $w$ . But maybe it wasn't really necessary to move  $S$  over to  $w$ . A different option is to move  $w$  over to actuality: to the place where the token of  $S$  that we want evaluated in fact occurs. [...] Just as when judging metaphysical necessity, we consider  $w$  as counterfactual, so, when judging conceptual necessity, we consider it as *counterfactual*. We consider it as a hypothesis about what *this* world is like. [...] Evaluating  $S$  with respect to counterfactual  $w$  is asking whether  $S$  holds on the hypothesis that  $w$  is (contrary to what we perhaps think) this very world.

The thought is that counterfactual evaluation proceeds by determining what  $S$  says, on the hypothesis that  $w$  is actual. In order to be able to engage in counterfactual evaluation, as Yablo points out, we need to know what to hold fixed about  $S$  and what to allow to change. Jackson and Chalmers give slightly different answers to the question of what is held fixed about  $S$  for the purposes of counterfactual evaluation, but, for our present purposes, we will say what is common to both of their views, namely, that what is held

---

<sup>33</sup> "Coulda, Woulda, Shoulda", in Gendler, T. and Hawthorne, J., eds., *Conceivability and Possibility* p. 448.

fixed is the property or description that is strongly associated with the proper name or natural kind term that is a constituent of S.<sup>34</sup> We will discuss their views on this more specifically in sections 1.1 and 1.2 of this chapter, respectively.

Chalmers and Jackson are committed to interlocutors themselves using something like counterfactual evaluation (perhaps only tacitly) to rule out possibilities. Counterfactual evaluation captures what is understood by the assertion, according to their view. In other words, on their view, it appears that counterfactual evaluation isn't just something that *theorists* can use to explain why a particular assertion led the audience to update their beliefs about the way that the world is, but rather also the procedure that individuals use (again, perhaps only tacitly) in updating in light of assertions.

We can see that this is something that Jackson and Chalmers think that individuals can do by looking at their discussion of our capacity to entertain hypotheses about the extensions of proper names and natural kind terms, on various hypotheses about how the world turns out in the relevant respects. For example, in “The Foundations of Two-Dimensional Semantics”, Chalmers writes:

The second key idea is that of scrutability: the idea that there is a strong epistemic dependence of an expression's extension on the state of the world. If we come to know that the world has a certain character, we are in a position to conclude that the expression has a certain extension. And if

---

<sup>34</sup> For our present purposes, we will only be concerned with evaluating counterfactually proper names and natural kind terms. We won't be concerned with what stays the same for other expressions, such as qualitative terms, sentential connectives, relational terms, etc. One reason to not be concerned with evaluating these expressions counterfactually is that these expressions aren't twin-earthable.

we were to learn that the world has a different character, we would be in a position to conclude the expression has a different extension. That is: we are in a position to come to know the extension of an expression, depending on which epistemic possibility turns out to be actual.

Similarly, in the opening of his “Two-Dimensional Semantics”, Chalmers writes:

Two-dimensional approaches to semantics start from the observation that the extension and even the intension of many of our expressions depend in some fashion on the external world. As things have turned out, my terms 'water' and 'H<sub>2</sub>O' have the same extension, and have the same (Kripkean) intension. But there are ways things could have turned out so that the two terms could have had a different extension, and a different intension. So there is a sense in which for a term like 'water', the term's extension and its Kripkean intension depend on the character of our world. Given that *this* world is actual, it turns out that 'water' refers to H<sub>2</sub>O, and its Kripkean intension picks out H<sub>2</sub>O in all possible worlds. But if another world had been actual (e.g. Putnam's Twin Earth world in which XYZ is the clear liquid in the oceans), 'water' might have referred to something quite different (e.g. XYZ), and it might have had an entirely different Kripkean intension (e.g. one that picks out XYZ in all worlds).

Both Jackson and Chalmers take it that individuals have the capacity to engage in counterfactual evaluation: if you describe a world to an individual in sufficient detail, and then ask them, on the hypothesis that the described world is the way that the actual world is, what is the extension of a particular proper name or natural kind term, they will be

able to use the property that they associate with that expression in order to determine the extension on that hypothesis (i.e., as that which, at that world, has the property that they associate with that name or natural kind term), as we will discuss in what follows.

Individuals are able to (at least tacitly) engage in counterfactual evaluation of proper names and natural kind terms, and sentences in which these expressions are a constituent, by virtue of strongly associating a property with that proper name or natural kind term. By *strongly associating*, they mean that an individual associates this property with that expression across all contexts. For example, one might strongly associate ‘water’ with the property of being the watery stuff, where the property of being the watery stuff is the property of having the qualitative (e.g., the property of being a clear odorless liquid under warm conditions, and of being a clear solid under cold conditions, and of being steam under hot conditions, and so on) and functional properties (e.g., the property of being the substance that falls as rain; that fills oceans, rivers and lakes; that quenches thirst; that comes out of taps; that we use to bathe, and so on) that we associate with ‘water’.

As we can see in considering the following example, it is important to include the functional characteristics that we associate with ‘water’ in addition to the qualitative characteristics that we associate with it. Consider a hypothesis in which our world turns out such that there is H<sub>2</sub>O on the surface of the planet (i.e., oceans, rivers, rain, etc.), and we intend to use ‘water’ to refer to it by virtue of associating with ‘water’ the property of being the predominant kind that has the qualitative properties that we associate with ‘water’. Suppose that we take it that the H<sub>2</sub>O on the surface *is* the most common clear

odorless liquid, etc., since it is the one we see. However, suppose also that (unbeknownst to us) the core of the world is full of XYZ (i.e., that the entire center of the Earth—everything under the immediate surface—is entirely full of XYZ), and that the ratio of H<sub>2</sub>O to XYZ at that world is actually one part H<sub>2</sub>O to each ninety-nine parts XYZ. At such a scenario, it is the subterranean XYZ, not the H<sub>2</sub>O on the surface, that has the property of being the predominant kind that has the qualitative properties that we associate with ‘water’, and so ‘water’ picks out XYZ. However, in including the functional characteristics, too, we avoid this issue, as H<sub>2</sub>O is the kind that has the property of being the predominant kind that has the functional *and* qualitative characteristics that we associate with ‘water’. Thus, in associating both qualitative and functional properties with ‘water’, if that scenario is actual, we are talking about the H<sub>2</sub>O (which is what we are taking ourselves to be talking about, though we aren’t in a position to know that it is H<sub>2</sub>O), rather than the XYZ (of which we are unaware).

On their views, we associate this property with ‘water’ in all contexts, including both contexts in which we are having a mundane conversation involving water (such as when you tell me that there is water in the refrigerator) as well as contexts in which we are considering an hypothetical scenario quite different from the situation in which we find ourselves (such as discussing how it would have impacted the musical influences that developed in Memphis over the course of the twentieth century if the city were filled with water, much like Venice).

This strong association allows individuals who possess the corresponding concept<sup>35</sup> to understand, in one respect, how the world must be in order for the sentence containing the corresponding term to be true. By *in a respect*, I mean that they take it that associating this property with the term gives the individual access to what we will call *the a priori truth-conditions*<sup>36</sup> for that sentence. On their views, the property provides a condition on which the term applies: the term applies to that which has the property. That is, even if we might not be in a position to know some things about water, such as facts about its underlying composition, we are still in a position to know that water, *however it turns out with respect to its underlying constitution*, has the property of being the watery stuff.

It is for this reason that counterfactual evaluation captures the contingent extension intuition: the extension of the expression is contingent on the way the world turns out in the relevant respects (i.e., the way that water turns out, for the extension of ‘water’). On their views, competence with an expression involves knowing the conditions under which the expression applies, which we can know without knowing precisely which world turns out to be actual. We know the conditions under which the expression applies if we know which property is associated with that expression: the expression applies to that which has the property associated with that expression. For proper names, the name applies to the individual that has the property, and for natural kind terms, the term applies to all samples of the same kind as the kind that has the property.

---

<sup>35</sup> The association seems to be with the concept *and* the corresponding term. They often switch between the level of thought and the level of language, and do not take this to be a concern. (See their comment on this point in their “Conceptual Analysis and Reductive Explanation”, section 3)

<sup>36</sup> Chalmers has called this several different things, including *primary truth-conditions* (in *The Conscious Mind*), *epistemic truth-conditions* (in “The Nature of Narrow Content” and in *The Multiplicity of Meaning*)

Note that I'm not being specific here about exactly what it is to be the property associated with an expression. Chalmers's and Jackson's views differ on this matter, as we will address in sections 1.1 and 1.2 in this chapter. Prior to those sections, we will continue to talk about it in a general way, as the property that individuals associate with the expression. Later in this chapter, in sections three and four, we will challenge the claim that we have a stable association between each proper name or natural kind term and a property.

Recall that the contingent extension intuition shares a feature with the uniform extension intuition, which, in chapter two, we called the unifying feature. According to the contingent extension intuition, once a particular world is considered as actual, that which has the property associated with that proper name or natural kind term is in the extension of that expression, and so that expression picks out that same individual or kind uniformly across all other worlds. For example, suppose that the property that we associate with 'Aristotle' is being the last great philosopher of Antiquity<sup>37</sup>, and that, at the world we consider as actual, the last great philosopher of Antiquity was the man that we call Speusippus, and so the man that we call Speusippus is the referent of 'Aristotle' at that world, and so, on the hypothesis that the world under consideration is actual, 'Aristotle' uniformly refers to Speusippus across worlds *even when Speucippus fails to have the property associated with 'Aristotle'*. Similarly, suppose that the property

---

<sup>37</sup> This is an oversimplification of their view. In at least many cases, the property that we associate with a proper name or natural kind term will include *many* qualitative and functional characteristics that we associate with that expression. I chose this very simple property only for the purpose of keeping the example relatively simple.



associated with 'water' is being the watery stuff, and that, at the world we consider as actual, XYZ is that which has the property of being the watery stuff. On counterfactual evaluation at that world, XYZ is the extension of 'water', and, on the hypothesis that the world under consideration is actual, 'water' uniformly picks out all and only samples of XYZ across worlds, *including samples that fail to have that property*. That is to say, more generally, that just as the uniform extension intuition tells us that, given that 'Aristotle' refers to Aristotle, it refers to him uniformly across all worlds at which 'Aristotle' has any extension at all, and that, given that 'water' refer to H<sub>2</sub>O, it refers to H<sub>2</sub>O uniformly at all worlds at which it refers to anything at all, according to the contingent extension intuition, whatever is in the extension of a proper name or natural kind term on counterfactual evaluation at a world considered as actual, that proper name or natural kind term will refer uniformly to *that* individual or all samples of *that* kind across all worlds at which it picks out anything at all.

Regarding proper names, what this shared feature of the uniform extension intuition and contingent extension intuition shows is that we intend, in using a proper name, to refer to the same individual across worlds, regardless of whether the individual at another world has the property that we associate with the name. For example, suppose that I know that the property associated with 'Aristotle' is the last great philosopher of Antiquity. Suppose, further, that Jackson and Chalmers are right that the reference of 'Aristotle' is mediated by this property, such that it refers to whoever has that property. Finally, suppose that I'm not sure *which* individual 'Aristotle' refers to because I'm not sure which of a few candidate individuals has the property of being the last great philosopher

of Antiquity. Suppose that still epistemically possible for me is a world at which the man we call Aristotle (though I don't know that he is so called) has the property, another world at which the man we call Plato (though I don't know that he is so called) has the property, and a third world at which the man we call Socrates (though I don't know he is so called) has the property. I can entertain the hypothesis that, if the first world is the way the actual world turns out, then 'Aristotle' refers to the man that we call Aristotle (though I wouldn't describe the case that way, since I don't know that he is so called), and that 'Aristotle' refers to *him* at all worlds at which it picks out anything at all, even at those worlds at which he fails to have the property that we associate with 'Aristotle'. Similarly, I can entertain the hypothesis that is the second world is the way our world turns out, then 'Aristotle' refers to the man that we call Plato (though, again, I wouldn't be in a position to describe the case in this way), and that 'Aristotle' refers to *him* across all worlds at which it picks out anything at all, even at those worlds at which he fails to have the property that we associate with 'Aristotle'.

Regarding natural kind terms, that the term applies to all samples of the same kind, rather than merely those samples that happen to have the property associated with the expression at the time a token of that expression is used, captures our intuition that in using a natural kind term, we intend to refer to all and only samples of that kind. For example, if the property associated with 'water' is being a clear odorless liquid that comes out of taps; forms rivers, lakes, and oceans; falls as rain; etc., not all samples of water have this property. Some water is frozen, for example, and so is not a liquid. If we intend to use 'water' as a natural kind term, such that it picks out all and only samples of

the same kind, it must have some feature that allows it to do this by virtue of picking out, in addition to samples that do have that property, samples of the same kind that do *not* have that property. We'll discuss one way of developing this feature of proper names and natural kind terms in section four of this chapter.

As with the 'Aristotle' example for proper names, we can use this understanding of the conditions under which the expression applies to consider various hypotheses about the extension of natural kind terms. Keeping with our familiar 'water' example, suppose that I know that the property of being the watery stuff is the property associated with 'water', and the two hypotheses that I am considering about the way that the substance that has the property of being the watery stuff are that it is H<sub>2</sub>O, and that it is XYZ. I can entertain the hypothesis that if the world turns out to be an H<sub>2</sub>O world, then 'water' refers to H<sub>2</sub>O (and we've been referring to H<sub>2</sub>O all along in our use of 'water'), and that it refers to all and only samples of H<sub>2</sub>O across all worlds at which 'water' picks out anything at all. I can also entertain the hypothesis that if the world turns out, instead, to be an XYZ world, then 'water' refers to XYZ (and we've been referring to XYZ all along in our use of 'water'), and that it refers to all and only samples of XYZ at all worlds at which 'water' picks out anything at all.

Chalmers and Jackson use two *intensions*, or functions from worlds to extensions, to capture these two kinds of evaluation. The *A-intension* (Jackson) or *primary intension* (Chalmers) of a proper name or natural kind term is a function from worlds to extensions that captures the counterfactual evaluation of the proper name or natural kind term. For

example, if the property associated with ‘water’ is the property of being the watery stuff, then the primary intension of ‘water’ is a function from worlds to extensions that picks out at the world considered as actual all and only samples of the same kind as that which has the property of being the watery stuff at that world. If the world considered as actual is one at which H<sub>2</sub>O has the property of being the watery stuff, then the primary intension of ‘water’ picks out all and only H<sub>2</sub>O. If the world considered as actual is one at which XYZ is the watery stuff, then the primary intension of ‘water’ picks out all and only XYZ at that world.

The *C-intension* (Jackson) or *secondary intension* (Chalmers) of a proper name or natural kind term captures the counterfactual evaluation of the proper name or natural kind term. Given that this world—the actual world—is an H<sub>2</sub>O world, and that the property associated with ‘water’ is the property of being the watery stuff, then the secondary intension of ‘water’ picks out all and only H<sub>2</sub>O at all worlds at which it picks out anything at all.

More needs to be said about these intensions on their respective views. Chalmers and Jackson interpret these intensions slightly differently. In section 1.1 we will discuss Jackson’s interpretation of A-intensions and C-intensions, and in section 1.2, we will discuss Chalmers’s interpretation of primary and secondary intensions.

### *Section 1.1: Jackson’s Intensions*

Jackson thinks that the intensions capture the conventions of use for the expression within the language in question<sup>38</sup>, though, here, we will focus only on proper names and natural kind terms. For each proper name or natural kind term, Jackson thinks that there is a property that we strongly associate with that expression, such that we associate it with that expression in all contexts, including not only familiar, everyday contexts, but also contexts in which we are discussing far-fetched hypothetical situations that we have little to no reason to think could actually obtain, as we saw in Jackson's quote about the square.

Jackson thinks that we discover these resilient properties through engaging in conceptual analysis.<sup>39</sup> One considers various situations, on the hypothesis that the situation is actual, and considers whether the expression applies in that situation and, if so, to what it applies. Once one has done this for multiple cases, some commonalities will emerge regarding the properties had by those things to which one judged that the term would apply in each case. Whatever property is had by all those things is the property that one strongly associates with the expression.

Note that doing conceptual analysis does not require us to use our current epistemic state. In at least some cases, it would be more useful to use a more impoverished epistemic state. For example, in doing conceptual analysis for 'water', if one already knows that water is H<sub>2</sub>O, it isn't epistemically possible that water might turn out to be XYZ, or

---

<sup>38</sup> It seems quite plausible to understand Jackson to be making claims about natural languages and, more specifically, about English. One can get this impression from observing that he says things about how we communicate in the sorts of familiar daily communicative exchanges that he uses for examples, such as informing one another of the time that the train arrives, or telling another what her new desk looks like.

<sup>39</sup> See, for example, chapter two of *From Metaphysics to Ethics*.

anything else other than H<sub>2</sub>O. However, one can bracket that knowledge and consider an epistemic state that one would be in if one had the same superficial experiences of water, but did not yet know its underlying structure. Imagine such a hypothetical case in which the year is 1800, and one is among the chemists researching the underlying chemical composition of water. Suppose that there are two competing hypotheses: that water is H<sub>2</sub>O, and that water is XYZ. In entertaining the supposition that that one is in that epistemic state, one can consider two hypotheses about the way the world might turn out, for all one knows, including the way the extension of our expression ‘water’ turns out. One can entertain the hypothesis that, if water turns out to be H<sub>2</sub>O then ‘water’ picks out all and only H<sub>2</sub>O, and one can entertain the hypothesis that, if water turns out to be XYZ then ‘water’ turns out to be XYZ. One can then consider the relevant similarities between H<sub>2</sub>O and XYZ<sup>40</sup>, namely their qualitative and functional characteristics. Whatever property is had by both is the property associated with the A-intension for ‘water’.

Ideally, one would consider a great many hypothetical cases (while adopting the supposition that one did not already know that water is H<sub>2</sub>O), and then consider the property had by all and only those samples in each hypothetical case were such that, if that case turns out to be actual then that sample would be water. This helps to avoid accidentally including properties that one might not think are relevant. For example, suppose that a particular fellow—let’s call him Jack—has spent his whole life in the UK.

---

<sup>40</sup> Or, at least how one represents XYZ to be. What matters for these purposes is that XYZ is hypothesized to be (superficially) qualitatively and functionally indistinguishable from H<sub>2</sub>O—the only difference is regarding their underlying constitution. It is intended to represent an alternative hypothesis about the way that water could turn out, for all the pre-discovery chemists knew.

When he engages in conceptual analysis about ‘water’, he might think of the Thames, of St. Nectan’s Kieve, and of Loch Ness. The problem is that not only do the samples in these locations have in common the property of being a clear odorless liquid, they also share the property of being a clear liquid *in the United Kingdom*. But, surely, having the property of being in the United Kingdom is not part of the convention of using water, and it is very plausible that it isn’t the property that Jack associates with ‘water’. Jack might have only done conceptual analysis on cases that are (superficially) qualitatively indistinguishable from his own experiences with water, but, if asked, he would presumably acknowledge that it would accord with the conventions of use for ‘water’ to say that there is water in, say, the Mediterranean. That he would say that the substance in the Mediterranean is in the extension of ‘water’ in that scenario demonstrates that he *doesn’t* associate the property of being the watery stuff in the UK with ‘water’, but, rather, the property of being the watery stuff.

One might wonder why this would be a problem, given that the property associated with a natural kind term has the unifying feature. Even if the property that Jack associates with ‘water’ *were* the property of being the watery stuff in the UK, since the property also has the unifying feature, such that it picks out all and only samples of the kind that is such that some samples are picked out by the property that Jack associates with it, then it will pick out all and only H<sub>2</sub>O. However, the property of being the watery stuff in the UK does not capture his grasp of the conventions of use for ‘water’, nor does it capture the cognitive significance of ‘water’ for Jack: though the cases of water that most readily come to his mind are those that he’s seen before, as note earlier, he would also

acknowledge that there is water in the Mediterranean, and in the Pacific, and in the Great Lakes of North America, and so on. If that property captured his grasp of the conventions of use of 'water' or the cognitive significance of 'water' for him, then he would not judge these samples outside the UK to be water. For this reason, it is important to consider a very large range of cases so that conceptual analysis can reveal the property that he associates with 'water'.

There are other potential worries in the vicinity, such as if Jack makes a misidentification of water based on how a mistaken belief he has regarding how water will present under certain circumstances. For example, if Jack doesn't know that water has a solid state, then in imagining a very cold scenario in which there is a clear odorless liquid, he will say that nothing in that scenario is water. Thus, if conceptual analysis reveals the property that *Jack* associates with 'water', and if that property is the property of being the watery stuff, then since each of us will associate different properties with 'water', there will be a different property of being the watery stuff determined by each individual's conceptual analysis. Perhaps it would be better to call each the property of being the watery stuff<sub>name</sub>, including the name of the individual who had engaged in the conceptual analysis that revealed that property. As we will see later in this chapter, using the property revealed by an individual's conceptual analysis will encounter subsequent difficulties.



Jackson thinks that, for at least some terms, the property that we associate with an expression defers to the relevant experts or to their best theory. For example, in *From Metaphysics to Ethics*, he writes:

In case of a theoretical term like this, before we know what the best theory says (and, of course, sometimes we never know what the best theory says; it eludes us), we have no paradigm cases for the term. It might be thought that we will not be able to give necessary and sufficient cases of K-hood in such cases before the relevant theory is with us. [...] But all that follows is that we need to state them in long disjunctions of longish conjunctions of the following kind:  $x$  is a fish if and only if (the best true theory of biology says that the important properties out of or descended from or explanatory of  $F_1, F_2, F_3, \dots$  are so-and-so, and  $x$  has so-and-so) or (the best true theory of biology says that the important properties out of or descended from or explanatory of  $F_1, F_2, F_3, \dots$  are such-and-such, and  $x$  has such-and-such) or..., where  $F_1, F_2, F_3, \dots$  are the properties that we initially associate with being a fish (the properties of exemplars)

Here, we can see that, though we take fish to be the things that have properties  $F_1, F_2, F_3, \dots$  we also intend 'fish' to pick out all and only the members of the natural kind that is such that at least some members (the ones we take to be exemplars) have properties  $F_1, F_2, F_3, \dots$ . From this, we can see that, according to Jackson's account, we intend to refer uniformly to all and only fish in using 'fish', i.e., that the property that we associate with 'fish' has the unifying feature.

The property that we discover through conceptual analysis is the property that captures (presumably what we take to be) the conventions of use for the expression for our language.<sup>41</sup> We needn't worry here about whether precisely the same property is associated with the expression for each speaker—so long as there is sufficient similarity between these properties that it allows for successful communication, this is adequate to serve as capturing their grasp of the conventions of use.

That alone, however, is insufficient to capture the conventions of use of the expression for the language. Since different speakers may associate (slightly) different properties with the expression in question, which, if any of those properties grounds the convention of use? It seems that none of the properties has greater claim than the others. Perhaps an answer to that challenge is that the property that determines the convention of the use for 'fish' is the property that is common to all of the samples that each speaker would judge, during conceptual analysis, to be in the extension of 'fish', on the hypothesis that the situation imagined turned out to be actual. That is, perhaps the convention of use is to use 'fish' to pick out all and only members of that natural kind that has the property that is commonly associated with fish, or to pick out all and only members of that natural kind that has the property that is in common among the exemplars for each individual (i.e., that property had by all exemplars for each individual), where this property is determined as the one that is in common across all samples that, in each scenario considered by each individual in engaging in conceptual analysis, each individual would judge to be in the extension of 'water'.

---

<sup>41</sup> For a short discussion of this, see Jackson's "Why We Need A-Intension".

That property determines the A-intension and the C-intension for Jackson. The A-intension for a proper name picks out, at each world, that individual that has the property associated with that name. The A-intension for a natural kind term picks out, at each world, all and only those samples of the kind that is such that at least some members of that kind have the property.

C-intensions for proper names and natural kind terms are *actually-involving*: they pick out that individual or members of that kind, respectively, that have the relevant property *at the actual world*. The C-intension for a proper name picks out, at each world, that individual that has the property *at the actual world*. For example, suppose, again, that the property that we associate with ‘Aristotle’ is being the last great philosopher of Antiquity. The C-intension for ‘Aristotle’ picks out, at each world, the individual who has the property of being the actual last great philosopher of Antiquity. Since the individual who has the property at the actual world of being the last great philosopher of antiquity is Aristotle, the C-intension of Aristotle picks out Aristotle at each world.

Similarly, if the property that we associate with ‘fish’ is being a scaly cold-blooded vertebrate that lives in water and moves around by swimming—let us abbreviate as the property of being F—then the C-intension of ‘fish’ picks out, at each world, all and only samples of the kind that has the property of actually being F. Since the kind that is such that at least some members have the property of being F at the actual world is the kind fish, then at each world, the C-intension of ‘fish’ picks out all and only fish.

Unlike C-intensions, A-intensions are not actually-involving. Rather, at each world, they pick out that which has the property associated with the expression *at that world*. For example, the A-intension of ‘Aristotle’ picks out at each world that individual that has the property of being the last great philosopher of Antiquity at that world. At a world at which Plato is the last great philosopher of Antiquity, ‘Aristotle’ picks out Plato.

The A-intension for ‘fish’ picks out, at each world, that which has the property at that world of being F. At a world at which some other kind of sea creatures have property F, then, at that world, the A-intension for ‘fish’ picks out all and only members of that kind of sea creature.

One might reasonably worry, given that Jackson’s view is that there are two kinds of meaning or intensions, that this allows for the possibility of ambiguity for expressions. However, Jackson reminds us<sup>42</sup> that the A-intension of an expression and the C-intension of an expression will always have the same extension at the actual world, and so will never allow for ambiguity. For example, the conditions under which ‘there are fish in the oceans’ and the conditions under which ‘there are actual fish in the oceans’ are the same. We will return to discussing A-intensions after discussing the two intensions—primary and secondary intensions—on Chalmers’s view.

### *Section 1.2: Chalmers’s Intensions*

---

<sup>42</sup> For example, see his discussion point two about actually involving expressions on p 260 of “Why We Need A-Intensions”.

On Chalmers's view, the primary and secondary intensions of expressions capture two different ways the extensions of expressions depend on the way the world turns out.

About this, he writes<sup>43</sup>:

First, the actual extension of an expression depends on the character of the actual world in which an expression is uttered. Second, the counterfactual extension of an expression depends on the character of the counterfactual world in which the expression is evaluated. Corresponding to these two sorts of dependence, expressions correspondingly have two sorts of intensions, associating possible states of the world with extensions in different ways. On the two-dimensional framework, these two intensions can be seen as capturing two dimensions of meaning.

The primary intension of expressions captures the way the extension of the expression depends on the way the actual world turns out. This is precisely the contingent extension intuition: the extensions of our expressions are determined in part by the way this world turns out. If, for example, the watery stuff turns out to be H<sub>2</sub>O, then 'water' turns out to refer to H<sub>2</sub>O, and if the watery stuff instead turns out to be XYZ, then 'water' turns out to refer to XYZ. The secondary intension of an expression captures the way that the extension of the expression is determined across counterfactual worlds. Given that this world is an H<sub>2</sub>O world, water refers to H<sub>2</sub>O at all worlds, including at a counterfactual world at which H<sub>2</sub>O is *not* the watery stuff.

The secondary intension of a proper name or natural kind term captures the uniform extension intuition: given the extension of a proper name or natural kind term, that

---

<sup>43</sup> "The Foundations of Two-Dimensional Semantics", p. 59

expression refers uniformly to the same individual, or all and only samples of the same kind, respectively, at all worlds. Given that ‘Aristotle’ refers to Aristotle, the secondary intension of ‘Aristotle’ picks out Aristotle at all worlds. Similarly, as we saw, given that ‘water’ refers to all and only H<sub>2</sub>O, the secondary intension picks out all and only H<sub>2</sub>O at each world.

Furthermore, the secondary intension of an expression serves as its contribution to the truth-conditions for assertions in which it is a constituent expression.<sup>44</sup> For example, the secondary intension of ‘Aristotle’ serves as its contribution to the truth-conditions of an assertion of, ‘Aristotle was the teacher of Alexander the Great’. This assertion is true just in case Aristotle was the teacher of Alexander the Great, and so it is true at all worlds that are the way the assertion says the world is, and false otherwise. Similarly, the secondary intension of ‘water’ serves as its contribution to the truth-conditions of an assertion of, ‘Seventy percent of the Earth’s surface is covered by water’. This assertion is true just in case seventy percent of the Earth’s surface is covered by water, and so it is true at all worlds that are that way, and otherwise false.

Knowing the truth-conditions of an assertion allows one to know how the assertion says the world is (or how it is not), and so which possibilities are excluded by the assertion. However, often, there is an important sense in which we aren’t in a position to know precisely how the assertion says the world is, and so we aren’t in a position to rule out worlds on the grounds that they aren’t the way the assertion says the world is. For

---

<sup>44</sup> Sometimes, Chalmers refers to these as the *metaphysical truth-conditions*, presumably in contrast with the *epistemic truth-conditions* to which the primary intensions of expressions contribute. See “Two-Dimensional Semantics”.

example, if I know that water is the most common clear odorless liquid (perhaps along with other facts about it, including functional characteristics such as that it falls as rain, comes out of taps, constitutes oceans, rivers, and lakes, and so on), but not that it is H<sub>2</sub>O, then there is a sense in which I know how an assertion, ‘Seventy percent of the Earth’s surface is covered by water’, says the world is, and a sense in which I do not. I don’t know enough to know that it says that the world is such that seventy percent of the Earth is covered in H<sub>2</sub>O. I do know enough to know that it says that the world is such that seventy percent of the Earth is covered in the watery stuff, whatever the underlying constitution of it turns out to be.

In cases such as this, we often are in a position to rule out *some* possibilities in light of the assertion. In the present example, I’m in a position to rule out those worlds at which it is not the case that seventy percent of the Earth’s surface is covered in the watery stuff at that world.

Chalmers’s primary intensions can offer one explanation of how we can rule out worlds in these kinds of cases. The property associated with the primary intension of an expression is discovered through conceptual analysis for that term, in the same way that, on Jackson’s account, the property associated with the A-intension of an expression is discovered through conceptual analysis for that term. Furthermore, conceptual analysis proceeds the same way on Chalmers’s account as it does on Jackson’s account, so we won’t revisit our earlier discussion of conceptual analysis here.

According to Chalmers, the primary intension of an expression captures the cognitive significance of that expression. The primary intension captures what we understand when the expression is used, and gives us access to what Chalmers calls the *epistemic truth-conditions* for assertions in which the expression is a constituent. About the epistemic truth-conditions, he writes:<sup>45</sup>

The primary intension of an utterance yields a condition under which the utterance will be true. For example, the primary intension of 'there is water in the glass' will be true at some scenarios and false at others, and the utterance will be true iff the primary intension is true at the scenario of the utterance (roughly, if the glass picked out by the individual at the center of the scenario contains the dominant watery stuff in the environment around the center). This can be seen as an *epistemic* truth-condition for the utterance, specifying how the truth of the utterance depends (epistemically) on which epistemically possible scenario turns out to be actual.

I understand these epistemic truth-conditions to serve as a kind of condition for ruling out possibilities that are still epistemically possible for the hearer of the assertion. The hearer may not know enough about the way the world is to know precisely how the world must be in order for the assertion to be true, yet she can still use the assertion to pare down the space of ways the world might be, for all she knows. The idea is that she can use these epistemic truth-conditions to rule out those worlds that aren't the way the assertion says the world is, according to the epistemic truth-conditions, and the epistemic truth-

---

<sup>45</sup> "Two-Dimensional Semantics"



conditions specify the way in which the world must be in order for the assertion to be true in a way that *doesn't* specify in full detail the way in which the world must turn out.

Suppose, for example, that, for a hearer of an assertion of 'Aristotle was the teacher of Alexander the Great', the cognitive significance of 'Aristotle' is the last great philosopher of Antiquity, and that both she has good reason for believing that Aristotle was the last great philosopher of Antiquity, and that this is all she knows about Aristotle. Since being the last great philosopher of Antiquity is the cognitive significance of 'Aristotle' for her, her primary intension of 'Aristotle' picks out, at each world still epistemically possible for her, the individual who is the last great philosopher of Antiquity at that world. Furthermore, because being the last great philosopher of Antiquity is what she associates with 'Aristotle', she can use this to determine the way the world must be in order for the assertion, 'Aristotle was the teacher of Alexander the Great', to be true: regardless of which of the worlds that are still epistemically possible for her turn out to be actual, in order for that assertion to be true, the world must be such that the last great philosopher of Antiquity was the teacher of Alexander the Great. This means that she can exclude any worlds at which anyone other than the last great philosopher of Antiquity was the teacher of Alexander the Great, as well as any worlds at which Alexander the Great didn't have a teacher, and so on. However, since she isn't sure yet which individual was the last great philosopher of Antiquity—though she knows that, whoever he is, he is called 'Aristotle'—she isn't yet in a position to rule out a world at which Plato is the last great philosopher of Antiquity, or a world at which Heraclitus was the last great philosopher of Antiquity, and so on.

The case is similar for natural kind terms. Suppose that the same individual hears the assertion, ‘Seventy percent of the Earth’s surface is covered in water’. Suppose, also, that the cognitive significance of ‘water’ for her is the watery stuff, and that she does not know the underlying constitution of water. Her primary intension of ‘water’ will pick out, at each world, all and only samples of the kind that is such that it is the watery stuff. Furthermore, she has access to her epistemic truth-conditions for that assertion: the assertion is true just in case seventy percent of the Earth’s surface is covered in the watery stuff. Since she doesn’t know the underlying constitution of water, she is not in a position to know precisely how the assertion says the world is. However, she is in a position to know that the assertion is true if the watery stuff covers 70 percent of the Earth’s surface. This allows her to rule out any possibilities still epistemically possible for her at which it is not the case that seventy percent of the Earth’s surface is covered with the watery stuff at that world. This also allows her to know that if the actual world is an H<sub>2</sub>O world, then the assertion is true just in case seventy percent of the Earth’s surface is covered in H<sub>2</sub>O, and that if the actual world turns out to be an XYZ world, then the assertion is true just in case seventy percent of the Earth’s surface is covered in XYZ, and so on, for each way the world could still turn out, for all she knows.

As with Jackson’s account, since Chalmers’s view allows for two kinds of intensions, one might reasonably worry that this allows for the possibility of ambiguity. More specifically, any expression for which the extension of the primary intension at the actual world differs from the extension of the secondary intension at the actual world would be

ambiguous. Thus, it is imperative that the extensions of the two intensions are the same at the actual world.

Recall that both the primary intension and the secondary intension have the unifying feature: they refer to all and only samples of the same kind. This means that, if the property associated with the primary intension of a proper name picks out the same individual as the secondary intension of that proper name then that proper name will not be ambiguous. Similarly, if the property associated with the primary intension of a natural kind term picks out at least some samples of the relevant kind (and no samples of any other kind), and so the primary intension of that expression picks out all and only samples of that kind, then the primary intension and the secondary intension will have the same extension at the actual world.

Furthermore, if the view is that the extension of the things that have the property associated with the primary intension is either co-extensive at the actual world with the extension of the secondary intension, or, in the case of natural kind terms, picks out at the actual world at least some samples of the same kind as the kind that is in the extension of the secondary intension of that expression, then we can see how one might find it plausible that the primary intension could give us access to the truth-conditions of the assertion. If the property that I associate with 'water' is the watery stuff, and at least some samples of water have the property of being the watery stuff, then, by virtue of having the unifying feature, the primary intension of 'water' picks out all and only H<sub>2</sub>O. Since the secondary intension of 'water' picks out all and only H<sub>2</sub>O, and the secondary

intension of 'water' is its contribution to the truth-conditions, and since the two intensions are co-extensive, then if I know what is picked out by the primary intension, this gives me access to the epistemic, or a priori, truth-conditions.

However, if the property associated with the primary intension doesn't uniquely pick out the same individual or kind as is picked out by the secondary intension, or if the primary intension picks out a different individual or kind, then the expression is ambiguous. Suppose that the property that one associates with 'Aristotle' is an ancient Greek philosopher who is commemorated with a marble bust at the local library. However, suppose that Plato is also commemorated with a marble bust at the local library. This means that the property that one associates with 'Aristotle', and that captures the cognitive significance of 'Aristotle', doesn't uniquely pick out Aristotle, and so the primary and secondary intensions of 'Aristotle' differ in extension, thereby making 'Aristotle' ambiguous. Similarly, suppose that the property that one associates with 'Proxima Centauri' is the nearest star to the earth. The secondary intension of 'Proxima Centauri' is Proxima Centauri, but because that which has the property associated with the primary intension is the Sun, the two intensions have a different extension.

There is another way, consistent with Chalmers's view, to respond to the ambiguity worry that also avoids the unique property worry. One could argue that the primary intension captures the cognitive significance of the expression, though does not determine the extension of the expression, and in no way contributes to the truth-conditions of the sentence. One can then think of the primary intension, and the epistemic truth-conditions

to which it contributes, as a means of approximating the truth-conditions of the sentence. On this way of responding to these worries, the primary intension doesn't play a role in determining the extension of the expression (save perhaps if it played a reference-fixing role), and so doesn't contribute to the truth-conditions. However, it does capture one's best guess about the truth-conditions of assertions, given one's impoverished epistemic state (e.g., as in our earlier example, knowing how the world turns out in some respects while still being ignorant of other relevant facts, such as knowing that water is the watery stuff without knowing that it is H<sub>2</sub>O).

In order to still properly call this a variety of two-dimensional semantics, one would have to argue that the primary intension of an expression captures the cognitive significance of the expression but does not determine the extension of the expression, and so does not play a role in determining the truth-conditions of assertions in which it is a constituent. Is capturing the cognitive significance sufficient for counting as part of the semantics? What other work, if any, can a primary intension do, such that it deserves to be called part of the semantics? Perhaps the best case that can be made is that primary intensions play a central role in modeling communicative exchanges, as we will see in the remainder of this chapter. Though I do not think this is an approach that he would adopt, we will explore this way of thinking about primary intensions in more detail in chapter six.

## *Section 2: Updating Hypotheses About the Way the World is in Light of Assertions*

It is our ability to entertain hypotheses about the way the extension of our expressions will turn out given various ways our world could turn out, for all we know, that enables us to rule out hypotheses about the way the world is, or to come to consider new hypotheses about the way the world is, in light of an assertion. Structurally, the procedure looks a lot like diagonalization on Stalnaker's account—the difference is that, on Jackson's and Chalmers's account, this is not a matter of *reinterpretation*, as it is on Stalnaker's account, but rather a semantic feature of the language,<sup>46</sup> and that we undergo this process all the time (perhaps only tacitly) as language users, rather than it being a way in which *theorists* model special cases but that is not the process that language users actually undergo.

As discussed in the preceding section, Jackson and Chalmers favor views on which a property is associated with a proper name or natural kind term, and an individual is competent with that proper name or natural kind term if she associates that property with that expression, such that she knows that, regardless of which of the worlds still epistemically possible for her turn out to be actual (in the epistemic sense of 'turn out'), the expression applies to that which has the property, whatever that thing that has the property turns out to be.

I can use this ability to engage in counterfactual evaluation to rule out some hypotheses about the way the world is in light of some assertions, and also to introduce new hypotheses for consideration. Before discussing this, however, it is important to note that

---

<sup>46</sup> Stalnaker has a nice discussion of this difference between his metasemantic interpretation, and Jackson's and Chalmers's semantic interpretations, in "Assertion Revisited" (page 301).

Jackson and Chalmers do not share Stalnaker's commitment to using externalism about content to exclude worlds such as the XYZ worlds from the context set. Recall that, on Stalnaker's view, such worlds are only brought into the explanation, i.e., of how the state of the conversation changes in light of assertions, as theoretical tools for the purposes of the pragmatic reinterpretation in the cases in which the standard interpretation won't explain why the assertion is informative. Jackson and Chalmers do not share Stalnaker's externalism about content, and so allow for the possibility that individuals can entertain hypotheses about XYZ worlds, even though they aren't in causal contact with XYZ.<sup>47</sup>

For example, suppose that, as in our earlier example in chapter two, I offer my guests a drink, gesturing at an opaque pitcher. Suppose that the guests know that the only beverages I have in the house are water, lemonade, and chardonnay. Thus, for all they know, the contents of the pitcher could turn out such that there is water in the pitcher, or there is lemonade in the pitcher, or there is chardonnay in the pitcher, and so there is at least one world in context set for each of these possibilities. Furthermore, since, for all they know, water could turn out to be H<sub>2</sub>O or XYZ, and so there is at least one world at which there is H<sub>2</sub>O in the pitcher in the context set, and at least one world at which there is XYZ in the pitcher.

When one of my guests asks me what's in the pitcher, I assert, 'there is water in the pitcher'. How do they go about ruling out possibilities about the way the contents of the

---

<sup>47</sup> Though, if Chalmers adopts his view of worlds as maximal coherent hypotheses qualitatively described, then, in entertaining such a hypothesis, one isn't thereby thinking about XYZ, but rather, at most, a scenario that qualitatively describes an XYZ world.

pitcher could turn out in light of my assertion? They engage in counterfactual evaluation, as represented in figure 1.

Figure 1

	<b>H<sub>2</sub>O</b>	<b>XYZ</b>	<b>Lemonade</b>	<b>Chardonnay</b>
<b>H<sub>2</sub>O</b>	<b>T</b>	<b>T</b>	<b>F</b>	<b>F</b>
<b>XYZ</b>	<b>F</b>	<b>T</b>	<b>F</b>	<b>F</b>
<b>Lemonade</b>	<b>F</b>	<b>F</b>	<b>F</b>	<b>F</b>
<b>Chardonnay</b>	<b>F</b>	<b>F</b>	<b>F</b>	<b>F</b>

My guests are able to rule out those worlds at which what is said at that world is false at that world, i.e., those worlds that are false on the diagonal. Since that which is said at the lemonade world is false at the lemonade world, and that which is said at the chardonnay world is false at the chardonnay world, those worlds are ruled out by my assertion. However, since that which is said at the H<sub>2</sub>O world is true at the H<sub>2</sub>O world, and that which is said at the XYZ world is false at the XYZ world, both of those are still live possibilities for ways the contents of the pitcher turn out. The guests are in a position to know that there is water in the jug, even though they don't know which way water turns out (e.g., as H<sub>2</sub>O or as XYZ).



However, as I will argue, we will need, not one, but *two* properties associated with primary intensions and with secondary intensions.<sup>48</sup> In order to accommodate the contingent extension intuition in the way that Jackson and Chalmers wish to do, we will need a first-order property and a second-order property each associated with the primary intension and the secondary intension of proper names and natural kind terms. These two properties will do different work for us.

The first-order property is the property that is associated with a token of an expression that captures the cognitive significance of the token of the expression. This is the work that is intended to be done with the properties commonly offered as examples by Jackson and Chalmers, such as, for a token of ‘Hesperus’, the property of being the brightest celestial body visible just after dusk, or, for a token of ‘water’, the property of being the watery stuff.

The second-order property is the property that ensures that, in using a proper name or natural kind term, we do not only pick out the individual when that individual has the first-order property, or pick out only those samples of the kind that has the first-order property, but rather pick out the individual even when that individual lacks that property, and all samples of the same natural kind including those that do not have the first-order property. This second property is what provides, as we have been calling it, the unifying feature of these expressions.

---

<sup>48</sup> Thanks to Paul Pietroski for helpful discussion here.

### ***Section 3: First-order property***

Let us say that the first-order property associated with an expression is that which captures the cognitive significance of the expression (on Chalmers's account), or the individual's grasp of the conventions of use for the expression (on Jackson's account), and is revealed through conceptual analysis. This property fixes the reference of the expression at each scenario considered as actual for the purposes of conceptual analysis. We can think of this property as playing a *reference-fixing role*: it is playing the same role that the initial description or property that was used to fix the reference of our term. The thought is that the description or property that was initially used to fix to reference of a term, e.g., 'water', makes use of a manifest property or cluster of properties exhibited by water, plausibly including qualitative and/or functional properties that the person who introduced the term associates with water. For example, suppose that the property used to fix the reference of 'water' was the description 'the predominant clear odorless liquid that fills the oceans, rivers, and lakes; that falls as rain; that quenches thirst; and so on. Using that property to introduce 'water', the person who introduced the term is in a position to know a priori that the stuff in the extension of 'water' has that property.

Similarly, the property that is revealed through an individual's conceptual analysis to be stably associated with that expression will likely also make use of manifest qualitative and/or functional properties that the individual associates with 'water', and so can play the same reference-fixing role at each scenario. The thought is that, just as knowing the reference-fixing property or description for a term allows the person who introduced the term to know a priori that whatever is in the extension of that term has that property,

similarly the person who is engaging in conceptual analysis knows a priori that whatever is in the extension of the expression at that world has that property that she associates with that term.

Keeping with our ‘water’ example, the first-order property associated with ‘water’ might be *being the watery stuff*. Since this property mediates the reference of the expression, then if we know which property is associated with an expression (which, intuitively, it seems plausible to think that we know it, if we associate that property with the expression), then if we know enough about the way the world is, namely, which things have the property that we associate with the expression, we’ll know what our expressions refer to.

Furthermore, Jackson and Chalmers tell us that knowing which first-order property is associated with an expression gives us a priori access to the truth-conditions of at least some sentences in which that expression is a constituent. For example, in *The Conscious Mind*, Chalmers writes:<sup>49</sup>

We can see this in a different way by noting that there are two sets of *truth conditions* associated with any statement. If we evaluate the terms in a statement according to their primary intension, we arrive at the *primary truth conditions*: of the statement: that is, a set of centered possible worlds in which the statement, evaluated according to the primary intensions of the terms therein, turns out to be true. The primary truth conditions tell us how the actual world has to be for an utterance of the statements to be true

---

<sup>49</sup> P. 63

in that world; that is, they specify those *contexts* in which the statement would turn out to be true.

This allows us to know some sentences to be true a priori in the sense that we can know them to be true independently of knowing how the world turns out in some relevant respect. For example, if we know that the property associated with ‘water’ is being the watery stuff then we can know ‘water is the watery stuff’ is true independently of knowing how the watery stuff turns out (i.e., whether it turns out to be H<sub>2</sub>O, or XYZ, or something else). Thus, on their views, we can know ‘water is the watery stuff’ to be true a priori, in this sense of a priority.

For the reasons under discussion, we can see the importance of being able to know which property is associated with each expression. As we’ve discussed, Jackson and Chalmers argue that conceptual analysis is the appropriate means of determining the property associated with the expression. Chalmers and Jackson tell us that the association between a property and that expression is stable or resilient, such that it is by virtue of something in a given scenario, considered as actual, having that property that we say that the expression applies to that thing in that scenario.

Before proceeding, I will make clear three qualifications about the worries that I will raise about appealing to conceptual analysis as a means of revealing the property that is purportedly associated with a proper name or natural kind term. My first qualification is that, for our purposes, we will understand scenarios to be maximally specific. Thus, any

uncertainty about the extensions of our terms at a given scenario will not be due to lack of details about the scenario in question: all the facts will be in, so to speak.

My second qualification is that we will restrict our discussion of conceptual analysis only to the consideration of scenarios that are superficially indistinguishable from our own (though differing from our own in the underlying constitutions in some respects). Further worries could be raised by considering scenarios that differ in appearance from our own, and the more significant the differences in appearance, the worse the worries that can be raised. For example, it seems very reasonable to think that the more that a scenario differs in appearance from our own, the less clear our intuitions become regarding what, if anything, is in the extension of one of our proper names or natural kind terms. I will leave such worries outside the scope of the present project—I can raise my intended worries by appealing to only those scenarios qualitatively indistinguishable from our own.

My third qualification is in regards to the results of conceptual analysis. As Chalmers and Jackson explicitly state in “Conceptual Analysis and Reductive Explanation”, they are not committed to the claim that conceptual analysis will yield a counterexample-free description of a property that is associated with the corresponding term. For present purposes, I will not take the lack of such a description to constitute a serious worry for their account. One reason to think that this isn’t a serious worry for the claim that conceptual analysis yields the property associated with a term is that any such description of the property may be incomplete because there are some cases yet to be considered.

However, if we *do* associate a property with a term in the invariant way that they claim we do, then we should expect that we will have an answer about what, if anything, has the property in a given scenario and so, what, if anything, is in the extension of the term at that scenario. For this reason, I will only require that, for each scenario that we consider, we have an intuition regarding what, if anything, is in the extension of that term at that scenario, on the hypothesis that the scenario is actual. That is, I will not take it to be a shortcoming of their view that conceptual analysis does not yield a counterexample-free description of the property associated with (or definition of) the term, though we will if we find that there are cases in which we find that we are unsure whether there is anything in the extension of the term at the scenario.

One can think of conceptual analysis as the process through which the property that is perhaps only tacitly associated with the expression can be made explicit. However, in order for conceptual analysis to be successful here, the individual must have a stable association between a property and the expression, such that the same property is used to pick out the extension at each scenario considered as actual. If there is no such stable association, then conceptual analysis will not yield any such property, and sentences in which that expression are a constituent are not knowable a priori as Jackson and Chalmers claim.

While I won't question here whether we have some such intuitions about the extension of our expressions across scenarios—that is, I won't challenge the claim here that there are

some scenarios,<sup>50</sup> described to us in some way, such that we have intuitions regarding whether a given expression, such as ‘water’, applies to something in that scenario—it isn’t obvious that we have these sorts of intuitions about *all* such scenarios. Because there are, plausibly, some scenarios about which we do not have a clear intuition about the extension of a given proper name or natural kind term, this poses a challenge both to determining the property by conceptual analysis, and casts doubt on either the claim that there *is* a specific property that we associate strongly with the expression, or the claim that if there is such a property, we don’t know precisely which property it is that we associate with the expression. Either way, if we do not have clear intuitions about all cases, this suggests that either there is no such property such that we associated it with the expression in the resilient way that Jackson and Chalmers think we do, or that, if we do associate some such property with the expression in this resilient way, we don’t know what the property is, such that we aren’t sure whether anything at that world has the property (because we aren’t sure what the property is), and so we don’t know what, if anything, is in the extension of the expression at that world. Let us call such cases in which the individual does not have clear intuitions about the extension of the expression at a particular scenario on the hypothesis that the scenario turns out to be actual *borderline cases*.

These borderline cases pose a serious worry for the claim that there is some property such that the association between it and the expression is stable across all scenarios. They also challenge the claim that there are some sentences, such as ‘water is the watery stuff’, that

---

<sup>50</sup> These needn’t come at an ontological expense—they can be, as Chalmers sometimes says, maximally coherent sentences, i.e. sentences such that, for any other sentence of that language, the former entails either the latter, or its negation.

we can know a priori, in the sense of a priori on which we can know it to be true independently of knowing which way the world turns out in the relevant respect.

Borderline cases challenge whether the property that we associate with ‘water’ is just the property of being the watery stuff, or if there are further constraints on the extension of ‘water’ provided by the property associated by water such that having the qualitative and functional properties that we associate with ‘water’ is not, in itself, a sufficient condition for being in the extension of ‘water’ at a scenario. Let us call the property of being the watery stuff the property W—such that, if we do associate a property with ‘water’ in this stable way, it is property W—and let us call the property of having the qualitative and functional properties that we associate with ‘water’ the property QF<sup>51</sup>. We can then ask the question whether property W is just property QF, or whether there are further constraints on the property of being in the extension of ‘water’ beyond merely having property QF, such that W cannot be just QF. Consideration of borderline cases demonstrates that we have good reason to think that W cannot be just QF,

One kind of borderline case arises when we consider scenarios at which there is more than one equally prevalent and very common kind that superficially resemble one another. One such scenario is a world at which there is both H<sub>2</sub>O and XYZ, and the two are equally prevalent.<sup>52</sup> What do we say about this case?

---

<sup>51</sup> ‘QF’ stands for qualitative and functional.

<sup>52</sup> Soames considers such a case in *Reference and Description: The Case Against Two-Dimensional Semantics*. See “Scenario 3” on p. 205.



If W contains a uniqueness condition, perhaps such that it is the predominant substance among those that have QF that is in the extension of ‘water’, then in engaging in conceptual analysis using that scenario, we should expect to find that we’d say that, if the world turns out as the scenario represents the world to be, then ‘water’ is an empty term. If we say this, then W cannot be just QF: it must contain the additional constraint that the substance must be the *predominant* substance that also has QF.

However, if we find that we say that both kinds would be in the extension of ‘water’ at that scenario, then we could, for all we’ve said so far, still argue that W just is QF. This is because this is a scenario at which there are two substances that have QF and that are equally prevalent, and both are in the extension of ‘water’. Chalmers and Jackson, it seems, would agree with this way of thinking about the scenario: they tell us<sup>53</sup> that on the hypothesis that our world turns out to be one at which there are equal parts XYZ and H<sub>2</sub>O, and they are the two most common clear odorless liquids by far, we will have the intuition that ‘water’ will turn out to refer to *both*.

I’m not confident that we have this intuition that, at such a scenario, ‘water’ refers to both, or I’m at least not confident that all of us do. But, for the sake of argument, let us assume that we do have this intuition. This has the immediate consequence that the property that we associate with ‘water’ is not such that it only applies to *one* kind. In other words, if this is right, we are wrong if we think that the property associated with water is being *the* watery stuff.

---

<sup>53</sup> “Conceptual Analysis and Reductive Explanation”, section 5.

Modifying our account of the first-order property associated with ‘water’ in this way by abandoning this uniqueness condition invites further worries. Consider a scenario at which there are *three* equally common clear odorless liquids. What about with four, or five, or six, or...a hundred? As we’ve already seen, procedurally, according to Jackson and Chalmers, we are supposed to consult our intuitions on the matter. What do our intuitions say about each of these cases? Would the world at which there are three equally common clear odorless liquids (and these are much more prevalent than other clear odorless liquids) be a world at which there are three kinds in the extension of ‘water’? What about the world with ten equally prevalent clear odorless liquids? What about the case with one hundred equally prevalent superficially indistinguishable clear odorless liquids?

Suppose that our epistemic state is such that we know that there is at least one kind that has QF, but we don’t know how many kinds have QF. Consider one such scenario that is consistent with the superficial appearance of the oceans. Suppose that what appears superficially to be one kind of stuff that has QF turns out to be a hodgepodge of one hundred equally prevalent superficially indistinguishable but chemically distinct clear odorless liquids. Suppose also that, upon consulting our intuitions, we conclude that it would not turn out that there was *a* kind—water— but rather one hundred different kinds, none of which deserves the name ‘water’ any more than the others. This is because, suppose, our intuitions tell us that ‘water’ is such that, whatever property W is associated with it, it will not be such that it can be had by so many as one hundred chemically distinct but superficially indistinguishable kinds. Since there is no *best* satisfier, and our

intuitions are against there being one hundred *equally good* satisfiers, then our intuitions tell us that, if our world turns out to be that way, ‘water’ is an empty term.

If this is what our intuitions tell us about such a case, then though this scenario is not a borderline case since we know that nothing is in the extension of ‘water’,<sup>54</sup> this scenario also gives us good reason to think that W is not QF. This is because having QF is insufficient for being in the extension of ‘water’: at that scenario, there are one hundred equally prevalent kinds that have QF, and yet none of them are in the extension of ‘water’ at that scenario.

There are other kinds of borderline cases that raise problems for the claim that we can know ‘water is the watery stuff’ to be true a priori, as well. One borderline case arises from the worry about whether there is a sharp cut-off, as could be discovered by investigating our intuitions, regarding how many chemically distinct kinds we will accept in the extension of ‘water’. Suppose that, in consulting our intuitions, we are willing to say, with Chalmers, that on the hypothesis that our world is such that there is an equal ratio of H<sub>2</sub>O to XYZ, ‘water’ picks out both H<sub>2</sub>O and XYZ. Suppose that we next consider a case with one hundred kinds of superficially indistinguishable but chemically distinct clear odorless liquids, and our intuitions tell us that this would not be a scenario in which ‘water’ picks out all, or even any (for the reasons I’ve just mentioned about deservers), of these types. This gives us good reason to believe that there is some cut-off point between two kinds and one hundred kinds. Suppose that ninety-nine is also a clear

---

<sup>54</sup> It would only be a borderline case if we were unsure about whether anything was in the extension of ‘water’ at that scenario.

case in which ‘water’ would not pick out the ninety-nine kinds. Maybe even *ten* is too many. But what about three, or five, or seven? Do we have clear intuitions about these cases?

If we do not have clear intuitions about the borderline cases, or, even worse, if we do not have clear intuitions regarding *which* scenarios are the borderline cases, then there are two possible concerns, both of which challenge the claim that there is a stable association between ‘water’ and W. First, it might be that there is not some property such that our association between it and the expression is stable in the strong way that Jackson and Chalmers defend. Perhaps, in those cases in which we are unsure what to say, we fail to associate W with ‘water’ in those cases, such that we aren’t sure what is in the extension at that scenario because this is a scenario at which the extension isn’t determined by the associated property. Secondly, perhaps there *is* some such property, but we don’t know what it is, or there is some other epistemic impediment to being able to use it to determine whether the expression has extension at a scenario, on the hypothesis that the scenario is the way the world is. If we don’t know which property we associate with ‘water’, then we cannot know what is in the extension of ‘water’ at that scenario, since the property mediates the reference.

There is a further related worry about the *ratio* of kinds. So far, we’ve been treating the various kinds of clear odorless liquids as equally prevalent. Suppose, hypothetically, that Chalmers is right about our intuitions about a scenario in which there are equal parts H<sub>2</sub>O and XYZ, such that if that scenario turns out to be actual, ‘water’ picks out H<sub>2</sub>O and

XYZ. Also, suppose that, upon consulting our intuitions about a scenario in which the ratio of watery stuffs is such that there is one part XYZ for every ninety-nine parts H<sub>2</sub>O, we discover that we've got a clear intuition that this is a case in which 'water' picks out H<sub>2</sub>O but not XYZ.<sup>55</sup>

This means that there is some cut-off point at which the ratio of one to the other is such that 'water' only picks out one. In order to determine this cut-off point, we must consider scenarios at which the ratio of H<sub>2</sub>O to XYZ vary, and to see at what ratio our intuitions tell us that 'water' no longer picks out both. What about a case at which there is a ratio of forty-nine to fifty-one? What about forty-eight to fifty-two? What about twenty-five to seventy-five? If we don't have clear intuitions about these cases then, again, either there is not some property such that our association between it and the expression is stable in the strong way that Jackson and Chalmers defend, or there *is* some such property, but we don't know what it is, or some other epistemic impediment to being able to use it to determine whether the expression has extension at a scenario, on the hypothesis that the scenario is the way the world is.

As noted earlier, these borderline cases cast doubt on the claim that individuals associate a property with a term in the stable fashion defended by Jackson and Chalmers. If such a case cannot be made then this, in turn, casts doubt on the claim that statements of the form '(the) N is F', where N is a proper name or natural kind term and F is the property stably associated with N, are a priori. Since the idea is that they are a priori for the individual that stably associates F with N, on the grounds that F mediates the extension of

---

<sup>55</sup> Chalmers and Jackson discuss this in "Conceptual Analysis and Reductive Explanation", section 5.3.

N for the purposes of counterfactual evaluation, such that anything in the extension of N has property F, then if there is reason to doubt that an individual *stably* associates F with N, then there is reason to doubt the claim that ‘(the) N is F’ is a priori for her.<sup>56</sup>

For this reason, one can see why it would be attractive to give a response that undercuts this challenge posed by the borderline cases. In particular, what would be needed is a response that explains why we don’t know what to say in borderline cases that is consistent with nonetheless stably associating a property with that term. We will now consider two such potential responses.

One potential response is to argue that perhaps the property associated with ‘water’ is vague. That is, perhaps ‘water’ is like ‘bald’: it clearly applies in some cases (e.g., someone with no hair), it clearly doesn’t apply in other cases (e.g., someone with a full head of hair), and it isn’t clear whether it applies in some range of cases (e.g., someone with three hairs, or four hairs, or five hairs, and so on). In order for this to be a response to the borderline cases, it must argue that the *property* associated with ‘water’ is vague. Thus, the indeterminacy in the borderline cases is metaphysical, rather than purely epistemological (though, of course, it has epistemological consequences). The idea is that, just as the property of being bald is such that it clearly applies in some cases, that is clearly does not apply in some cases, and there are some cases in which there isn’t a fact

---

<sup>56</sup> There is a potential response that involves arguing that the property F is not the property revealed solely through that individual’s conceptual analysis, but rather the property that is in common among the properties revealed through conceptual analysis across at least many speakers of the language. On the one hand, this resolves the borderline cases issue. On the other hand, this fares no better than the preceding approach in regards to defending the claim that ‘(the) N is F’ is a priori: arguably, polling others about the results of their conceptual analysis is not an *a priori* method for access to the property F.

of the matters whether it applies (i.e., the range of cases for which it is indeterminate), the same may be said of the property W associated with 'water'.

In reply to this response, I will only suggest that 'water' is not a special case in this way. I suspect that, in engaging in conceptual analysis about at least many natural kind terms, we would find borderline cases. My point is that this approach will commit one to taking any term for which there are borderline cases to be a vague term, which will, as I noted, plausibly include at least a great many natural kind terms. Vagueness in biological kinds may not bother us greatly, especially when we consider questions involving the transition between species over the course of evolutionary history, but vagueness in chemical kinds or the kinds in physics might. If what guides our intuitions about the extension of these terms is the property stably associated with them, then the uneasiness one may feel about taking terms such as 'water' or 'silver' to be vague may suggest that the property that we associate with these terms is not vague.

A second response is that if the property that is determined through conceptual analysis is the property that is in common across that which is in the extension of the expression at each scenario, considered counterfactually, then since the number of kinds and relative percentages of kinds will vary across cases, these won't be in common, and so won't this thereby fail to be a feature of the property? For example, if at a scenario at which there are two equally prevalent kinds, both are in the extension, and if at a scenario at which there are three equally prevalent kind, then all three are in the extension, and if at a scenario at which there are three kinds, two of which are significantly more prevalent

than the third, then only the first two are in the extension, and so on, then the number of kinds, and the relative percentages of kinds, vary across scenarios. That is, unlike property QF, which (we are assuming for the sake of argument) is invariant across all scenarios, the number of kinds, and the relative percentages of the kinds, in the extension of the term varies across scenarios. If it is what is in common across that which is in the extension of the expression across scenarios considered counterfactually, then the number of kinds, and the relative percentages between kinds, isn't part of the property revealed through conceptual analysis.

I have two considerations to raise in reply. First, while the number of kind, and the relative percentages of kinds, that are in the extension varies across scenarios considered counterfactually, the extension determined may well be consistent with an additional constraint on property W beyond merely having QF. This constraint might be quite complex: it might require, for example, that if there is more than one kind that has QF, then if one kind is 75% or greater relative to the other kinds, then only that kind is in the extension; but if two kinds are equally prevalent, then both are in the extension; and if there are three kinds, and two are significantly more prevalent than the third, then the first two are in the extension, and so on. Thus, this would be a property that would be in common among that which is in the extension at each scenario considered counterfactually.

The second consideration is that, in the absence of such a constraint, we are left in a poor position to explain our intuitions about how many kinds may be in the extension of the



expression at a given scenario, and what the relative percentages allow. That is, if the property is what determines our intuitions about the extension of the expression at each scenario considered counterfactually, then *unless* the property has such a constraint, it seems that these intuitions are left unexplained on this view.

At this point, it is time to make a more general remark about failing to have clear intuitions about “borderline” cases. If we do not have firm intuitions about these “borderline” cases, then this counts against our competency with the expression because we do not have an a priori grasp of the reference-mediating property for the expression. If the requirement for competence<sup>57</sup> with ‘water’ is knowing what property is associated with the expression, and thus knowing what property mediates the reference of ‘water’ because we thereby know that ‘water’ applies to that which has the property, then since we don’t know the property, we aren’t competent with ‘water’.

Furthermore, importantly, the reason that we don’t know whether a given proper name or natural kind term applies in a given borderline scenario cannot have to do with ignorance about further facts about the way the world turns out for two reasons. The first reason is because all of the scenarios under consideration are superficially indistinguishable from

---

<sup>57</sup> Jackson briefly mentions competence on p. 64 of *From Metaphysics to Ethics*, in which he says that our classification of things into categories capture patterns that underlie our conceptual competence. (Please note that what he means by ‘conceptual competence’ is roughly knowing what our *words* apply to. For example, see p. 33 of *From Metaphysics to Ethics*.) Chalmers and Jackson also discuss this in their “Conceptual Analysis and Reductive Explanation” in terms of concept possession: they claim that possessing the relevant concept allows one to be able to engage in what we are calling counterfactual evaluation (which, as we know, proceeds through using the property to determine whether the term applies at a given scenario as actual). From this, it stands to reason that if an individual lacks the ability to do this because that individual does not associate a property with the term in the stable way that they require, then that individual thereby does not possess the corresponding concept. (Remember that, in “Conceptual Analysis and Reductive Explanation”, they explicitly state that they move freely from discussion at the level of thought to the level of language, and vice versa.)

our world, and we are supposed to be able to know, *in that epistemic state*, whether ‘water’ applies to anything (and, if so, to what) in each of these worlds. Recall, for example, the epistemic state we are assuming ourselves to be in while considering all these scenarios about ways that water could turn out. We only know that there is at least one kind of clear odorless liquid in the oceans, and so this constrains the space of epistemically possible scenarios to those at which there is at least one kind of clear odorless liquid in the oceans (i.e., each scenario is at least superficially indistinguishable from the actual world with respect to the way the oceans appear). Still, if this epistemic state suffices to determine whether, at a scenario superficially indistinguishable from our own at which there is one kind that has QF, ‘water’ picks out that kind, and to determine whether, at a scenario superficially indistinguishable from our own at which there are 100 equally prevalent kinds that each have QF, ‘water’ is an empty term, then it is unclear why this wouldn’t work for determining what, if anything, is in the extension of ‘water’, evaluated counterfactually at a scenario at which there are a borderline number of equally prevalent kinds that all have QF.

The second reason is because, at least according Chalmers, these scenarios are maximal sentences that detail every fact about the world in qualitative terms,<sup>58</sup> such that, using this method of considering these scenarios as actual, there are no further facts to find out about the world, and so our uncertainty cannot be due to a posteriori ignorance of this sort. Rather, the shortcoming must be due to our uncertainty about whether the

---

<sup>58</sup> That these scenarios are described qualitatively is for the purpose of avoiding trivializing sentences, such as ‘there is water in the oceans’, in the description. See his discussion of this in “Does Conceivability Entail Possibility”, section 8, under the heading, ‘Scrutability’.

expression applies to anything at that scenario, on the hypothesis that that scenario is actual, *even when we know all the a posteriori facts about that world.*

There is a different, further problem with the first-order property—it cannot do all the work that Jackson and Chalmers want it to do regarding picking out the same individual or kind uniformly. Consider again the property QF. Since something has property QF just in case it has the qualitative and functional characteristics that we associate with ‘water’, QF will be determined by that individual’s expectations for how water is presented under various conditions. However, in acknowledging that water may present in such a way as to fail to have QF under novel conditions, the individual demonstrates that he or she does not intend for ‘water’ to refer to *only* that which has QF, but rather all and only samples of the same kind. Thus, we clearly do not intend for ‘water’ to pick out *only* things with the property of being a clear odorless liquid. Rather, we intend to use ‘water’ to talk about that which has the property of being the watery stuff *and everything else of the same kind (and nothing else)*. If ‘water’ picks out that which has the first-order property (e.g., the property of being the watery stuff), and at least some H<sub>2</sub>O fails to have that property at the time a given token of ‘water’ is used, then how is it that that token of ‘water’ is able to refer to *all and only* H<sub>2</sub>O, as Jackson and Chalmers want it to do, rather than just the H<sub>2</sub>O that has the first-order property at the time the token is used?

Furthermore, we intend for our use of ‘water’ to pick out the same kind across worlds as it picks out in the actual world. That is, when I say, ‘water might have been in this glass’, while gesturing at a glass, I intend to say something that is true just in case the same

substance that is in the extension of ‘water’ at the actual world is of the same kind as the substance in the glass at the relevant counterfactual world. This is true for the uniform extension intuition and for the contingent extension intuition: as we’ve discussed, both share commitment to proper names and natural kind terms uniformly referring to the same individual or kind, respectively, across all worlds at which the expression picks out anything at all. It is for this purpose—making it the case that our proper names and natural kind terms refer to the same individual or kind, respectively, not only when they have the property associated with the proper name or natural kind term, but also when they do not, and also to make it the case that such expressions uniformly pick out the same individual or kind, respectively, across all worlds at which they pick out anything at all—that we need the second-order properties associated with the primary and secondary intensions of proper names and natural kind terms.

#### ***Section 4: Second-order property***

I claim that, rather than the first-order property, as Jackson and Chalmers seem to believe determines the extension of primary and secondary intensions, it is the second-order property that is associated with a proper name or natural kind term that determines the extension of the primary and secondary intensions of expressions. That is, whatever has the second-order property associated with the primary intension or secondary intension of that expression is in the extension of the primary intension or secondary intension, respectively. The reason that it cannot be the first-order property associated with a given proper name or natural kind term that determines the extension of the primary intension

and secondary intension of that expression is that the first-order property will only pick out the relevant individual if that individual has the property at the time the token of the name is used, or samples of the relevant kind(s) that have that property at the time the token of the natural kind term is used. For example, as we noted at the end of the preceding section, the first-order property of being the watery stuff can only be had by that which has the property of being *a* clear odorless liquid. Thus, if we want to somehow pick out the same individual (even when that individual fails to have the property associated with the name) or all samples of the same kind as those samples that have the first-order property, we need a second-order property that is had by that individual or all samples of the same kind, respectively.

Since this problem is particularly challenging in the case of natural kind terms, we will focus on developing an account of the second order property associated with natural kind terms here. As a first pass at such a property, the second-order property is the property of being of the same kind as those samples that have the first-order property associated with the expression. In order to see the difference between the second-order property associated with the primary intension of the expression and the second-order property associated with the secondary intension, we need to distinguish between the first-order property associated with the primary intension of the expression, and the first-order property associated with the secondary intension. The first-order property associated with the secondary intension of an expression is actually-involving—that is, it is the property of being (the) actual *F*, where *F* is the property associated with the expression, but the first-order property associated with the primary intension is not (i.e., *the actual F*

versus *the F*, respectively). On this way of thinking about it, the second-order property associated with the primary intension of 'water' is the property of being of the same kind as those samples that have the property of being the watery stuff, such that the kind(s) in the extension of the primary intension vary across worlds. In contrast, the second-order property associated with the secondary intension of 'water' is the property of being of the same kind as those samples that have the property of being the actual watery stuff, such that the same kind is in the extension of the secondary intension across worlds as the kind that is in the extension of it at the actual world.

The two second-order properties do different work. The second-order property associated with the primary intension of an expression is that which determines the extension of the primary intension of that expression at each world still epistemically possible for that individual or community, which is used to capture the cognitive significance of the expression, on Chalmers's account, or the conventions that govern the use of that expression, on Jackson's account. The second-order property associated with the secondary intension of an expression is that which determines the extension of the secondary intension, which captures the modal profile of the expression.

That the second-order property associated with the secondary intension has to do with what has the property at the actual world, such that it picks out the same kind across all counterfactual worlds as it picks out in the actual world, guarantees that the secondary intension captures our uniform extension intuition about the modal profile of the expression. That is to say, the second-order property associated with the secondary

intension of an expression ensures that the modal profile of the expression picks out all and only samples of the same kind across worlds.

As we discussed earlier, the work done by the primary intension is to capture the cognitive significance of the expression for an individual. One way to think about this is that the cognitive significance of the expression is captured by the extension of the primary intension of the expression at each world that is consistent with what that individual believes about the way our world is, evaluated counterfactually at that world under consideration. Since each of these worlds is a way the world could turn out, for all that individual believes, the extension of the primary intension at each of these worlds is a way the extension of that expression could turn out to be, for all that individual believes.

There is reason to restrict counterfactual evaluation to the doxastically possible worlds, rather than to allow counterfactual evaluation across all worlds. One significant reason to do this is that it makes much more sense of how sentences such as ‘water is the watery stuff’ are thought to come out a priori, in the sense of a priority on which we can know the sentence to be true independently of knowing how the world turns out in the relevant sense. To see this, consider an individual who believes that the oceans are filled with the watery stuff, though she is unsure what the liquid is. This means that all the worlds that are doxastically possible for her are such that the oceans are filled with the watery stuff at that world. This means that among the worlds that are doxastically possible for her are worlds at which the watery stuff is H<sub>2</sub>O, and so the oceans are filled with H<sub>2</sub>O, and

worlds at which the watery stuff is XYZ, and so the oceans are filled with XYZ, and so on. If we evaluate counterfactually ‘water is the watery stuff’ at all the worlds simpliciter, there will be some worlds at which there isn’t a watery stuff, and so ‘water’ is empty, and ‘water is the watery stuff’ is false. However, if we evaluate counterfactually ‘water is the watery stuff’ at all and only the worlds that are doxastically possible for her, then each of those worlds has oceans that are filled with the watery stuff and so, evaluated counterfactually, the sentence is true at each doxastically possible world. This suggests that the sentence is true no matter which of the doxastically possible worlds turns out to be, in an epistemic sense of ‘turns out’, the way the actual world is. We can see from this the conception of a priority they have in mind: the sentence is true on counterfactual evaluation, no matter which of the doxastically possible worlds is the way the actual world is.

However, there is a problem that arises, potentially quite often, in restricting counterfactual evaluation to only the doxastically possible worlds for this purpose. If all worlds under consideration are consistent with what the individual *believes* about the world, but the individual’s belief about a relevant matter is false, then there are two problematic considerations. First, the actual world will not be among the worlds used for counterfactual evaluation, and so even though there is something in the actual world that has the first-order property associated with the expression, nonetheless the primary intension won’t have extension at the actual world if we require that counterfactual evaluation is only performed at the doxastically possible worlds. Since the primary



intension and the secondary intension are supposed to be co-extensive at the actual world, it is problematic if the primary intension fails to have extension at the actual world.

Second, the sentence is true on counterfactual evaluation at each of the possible worlds, such that it comes out as a priori in the relevant sense, and yet it is false because it is false at the actual world. Since the second-order property that determines the extension of the expression is the property of being of the same kind as that which has the first-order property at the actual world, and nothing has that first-order property at the actual world, then the expression is empty at the actual world, and the sentence in which it is a constituent is false.

To see this, consider an individual that believes that water is the watery stuff, and so associates the property of being the watery stuff with 'water'. Is the sentence 'water is the watery stuff' a priori for her? Based on what we've said so far, it might sound like it counts as a priori in the sense that interests Jackson and Chalmers. All the worlds that are doxastically possible for her are worlds at which there are oceans, and they are full of the watery stuff at that world.<sup>59</sup> However, there are a number of ways that the actual world could turn out that are inconsistent with the way she believes the world is, such that the sentence could be true on counterfactual evaluation at each world that is doxastically

---

<sup>59</sup> Note that it is because she *believes* that the oceans are filled with the watery stuff that the doxastically possible worlds for her are all such that the oceans are filled with the watery stuff. If her beliefs about the liquid(s) in the ocean are entirely based on the superficial appearance of the ocean, then she isn't justified in her belief (who knows why she holds the belief, or from where she might have gotten it). Note also that if we said that, rather than the doxastically possible worlds, we will do counterfactual evaluation with epistemically possible worlds, we could not narrow down the space of worlds to just those at which the oceans are filled with *the watery stuff*, since, as we've noted, she isn't justified in believing that the oceans are filled with a single watery stuff.

possible for her, and so it counts as a priori on this understanding of a priority, and yet the sentence is false because the world isn't that way.

The first kind of case is if the actual world is such that there is more than one equally prevalent clear odorless liquid in the oceans. Since she believes that water is the watery stuff, and yet there isn't *a watery stuff*, then nothing in the actual world has the first-order property associated with 'water'. Since nothing in the actual world has the first-order property associated with 'water', the extension of the secondary intension of is empty at the actual world, meaning that 'water' is an empty term, and so the sentence is false. Nonetheless, since all of the worlds doxastically possible for her are such that they all have oceans, and those oceans are filled with the watery stuff, her sentence is true on counterfactual evaluation at each doxastically possible world, and so a priori for her.

The second kind of case is one in which the liquid in the oceans is homogeneous, but there is, unbeknownst to this individual, a yet more prevalent clear odorless liquid than that which fills the oceans. Suppose that, as on our earlier example, we are radically mistaken about our beliefs about the interior of our world, such that, rather than having layers of various kinds of rocks in various states, the entirety of the center of our planet is full of XYZ. As before, the individual believes that water is the watery stuff, and so associates the property of being the watery stuff with 'water'. Furthermore, she believes that there are oceans, and she believes that the oceans are filled with the watery stuff, so all worlds that are doxastically possible for her are ones at which the oceans are filled with the watery stuff at that world. This, of course, excludes the actual world from the

set of doxastically possible worlds since its oceans are not filled with the watery stuff. This means that, on counterfactual evaluation at each doxastically possible world, the sentence is true, and so a priori, but, as before, the sentence is false.

What this shows us is that counterfactual evaluation cannot be restricted to just the doxastically possible worlds for the relevant individual. However, in not restricting counterfactual evaluation in this way, we are unable to use it to motivate the notion of a priority that we've been discussing.

#### ***Section 5: Picking Out All and Only Samples of the Same Kind***

Another serious worry for Chalmers's and Jackson's epistemic interpretations is that, since both intensions are part of the semantics for an expression, this raises the possibility that any such expressions will be ambiguous if the two intensions pick out different extensions at the actual world. The second-order property associated with the primary intension is the property had by all samples of the same kind as that which has the first-order property at a given world considered counterfactually, whereas the second-order property associated with the secondary intension must involve whatever has the property at the actual world, such that the second-order property picks out uniformly, at the world considered as counterfactual, all and only samples of the same kind as that kind that has the first-order property at the actual world. In order to avoid ambiguity, it must be the case that either both intensions pick out all and only samples of the same kind that has the first-order property, or that they both pick out all and only the samples that have the first-

order property (but *not* also the samples of the kind that fail to have the first-order property). That is to say, the second-order property associated with the primary intension and the second-order property associated with the secondary intension must be such that either both of these second-order properties pick out *all and only* samples of the same kind, or else both second-order properties must pick out *only* those samples of the kind that have the first-order property.

For example, it would avoid ambiguity if both the second-order property associated with the primary intension of 'water' picks out all and only samples of the same kind as that which has the property of being the watery stuff (or whatever turns out to be the first-order property associated with the primary intension of 'water') at each world considered counterfactually, and the second-order property associated with the secondary intension must be such that it picks out uniformly at each world considered counterfactually all and only samples of the same kind as that which has the property of being the watery stuff at the actual world. Similarly, it would avoid ambiguity if the second-order property associated with the primary intension of 'water' picks out all and only those samples that have the property of being the watery stuff at the world considered as counterfactual, and the second-order property associated with the secondary intension of 'water' picks out *all and only* samples of the kind that has the property of being the actual watery stuff.

What would allow for the possibility of ambiguity is if one of the second-order properties picked out *only* those samples that had the first-order property (i.e., whatever is the clear odorless liquid at a world, but not samples of the same kind in another physical state,

such as, at an H<sub>2</sub>O world, picking out liquid H<sub>2</sub>O but not also ice), and the second-order property associated with the other intension picked out all and only samples of the same kind (i.e., all and only samples of the same kind that has the first-order property, such as *all* and only H<sub>2</sub>O, at an H<sub>2</sub>O world). In other words, if the second-order property associated with the primary intension of 'water' picks out only liquid H<sub>2</sub>O, and the second-order property associated with the secondary intension of 'water' picks out H<sub>2</sub>O in all of its states, then 'water' will thereby be ambiguous.

More generally, failure to have this similarity between the second-order properties results in allowing for ambiguity. Furthermore, since we need the second-order property to pick out all and only samples of the same kind as that which has the first-order property, in order to satisfy the contingent extension intuition, this means that both intensions must pick out all and only samples of the kind that has the first order property.

The second-order property is to do two things. First, it is to capture the cognitive significance of expressions. This is done by using the first-order property in determining the second-order property, such that among the things that have the second-order property, and so are in the extension of the relevant intension, are things that have the first-order property that is associated with the expression. Second, it is to capture the shared feature of the uniform extension intuition and contingent extension intuition: that proper names and natural kind terms pick out the same extension uniformly. This is done by developing the second-order property to be such that it picks out the same individual or kind as that which has the first-order property.

The motivation for these two features of the second-order property is that (a) in using a proper name or natural kind term, we intend to refer to whoever or whatever has the first-order property that is associated with the expression, and (b) we also understand that, often, individuals and kinds have properties beyond those we recognize, but we intend to refer to the same individual even when that individual does not have that first-order property, and to refer to the same kind including samples that aren't presented as having the first-order property. Here, we will continue to focus on natural kind terms, as I think the problems that arise in formulating the second-order properties are clearer in the case of natural kind terms.

Now we will investigate what the second-order property must be like in order to pick out all and only samples of the same kind as those samples that have the first-order property associated with that natural kind term. For example, suppose that we take the first-order property associated with 'water' to be *being the watery stuff*. This first-order property is had by all samples of H<sub>2</sub>O that are in the liquid state (and at least appear superficially not to be contaminated). The second-order property must then be such that it picks out not only those samples of H<sub>2</sub>O, but all samples of the same kind. Let us call this feature of the second-order property that is such that it picks out all and only samples of the same kind *the unifying feature* of the second-order property.

As we will see, the unifying feature will prove problematic for formulating the second-order property. First, there is a worry because there are a few things we could mean by

‘kind’ here, which will make a difference for what we mean by *all and only samples of the same kind*. Depending on the context, it might be a natural kind (e.g., in the context of a science discussion) or a functional kind (e.g., in the context of an economic discussion). We’ll look first at natural kind terms, and then functional kind terms.

In giving an account of natural kind terms, not only are we interested in what it means for things to be members of the same kind, but also what makes it the case that what we take to be natural kind terms is such that *they* constitute a semantic kind. That is to say, the latter concern is that we have a unified semantic kind here. Whether or not we have a unified semantic kind—i.e., the semantic kind, *natural kind terms*—depends on whether or not we can give a unified account of the second-order properties across natural kind expressions. By this, I mean whether or not we can give an account of the nature of the second-order properties that is sufficiently general to capture all the second-order properties associated with the primary and secondary intensions of natural kind terms, but not so general that it also captures the second-order properties associated with the primary and secondary intensions of other kinds of expressions.

Finding a unified account of second-order properties for natural kind terms proves challenging. It looks like, for the purposes of science, what matter for sameness of kind is sameness of natural kind, as determined by some sort of underlying sameness (e.g., in chemistry) or similarity in important respects (e.g., discussion of species) of constitution. One way to do this is to make the second-order property the property of natural kind

terms the property of having of the same underlying constitution (in some sense that will be challenging to spell out) as those samples that have the first-order property.

What do we mean by *sameness of underlying constitution*? It might be that this means different things for different contexts or for different domains of discourse. For example, the underlying constitution relevant for 'water' or 'silver' is likely the chemical constitution, whereas for 'dog', it is likely the genetic similarity.

One might think that sameness of underlying constitution requires sameness in *all* underlying constitution (e.g., sameness in chemistry guarantees sameness in underlying physics, etc.) This won't work in the case of biological kinds, especially animals, e.g., dogs, as they differ in underlying constitution in some senses, e.g., size of organs, location of organs, etc., even within the same breed. To adopt a general view of kinds as determined by sameness of underlying constitution in the way that might be appropriate to chemistry or physics is to make biological kinds very fine grained indeed: the consequence for, say, kinds of dogs, might well be that each individual dog is of a unique kind, since each dog differs in underlying constitution (e.g., size and/or relative location of organs) in at least some respect from other dogs, even those of the same breed.

Thus, a different account of what is meant by sameness of kind will be needed for biological kinds than for chemistry and physics. Though it's not an exhaustive list by any means, some alternative accounts for biological kinds include, for two organisms A and B, (a) A and B have the same underlying genetics, and (b) A and B have sufficiently



similar underlying genetics, (c) organism A and organism B are of the same kind if A and B can have fertile offspring.

The first approach has the attractive feature that what it takes to be a member of the same kind has *something* to do with underlying sameness of some feature of the composition of the members of the kind, namely, the sameness of the underlying genetics. In this way, it would allow us to salvage something about the general claims that sameness of kind is a matter of underlying sameness of constitution. However, strictly speaking, this cannot be right, as there are very many cases in which there is some, however slight, genetic variation across what we, intuitively, take to be members of the same kind. For example, intuitively, I am of the same kind as you, and yet, strictly speaking, we do not have the *same* underlying genetics.

What one who suggests that account probably has in mind is that there is a strong genetic *similarity*—that there is sameness in *some* of the underlying genetics between two members of the same kind. However, there is some difficulty in specifying precisely which underlying genetics must be the same. There is also a worry if we try to specify sameness of underlying genetics as some percentage of similarity without specifying which aspects of the genetics must be the same because it might not be the *mere degree of similarity* that matters, but also the respect in which the compared genetics are similar.

The third approach—that sameness of biological kind is determined by the ability to produce fertile offspring—is problematic for at least two reasons. First, it does not allow

us to define species for those organisms that reproduce asexually. Secondly, it has been discovered that some “ring species” are able to cross-breed with other birds that are genetically less similar to them than other birds more genetically similar to them. That the “ring species worry” is raised by some who are engaged in this discussion gives us reason to think that our intuitions about natural kinds do not obviously favor the “producing fertile offspring” version of species, as the fact that this is described as a case in which this can be done across less genetically similar species but not more genetically similar species—this suggests that at least some engaged in this debate are at least somewhat inclined to take genetic similarity to be more important for the purposes of categorization of species. Furthermore, that there is so much debate about the matter of how to define sameness of species suggests that either we do not have clear intuitions about the matter, or, if we do, they are not universal.

Unfortunately, it isn't just biological kinds that don't allow for the sameness of underlying constitution constraint on kind membership. Geological kinds are also determined by something other than sameness of underlying constitution. Kinds of rocks are typed by their origins: whether they solidified from a molten state, whether they were formed by erosion and deposition, or whether they were altered by heat, pressure, or reactions caused by the actions of gasses and/or fluids. The very same rock could be first formed by deposition, and then altered by pressure, in which case it would change kinds, even though nothing has changed about its underlying constitution—it's still the same components, though altered by the pressure.

Let us set these worries regarding sameness of kind aside for the present so that we can discuss other features of the view. I think that among the things that this discussion shows is that a unified account of the second-order property cannot be given for all natural kind terms. Instead, we'll need some other way of motivating the idea that they are all expressions of the same semantic kind. One way to do this is to stipulate that natural kind terms are those expressions that refer to all and only those expressions of the same natural kind, while allowing for what it is to be of the same kind to be determined by the second-order properties of the natural kind terms of a given domain of discourse.

Not only do we face problems in developing the second-order property due to challenges for the unifying feature of the second-order property, we also must consider problems that carry over from the problems with the first-order property. First, if there is no uniqueness feature (e.g., *the watery stuff* vs. *either the watery stuff or the equally prevalent clear odorless liquids* vs. *any clear odorless liquid(s) of such-and-such-ratio of the amount of clear odorless liquids at the world* vs. *any of the clear odorless liquids of such-and-such ratio provided that the total number of kinds of such liquids is less than [number]*) then we need a second-order property that allows us to pick out all and only those samples of each kind that has the first-order property.

For example, suppose that the first-order property associated with 'water' is *the watery stuff or the equally prevalent clear odorless liquids*. We need a second-order property that picks out all and only samples of each of the kinds that have the first order-property at each world considered counterfactually. This may not, in itself seem any more

problematic than being able to pick out all and only samples of one kind uniformly. However, consider the hypothesis that the actual world is such that it has four equally prevalent superficially indistinguishable but chemically distinct clear odorless liquids. Suppose, also, that the first-order property associated with ‘water’ is *being among the most prevalent clear odorless liquids* and so the second order property picks out all and only samples of all four kinds. Call the four kinds A, B, C, and D. Now, consider a counterfactual world (relative to the world that we are hypothesizing to be actual) that is such that it has only kinds A, B, and C. What is the extension of the secondary intension of ‘water’ there?

In order to determine this, we need to think about the relationship between these kinds: is this a kind of disjunctivism, such that, so long as at least one of these kinds is at a world, ‘water’ has an extension at that world? Is it a “conjunctivism”, such that if not all four kinds are at a world, ‘water’ is empty? Is it something in-between? I won’t try to settle the matter here. Instead, I’ll just confess that I don’t have clear intuitions on the matter, and I suspect that at least some others will be in the same boat—this casts doubt on whether we have the contingent extension intuition in these cases, since we don’t have clear intuitions about how ‘water’ would turn out, had that counterfactual world instead been actual.

A related worry is that if we either don’t associate a first-order property with a given natural kind term, or we are unsure which property is associated with that natural kind term, such that we don’t know whether the expression applies to anything at the scenario

considered as actual, as in the borderline cases, then there either isn't a first-order property to use in constructing the second-order property, or we don't know which property is the first-order property to be used in the construction of the second-order property. The first poses a challenge to there being a second-order property associated with the primary intension and secondary intension of that natural kind term because the first-order property is required for the second-order property. If there isn't a first-order property, then there can't be a second order-property. The second poses a challenge for us knowing which property the second-order property is. That is, this would be a case in which, even though there *is* a first-order property, we don't know what it is. If we don't know what it is, we can't know what the second order property is, either.

There is also a worry regarding expressions that behave like natural kind terms in some contexts and functional kind terms in others. 'Silver', in a science classroom, behaves more like a natural kind term (e.g., it is the underlying constitution that determines sameness of kind). 'Silver', in an economics classroom, behaves more like a functional kind term (e.g., what matters for sameness of kind is what it is worth as a currency, etc.)

Perhaps whether a given token of 'silver' is a natural kind term or a functional kind term is a matter of the context. That is, perhaps tokens of 'silver', as used in a chemistry class, are a natural kind term, and tokens of the very same kind, as used in an economics class, are a functional kind term.

However, we can see that, in some contexts, within that same context, one token of ‘silver’ could behave like a natural kind term, and another could behave like a functional kind term. For example, consider an economics class in which they are discussing the value of silver currency in some culture. A student arrives to class late and asks what the coins are made of. Another student tells her that it is silver. Question: are the tokens of ‘silver’ in the preceding discussion about the value of the currency tokens of the same type as the student’s answer that the coins are made of silver?

This raises questions about individuating types of expressions. Are there two different semantic types, the natural kind term ‘silver’ and the functional kind term ‘silver’? Is it the same type but different tokens of the same type behave differently across contexts? If the former, how is it that the student doesn’t change the subject, but rather manages to talk about the same thing, in both the discussion about the value of silver and also when she tells the late-comer that the coins are made of silver? If the latter, how do we explain how it can be used as both a natural kind term and a functional kind term within the same context, as we saw in the preceding economics class discussion?

There is a further worry regarding the possibility that the same expression may be used as a natural kind term in some contexts and a functional kind in other contexts. A token of ‘silver’ may be a natural kind term in one context due to its first-order property *or* its second-order property, and a token of ‘silver’ may be a functional kind term in another context due to its first-order property *or* its second-order property. That the two can have

different first-order properties allows for the very same expression to have different issues of vagueness that arise in different contexts.

Let's start with looking at how worries for the first-order properties can introduce different kinds of vagueness. Consider, first, the first-order property associated with 'silver' when we think of it as a natural kind term. Suppose that, for the sake of simplicity, we allow the second-order property associated with a token of 'silver' used in a chemistry class to be the property of being of the same kind as those samples that have the first-order property associated with the expression (where we ignore, for the present, the preceding worries about sameness of kind). Suppose that the first-order property is *being either the most prevalent shiny silver metal or among the most prevalent shiny silver metals*. This means that the second-order property is the property of being of the same kind(s) as that which has the first order property.

That 'silver', as a natural kind term, may be vague is similar to our earlier worry about how many superficially indistinguishable kinds of clear odorless liquid may be in the extension of 'water' at a scenario considered counterfactually. Consider a world at which there are ten chemically distinct kinds of equally prevalent superficially indistinguishable shiny silver metals. Evaluated counterfactually at this world, are these ten chemically distinct kinds of metals in the extension of the primary intension of 'silver'? What about a world at which there are one hundred chemically distinct kinds of shiny silver metal? What about a world at which there are two chemically distinct kinds of shiny silver metal, but the ratio is four to six between them? What about a world at which the ratio is three

to seven, or two to eight? If there are at least some cases in which we are unsure whether, on counterfactual evaluation, the primary intension of 'silver' has extension at that world, or we aren't sure what the extension of the primary intension is at that world, this introduces one kind of vagueness for 'silver'.

Now, consider the ways that the first-order property associated with 'silver' when we think of it as a functional kind term can introduce vagueness in a different way. A token of 'silver' may be a functional kind term in another context by virtue of its first-order or second-order properties. Suppose that the first order property associated with 'silver' in an economics class is the property of being worth \$30 an ounce. Also, suppose that silver is the only metal that has that value at the actual world. Consider a world in which they trade, not in coins, but in rocks, such that, say, one ounce of slate is worth \$30 an ounce. At that world, slate plays the functional role that silver plays in our world, and so if silver is a functional kind in the economics context, then at that world, considered as actual, slate is in the extension of the primary intension for 'silver', evaluated counterfactually at that world at which slate plays the same functional role that silver plays at ours.

Next, consider a world at which they don't really have a concept of currency, but rather only barter. Suppose that, at that world, bartering one ounce of Gouda cheese will buy you exactly the same sorts of goods that one ounce of silver would allow you to buy at our own world. Does Gouda play the same functional role at that world that silver plays at ours, such that, evaluated counterfactually at that world, Gouda is in the extension of the primary intension of 'silver'?



If you have a clear intuition (either way) in that case, consider a world very much like the one just described except that thirty different goods have precisely the same buying power at that world that silver has at our world. Would all of those thirty goods be in the extension of the primary intension of 'silver', as evaluated counterfactually at that world? If you said yes to the preceding Gouda case, and no to this case, then consider a case in which there are  $n$  goods such that they have the same buying power at that world as silver has at this world, where  $n$  is the number of goods such that you don't have a clear intuition about whether you would group them together as the same functional kind because they are otherwise quite disparate.

This is a second way of introducing vagueness to the primary intension of 'silver'. But notice that now we have two different ways in which the primary intension of 'silver' can be vague. Furthermore, given that Jackson and Chalmers allows that the primary intension of a proper name or natural kind term may vary across contexts, this allows for the ways in which such expressions can be vague to only be limited by the ways in which their primary intensions can vary across contexts.

One way to resolve this is to use a first-order property that captures both the qualities of a kind's superficial appearance *and* functional properties. For example, we could use the as the first-order property associated with 'water' the property of being the watery stuff that falls from the sky as rain; comes out of taps; composes rivers, lakes, and oceans; quenches thirst; can be used to bathe and water plants; etc. Similarly, for 'silver', we

could use the property of being the shiny metal that is worth \$30 an ounce. Rather than having many primary intensions, this would allow us to use the same primary intension across contexts, avoiding the problem of many ways the primary intension of the expression could be vague.

The problem with this is that it allows for a modified problem of unwanted necessity to arise. The original problem of unwanted necessity is a challenge to descriptivism: in using ‘the last great philosopher of Antiquity’ as the description that mediates the reference of ‘Aristotle’, we make ‘Aristotle was the last great philosopher of Antiquity’ a necessary truth. However, it seems reasonable to think that this sentence is contingent: it seems possible that Aristotle might well have gone into another profession, or to have not been such a great philosopher.

To see the modified version of the problem of necessity, think about a case in which we want to engage in counterfactual evaluation for ‘silver’ as a natural kind term, such that we can consider how the extension of ‘silver’ would have turned out differently had another world been actual. It seems that, in this case, what is relevant is the superficial appearance, and *not* what we do with it (e.g., treat it as currency). If we have a first-order property that includes both superficial and functional features, then at a world at which there is a predominant silvery shiny metal, but no one makes currency of it, then the natural kind term ‘silver’, evaluated counterfactually at that world, is empty. This seems an odd result. For example, our world could have been that way—it might have been that, though in fact there is silver at our world, and we successfully introduced ‘silver’ to

refer to it, no one ever bothered to make currency out of silver—and yet that would be a case in which the natural kind term, ‘silver’, would not be empty.

On the other hand, if we want to engage in counterfactual evaluation for ‘silver’ as a functional kind term, then what matters is what we do with silver, and not its superficial appearance. At a world at which there was something that played the same role in our currency system, but failed to have the same superficial appearance as silver, at that world considered as actual, the functional kind term, ‘silver’ is empty. But if what is important for that which is in the extension of functional kind terms is that they play the relevant role, rather than their superficial appearance, then this result seems odd, too.

### ***Section 6: Two Ways of Developing the Second-Order Property***

It seems that there are two ways of determining a second-order property. In the first case, we have some first-order property that we associate with a natural kind term that doesn’t include how many chemically distinct kinds can be in the extension of that natural kind term, and we say that across worlds, however many kinds at that world have this first-order property also have the second-order property associated with the primary intension, and so to be in the extension of the primary intension at that world.

For example, suppose that the first-order property that we associate with ‘water’ is the property of being a clear odorless liquid that falls as rain; constitutes rivers, lakes, and oceans; quenches thirst; is used to bathe; etc. However, the property isn’t such that it

restricts the number of chemically distinct kinds that can be in the extension of the primary intension of ‘water’, evaluated counterfactually at a world, provided that all the superficially indistinguishable but chemically distinct kind have that first-order property. The second-order property associated with the primary intension of the natural kind term is the property of being a member of any of those kinds that have the first-order property associated with the primary intension of that natural kind term. For example, at a world at which there are one hundred chemically distinct kinds that have the first-order property, the second-order property uniformly picks out all samples of each of those one hundred kinds. In this way of thinking about the second-order property, we let the world determine how many chemically distinct kinds are in the extension of the primary intension, namely, however many kinds at that world have the first-order property.

The second way of thinking about the second-order property is that we have some property in mind, and that property includes some cut-off point regarding how many chemically distinct kinds there are. Thinking about second-order properties in this way would be motivated if we find that we have clear intuitions that, say, at a world at which there are three chemically distinct kinds that have the first-order property associated with ‘water’, all three kinds are in the extension of the primary intension of ‘water’, but at a world at which there are four such kinds, we have a strong intuition that not all four kinds will be in the extension of the primary intension (and that either no more than three of them, or none of them, will be in the extension).

I take it that the way that Jackson and Chalmers would advise that we adjudicate between which of these two ways we should think about second-order properties is that we should consult our intuitions about the matter. However, there is reason to worry that our intuitions underdetermine which way of thinking about second-order properties is most in accordance with our intuitions. Borderline cases provide examples in which we don't know whether there is anything in the extension of a natural kind term because we are unclear about how many kinds may be in the extension of that natural kind term, evaluated counterfactually. A given borderline case might be a borderline case because we have the intuition that there is a cut-off point, but we aren't sure what it is, or it might be a borderline case because we aren't sure whether there is a limit to how many kinds may be in the extension of the natural kind term, evaluated counterfactually. We may well find that we are uncertain which of these is the reason that a given case is a borderline case. This gives us reason to think that which way of thinking about second-order properties is most in accordance with our intuitions is underdetermined by our intuitions in such cases, since our intuitions do not settle the matter for us.

### ***Section 7: Bad Predictions***

In this section, I argue that since both Chalmers and Jackson are committed to the uniformity feature of the contingent extension intuition, such that the primary intension of proper names picks out the same individual even when that individual does not have the first-order property associated with the name, and the primary intension of a natural kind term picks out the same kind uniformly, including those samples that don't have the

first-order property associated with that natural kind term, this allows for the same problem to arise for them regarding explaining the mistake made by Bruce's students that we saw arise for Stalnaker in the preceding chapter. What this means is that Jackson and Chalmers will run into the same problems as Stalnaker in modeling the force of assertions in cases in which that which is in the extension of the expression is presented in ways that violate the interlocutors' expectations.

Recall Jackson's and Chalmers's procedure for ruling out worlds in light of assertions: we evaluate the assertion counterfactually at each world that is epistemically possible for the audience, and we rule out those worlds at which what is said at that world, evaluated counterfactually, is false at that world (i.e., those worlds that are false on the diagonal). Remember that our challenge is to explain how the students ruled out the actual world—a world at which there is ice in the basin—in light of Bruce's assertion, 'there is water in the basin'.

Evaluated counterfactually at our world, what is said at our world is true at our world. In other words, the primary intension of his assertion maps this world to *true*. This is because that which is in the basin is of the same natural kind as that which has the property of being the watery stuff. Thus, this doesn't explain why his students ruled it out.

One might think that it could be explained by saying that the students took his assertion to rule out this world because they are mistaken about the way this world is: they think

this world is one in which the watery stuff is of a different kind than that which is the predominant cold, clear, solid. Perhaps the students thought that our world is *that* way, and so thought that what he said was false here, and so that is why they ruled out this world.

While it's true that this is a good explanation of why the students would have ruled out *that* sort of world, it is not an appropriate explanation for why they took it to rule out *this* world. Remember, the means for excluding worlds in light of assertions is that those worlds at which what is asserted is false on the diagonal are ruled out. The procedure is not *those worlds that individual take what has been asserted to be false on the diagonal* to be ruled out. In this respect, the approach has a component that goes beyond, and is independent of, what the interlocutors know. Given that there is a determinate primary intension for 'water'—which is somewhat of a perilous assumption, given our earlier discussion, but we will provisionally assume it here—there is a determinate primary intension for his assertion, and then there is a fact of the matter whether, at each world considered as actual, that primary intension maps that world to true or false. Given the second-order property associated with the primary intension determined by the first-order property that the students associate with 'water', the second-order property picks out all and only H<sub>2</sub>O, and so the primary intension of his assertion maps this world to true because there is H<sub>2</sub>O in the basin.

Thus, like Stalnaker's view, Jackson and Chalmers's views predict that Bruce's students would not rule out this world in light of his assertion. In the next section, I offer a

diagnosis of what I take to be the common feature shared by these three views that allows this problem to arise, and then argue that we can avoid this problem by excluding this feature from their views.

### ***Section 8: Diagnosis of the Problem***

Stalnaker, Jackson, and Chalmers all encounter this difficulty in accounting for the force of such assertions because they are committed to using the unifying feature of primary intensions, or metasemantic facts for the purposes of reinterpretation, for the purposes of explaining how the audience rules out worlds in light of assertions. That is to say, they are committed to the primary intension (or the metasemantic facts, on Stalnaker's account, though I will subsume them under 'primary intensions' for the purposes of this section, unless otherwise specified) behaving as the reference-fixer for the extension at a given world on the supposition that that world is actual (on reinterpretation, for Stalnaker, or as considered counterfactually, for Jackson and Chalmers), and the extension being uniform, such that the expression picks out all and only samples of that kind, in the case of natural kind terms, or that individual, even when that individual fails to have the property associated with the name, in the case of proper names. It is for this reason that, for example, the primary intension of Bruce's assertion, evaluated at the actual world or another H<sub>2</sub>O world at which there is ice in the basin (or reinterpreted at such worlds, in Stalnaker's case), maps that world to true. It is because it maps such worlds to true that their views result in making bad predictions of the force of the assertion. This is because, in saying that the primary intension of his assertion maps this world to true, their view



predicts that the students would not rule it out. Yet that is precisely the phenomenon we are trying to explain: they *do* rule it out, and it is the sort of mistake that can quite plausibly be made. For this reason, if we are to explain such cases, we are required to either give up the unifying features of the primary intensions in some cases, or to find a different way of accommodating the intuition in these cases, for the purposes of explaining why his students ruled out this world in light of his assertion.

We will adopt the latter approach in the views developed in chapters five and six. As I will discuss in my general diagnostic remarks in chapter four, there are *two* ways of understanding and applying the contingent extension intuition. The views that we've been discussing so far adopt one way of understanding and applying it, and the two views that I develop in chapters five and six adopt the other way of understanding and applying it.

## **Chapter Four: Diagnostics and Background for My Three-Dimensional Proposal**

In this chapter, I introduce my solution to these problematic cases in a general way. In chapter five, I offer an interpretation of the general solution that I present in this chapter that is a proposed extension to Stalnaker's account. In chapter six, I offer one interpretation of this general solution that is a proposed extension to Jackson's account, and another that is a proposed extension to Chalmers's account.

### ***Section 1: An Introductory Sketch of the Proposal***

I propose an extension to Stalnaker's, Jackson's, and Chalmers's accounts in order to accommodate the problematic cases that we've been discussing. My proposed extensions to their accounts does not require them to take on any new semantic or ontological commitments, but only to use in a new way that to which they are already committed. The proposed extensions also allow them to retain their commitment to claiming that individuals have (perhaps only tacitly) both the uniform extension intuition and the contingent extension intuition. In this chapter, I introduce my general proposal for extending their accounts, and, in the chapters that follow, I offer interpretations of the framework that are more specifically tailored to the commitments of their respective views.

In all three cases, my proposed extension of their accounts is part of the pragmatics, not the semantics. In order to accommodate these problematic cases, I am proposing a kind

of pragmatic reinterpretation that, in the following respects, is very much like Stalnaker's metalinguistic diagonalization. We've just mentioned one similarity: both are pragmatic reinterpretations, and so are not part of the semantics. Though I will discuss the details for the proposed extensions for each of their accounts in chapters five and six, there are two important commonalities between my proposed extensions for each of their accounts and diagonalization. First, recall that the reassignment of the extension of expressions for the purposes of diagonalization does not determine the modal profile of the expression, according to the semantics, and does not serve as a contribution of the expression to the truth-conditions of the assertion. Similarly, on the proposed extension for each of their accounts, the reassignment of the extension of the expression does not determine the modal profile of the expression, according to the semantics on their views, and does not serve as the contribution of the expression to the truth-conditions of the assertion. Second, recall that the reinterpretation of assertions for the purposes of diagonalization does not determine the truth-conditions of the assertion. Similarly, the reinterpretation of assertions on my proposed extensions does not determine the truth-conditions for that assertion.

Second, both use that description or property that the interlocutors associate with the relevant expression for the purposes of pragmatic reinterpretation—that description or property is used to reassign the extension of the expression at each world in the context set. Recall that, on metasemantic diagonalization, the description or property that the interlocutors associate with that expression is used to reassign the extension of that expression at each world in the context set, and then the expression, on reinterpretation,

refers uniformly to that individual, or samples of that kind, at that world, even when that individual or some samples of the kind, fail to satisfy that description or to have that property. On my proposed extensions to their accounts, the description or property that the interlocutors associate with the relevant expression is used for the purposes of pragmatic reinterpretation, but, unlike on diagonalization, the reinterpretation does not include the unifying feature: a proper name refers only to that individual when that individual satisfies the description or has the relevant property, and a natural kind term only refers to all and only those samples that satisfy the description or have the relevant property.<sup>60</sup>

Third, whether what is said, on pragmatic reinterpretation at world in the context set, is true at that world, determines whether the conversation should proceed as though that world is ruled out by the assertion. On the face of it, this might seem to be a point of dissimilarity with diagonalization, rather than a point in common. That is, when engaging in diagonalization, don't we look at the diagonal to determine which worlds are *ruled out*? The answer is: it depends on what you mean by 'ruled out'. If you mean *ruled out by virtue of the way the assertion says the world is*, then we already know that the answer is 'no'. Diagonalization, as a kind of pragmatic reinterpretation, is thereby not part of the semantics, and so doesn't capture, on Stalnaker's account, what the assertion says. If you mean that the conversation proceeds as though such possibilities have been excluded, then the answer is 'yes'.

---

<sup>60</sup> Also, unlike on metalinguistic diagonalization, my proposed extensions can accommodate cases in which there is variation in which property or description the interlocutors associate with expression. This is important in cases in which there is variation among the interlocutors regarding which worlds are ruled out in light of the assertion. We will discuss this feature in more detail in section 4.

To see this, consider an assertion of ‘water is H<sub>2</sub>O’, said in conversation with interlocutors who do not already know this. We model the way the assertion changes the state of conversation by having both H<sub>2</sub>O worlds and at least one XYZ world<sup>61</sup> in the prior context set, and then excluding any XYZ worlds from the posterior context set. However, the XYZ world isn’t ruled out by virtue of what the assertion *says* because what the assertion says is necessarily true, and so true at the XYZ world. So, on diagonalization, the XYZ world isn’t *ruled out by the assertion*, if, by ‘ruled out by the assertion’, we mean ruled out by virtue of what the assertion says. Rather, it is ruled out on the pragmatic reinterpretation of the assertion at that world. Importantly, the interlocutors understand that any such alternative hypotheses about the way water could turn out have been excluded. The conversation is to proceed as though these alternative ways the world might turn out (on the epistemic sense of ‘turn out’) have been ruled out, such that, for example, the interlocutors ought to not make subsequent assertions attempting to exclude them.

Similarly, on my proposed three-dimensional pragmatic reinterpretation, worlds aren’t ruled out by assertions by virtue of what that assertion says. Instead, this pragmatic reinterpretation helps us to model how the assertion changes the state of the conversation in such a way that it captures how the conversation should proceed based on what possibilities are excluded, where this is determined not only by the reinterpretation itself but also by which interlocutors, if any, understood what is said on reinterpretation, as

---

<sup>61</sup> Or other world that is superficially indistinguishable from our own, but the predominant clear odorless liquid differs in underlying constitution from H<sub>2</sub>O.

well as other considerations based the interests of the interlocutors, such as the aims of the project(s) in which they are currently engaged by virtue of having the conversation. We'll discuss this in much greater detail in this chapter. At present, the important point to note is that, as with diagonalization, if a world is marked as false on the diagonal, that world is excluded, though the procedure for determining whether a world is ruled out according to the diagonal is more complex on my proposal. Nonetheless, as with diagonalization, the diagonal is very important on my proposed pragmatic reinterpretation.

Fourth, my proposed reinterpretation doesn't require them to adopt any further commitments to worlds beyond what they have already adopted on their current views. It merely requires them to use the same worlds in a different way. For the proposed extension to Stalnaker's account, he only needs the worlds to which he is already committed to for the purposes of appealing to the standard interpretation and, where required, the worlds that theorists employ to model the state of conversation in cases in which they must appeal to diagonalization. Similarly, for my proposed extension to Jackson's and Chalmers's accounts, they need only those worlds required by their current view. Importantly, neither is required to appeal to metaphysically impossible worlds, such as worlds at which Hesperus is Phosphorus or at which water is not H<sub>2</sub>O.<sup>62</sup>

---

<sup>62</sup> In an exchange with Jeff King, Stalnaker concedes that, given his commitment to worlds as maximal properties, he might thereby be committed to a world at which two properties, each of which are metaphysically possible, might be had by the same thing, though that a thing has *both* of these properties is metaphysically impossible, though he argues that even if King is right that his view of worlds commits him to a metaphysically impossible world, it need only commit him to *one* such world. See King's "What in the World are the Ways a World Might Be?" and Stalnaker's reply in his "Replies". That Stalnaker might be committed to one metaphysically impossible world isn't relevant for our interests here. My concern is that my proposed extension to his view doesn't require that he is committed to such worlds. I mention this

Lastly, being able to account for these cases vindates their respective views. That is, not being able to account for the way in which these problematic assertions change the state of the conversation poses a serious challenge to their views. Recall that appealing to diagonalization on Stalnaker's account allowed him to account for the fact that necessary a posteriori assertions are informative. This is important because he is unable to account for this on his preferred semantics. Being unable to account for the fact that these assertions are informative would count as a serious consideration against his preferred semantic view, unless he could somehow account for them in a way that largely fit with his preferred semantics.

This is why the fact that diagonalization is parasitic on standard interpretation, on his view, is so important. Diagonalization allows him to account for the fact that these assertions are informative, and to do so in a way that is very similar to, though not conflicting with, standard interpretation. Since diagonalization isn't part of the semantics, it doesn't conflict with the standard interpretation for these assertions. Rather, in using a kind of pragmatic reinterpretation that proceeds much like standard interpretation, Stalnaker is able to explain why these assertions are informative.

Similarly, appealing to my proposed extensions of their accounts allows them to account for the way in which these assertions change the state of the conversation, which their preferred semantics does not allow them to do. In adopting this pragmatic

---

because I want to make it clear that I'm not claiming that Stalnaker isn't committed to such worlds—I'm only making the weaker claim that my proposal doesn't require that he is committed to such worlds.

reinterpretation that is parasitic on their preferred standard interpretations, this allows them to account for these problematic assertions in a way that both proceeds largely in the same way as standard semantic interpretation, on their account, while failing to conflict with the standard semantic interpretation of these assertions. The latter point is very important because, without the standard interpretation of these assertions, we cannot explain why the interlocutors, in ruling out the worlds that they do in light of the assertion, are making a mistake. I will discuss the details of the way in which my proposed extension of Stalnaker's view is parasitic on his preferred semantics in section 2 of chapter five, the way in which my proposed extension of Jackson's view is parasitic on his preferred semantics in section 2 of chapter six, and the way in which my proposed extension of Chalmers's view is parasitic on his preferred semantics in section 3 of chapter six.

As part of the process of pragmatic reinterpretation of this sort, which I will call the *3D approach*, I appeal to what I call a *three-dimensional framework*, or *three-dimensional matrix*. It is called a three-dimensional matrix because it is constructed by lining up at least two two-dimensional matrices (more precisely, there will be one two-dimensional matrix for each interlocutor, for reasons to be explained in section 5.) in a way that extends along the Z-axis in addition to the X-axis and the Y-axis, as on the two-dimensional matrices. We will discuss the three-dimensional matrix in more detail in section 3, including how to construct them, and the benefits of employing a three-dimensional matrix of this sort.



Once we have discussed my proposed extensions to Jackson's and Chalmers's accounts, I will discuss how their views, extended in this way, can accommodate the assertions that previously proved to be problematic. Next, I will discuss whether their accounts, modified in the ways I propose, have interesting consequences for their claims that sentences such as, 'water is the watery stuff', are knowable a priori, and that conceivability is a good guide for possibility.

### *Section 2: On Knowing How an Individual or Kind is Presented*

As we've seen, the proposals that Stalnaker, Jackson, and Chalmers offer for appealing to the two-dimensional framework to model the force of assertions<sup>63</sup> work well in cases in which the interlocutors know how the individual or kind presents itself under the relevant conditions. However, as we've also seen, this approach cannot model the way that an assertion changes the state of the conversation when the interlocutors are either unsure of, or have false beliefs about, how the individual or kind presents itself under the relevant conditions. In such cases, we will say that the interlocutors do not know how the individual or kind presents under the relevant conditions.

In this section, we'll first develop this notion of knowing how the individual or kind presents under the relevant conditions, and then we'll discuss the importance of being

---

<sup>63</sup> Though we won't discuss this in detail until we cover Stalnaker's view in chapter two, Stalnaker aficionados might be concerned that what I've said here suggests that Stalnaker always uses the two-dimensional matrix to model the force of assertions. To clarify briefly, Stalnaker's view is that we appeal to the two-dimensional matrix only in those cases in which the force of the assertion cannot be explained on the one-dimensional matrix (i.e., as ruling out those worlds at which what is said is false on the horizontal, as discussed above). For a much more detailed discussion of this, see chapter two.

able to model the way that assertions change the state of conversations in cases in which the aim of the conversation is to discover how the world is in some respect, but in which the interlocutors either are unsure about, or have a false belief about, the way that the individual or kind presents under the relevant conditions.

We begin with several questions. First, what is it to know how an individual or kind presents under the relevant conditions? Second, what are the prospects of modeling the force of assertions made in a context in which the audience does not know how the relevant individual or kind is presented, and why should we be concerned with such a project?

I'll begin with the first question. For assertions in which a proper name or natural kind term is a constituent, knowing how the individual or kind presents itself under the relevant conditions plays an important role in allowing the interlocutors to know how the assertion says that the world is, and what possibilities are ruled out by that assertion. If one has the correct expectations regarding how an individual or kind will present under the relevant conditions, then one understands how the world must be in order for what is said to be true.

Recall our example in which Bruce's students are trying to guess which statue, from the series of pictures that he has shown them, is his favorite. When he gives them the hint, 'My favorite statue is made of copper', the students—having the false belief that things made of copper always exhibit the familiar shiny coppery-colored metallic appearance—

rule out the possibility in which the Statue of Liberty is Bruce's favorite statue. They didn't know how copper presents itself under some of the relevant conditions, namely the conditions in which the Statue of Liberty is found. Had they known how copper presents under the relevant conditions, namely those under which the Statue of Liberty is found (i.e., having been outdoors under conditions in which oxidation of copper readily occurs), they would know that copper statues that have been outdoors for extended periods of time exhibit a dull greenish appearance. This would allow them to know that in order for what Bruce said to be true at the possibility at which the Statue of Liberty is his favorite, his favorite statue, given the conditions under which it is found, must present as having a dull green appearance.

Similarly, it was because Bruce's students did not know how water presents under the relevant circumstances that they mistakenly ruled out Alpha. Had they known that water presents as a clear solid under cold conditions, they would have been in a position to know not to rule out Alpha in light of what Bruce said. In this way, the students did not know how the world must be, if the ambient conditions are cold, in order for what Bruce said to be true

We can see the importance of being able to model the way that these assertions change the state of conversations if we consider the important similarities between these cases and the kinds of cases that *can* be successfully accommodated on Stalnaker's, Jackson's, and Chalmers's views, as they presently stand. In both kinds of cases, the interlocutors understand themselves to be engaged in a conversation in which the aim is to collaborate

in narrowing down the ways the world might turn out in some respect (in the epistemic sense of ‘turn out’, i.e., for all they know). Furthermore, in both kinds of cases, the interlocutors understand that, if an assertion is accepted, they ought to rule out those possibilities that aren’t the way the assertion says the world is, and they are to introduce those possibilities that are the way the assertion says the world is, if those possibilities aren’t already under consideration.

However, from the fact that the interlocutors understand that they ought to rule out those worlds that aren’t the way the assertion says the world is, it doesn’t follow that we, as theorists, ought to model the way that assertion changes the state of the conversation by ruling out all and only those possibilities in the prior context set at which what is said is false (or, for Jackson and Chalmers, false on the diagonal). In the problematic cases of interest to us, modeling the way in which the assertion changes the conversation in that way is inappropriate—given what the interlocutors do (or, rather, what they do not) know, they *aren’t in a position* to be able to rule out all and only those worlds at which the assertion is false, on Stalnaker’s view, or false on the diagonal, on Jackson’s and Chalmers’s view, for reasons that we will discuss at length in what follows. The idea is that, even though the interlocutors’ failure to rule out all and only those possibilities in the prior context set at which what is said (or said on the diagonal) is false is what makes it the case that they have made a mistake in ruling out a world that is the way the assertion says the world is, or in introducing a world that isn’t the way the assertion says the world is, it is a mistake to which we cannot hold them accountable because they aren’t in a position to act rule out all and only those possibilities. In other words, in cases

in which the interlocutors don't know how the individual or kind presents under the relevant conditions, they aren't in a position to know what the world must be like in order for that assertion to be true, and so they aren't in a position to rule out all and only those possibilities at which what is said is false.

Instead, the best that the interlocutors are able to do in their aim to rule out all and only those possibilities at which what is said is false is to *approximate* this by ruling out worlds on the basis that what that assertion says at that world, on a kind of pragmatic reinterpretation that we will call *approximating evaluation*, is false at that world. Approximating evaluation is very much like both diagonalization, on Stalnaker's account, and counterfactual evaluation, on Jackson's and Chalmers's accounts, except that it lacks the unifying feature. Recall that, for the purposes of diagonalization and counterfactual evaluation, the extension of a proper name is reassigned at each world on the diagonal as that individual that, on Stalnaker's view, either has the property, or satisfies the description, that we use as the reference-fixer; or, on Jackson's and Chalmers's views, has that property, or satisfies that description, associated with the A-intension or primary intension, and that, on both diagonalization and counterfactual evaluation, the expression picks out that individual uniformly. Recall also that, for the purposes of diagonalization and counterfactual evaluation, the extension of a natural kind term is reassigned at each world on the diagonal as all and only samples of that kind that, on Stalnaker's view, either has the property, or satisfies the description, that we use as the reference-fixer; or, on Jackson's and Chalmers's views, has that property, or satisfies that description, associated with the A-intension or primary intension, and that, on both diagonalization

and counterfactual evaluation, the natural kind term picks out all and only samples of that kind (i.e., picks out the same kind uniformly).

On approximating evaluation, as I noted earlier in this chapter, we also use that property or description that the interlocutors associate with the expression in order to reassign the extension of the expression at each world on the diagonal. Let us call *the approximating intension* the function from worlds to extensions that picks out, at each world, that which either satisfies the description that the interlocutors associate with that expression, or the property that the interlocutors associate with that expression. As I mentioned earlier, the approximating intension does not have the unifying feature: it only picks out the relevant individual when that individual satisfies the description, or has the property, that the interlocutors associate with the proper name, and it only picks out all and only those samples of a kind that satisfy the description, or has the property, that the interlocutors associate with that natural kind term.

Let us call the conditions that the interlocutors use, in these cases in which they don't know how the individual or kind presents under the relevant conditions, to assess whether they should rule out a world in light of an assertion the *approximating truth-conditions* for that assertion. This notion is parasitic on the contingent extension intuition. Recall that the contingent extension intuition is the intuition that, had our world turned out differently in the relevant respects, the extension of our expressions would have turned out differently, too. Having this intuition allows us to entertain how the extensions of our

expressions would turn out on various hypotheses about the way the world turns out in the relevant respects.

For example, as we've said, if we don't know the underlying chemical constitution of water, we can entertain the hypothesis that if the world turns out such that the substance has the qualitative and functional characteristics that we associate with 'water' is H<sub>2</sub>O, then 'water' refers to H<sub>2</sub>O, and we can also entertain the hypothesis that if the world turns out such that the substance that has these qualitative and functional characteristics instead turns out to be XYZ, then 'water' picks out XYZ. This ability to consider hypotheses about the extension of our expressions, on various hypotheses about how the world might turn out (in the epistemic sense) in the relevant respects allows us to have approximating truth-conditions: the conditions that we use to approximate the truth-conditions for an assertion. The approximating truth-conditions are determined by the approximating intensions for the constituent expressions.<sup>64</sup> Since interlocutors who don't know how the individual or kind presents under the relevant conditions, the best they can do is to rule out those worlds that fail to satisfy their approximating truth-conditions for the assertion. Under which if that world turns out to be actual and to satisfy the conditions, then that assertion is true. That the approximating intension for a proper name or natural kind term serves as that expression's contribution to the approximating truth-conditions is the feature of the approximating truth-conditions that makes them useful for explaining the way in which assertions change the state of conversations in cases in which the

---

<sup>64</sup> For present purposes, we will treat all expressions other than proper names and natural kind terms as unproblematic. Let us suppose that these unproblematic expressions make the same contribution to the approximating truth-conditions as they make to the truth-conditions. This simplifying assumption allows us to focus on proper names and natural kind term for our present purposes.

interlocutors don't know how the individual or kind presents under the relevant conditions.

To see this, let's consider Bruce's assertion again. The students believe that water always presents as a clear odorless liquid—they don't know that it presents as a clear solid under cold conditions, or as a gas under very hot conditions—and they believe it to be the *predominant* clear odorless liquid. Thus, the property that they associate with water is being the predominant clear odorless liquid. Given the contribution that the approximating intension makes to the approximating truth-conditions, the approximating truth-conditions for Bruce's assertion are that there must be a sample of the predominant clear odorless liquid in the basin, presented as a clear odorless liquid. Since, at Alpha, there isn't a clear odorless liquid in the basin, they rule out Alpha in light of Bruce's assertion because it fails to satisfy the approximating truth-conditions for Bruce's assertion.

Let's consider another example. Recall again the favorite statue exercise. The students don't know how copper presents itself when it has been exposed to conditions under which it oxidizes—they expect copper to always present as a shiny copper-colored metal. For this reason, being a shiny copper-colored metal is the property that they associate with 'copper', and so this is the property that we use to determine the extension of the approximating intension for 'copper' for the purposes of approximating evaluation. The approximating truth-conditions for Bruce's assertion, 'My favorite statue is made of copper', guides the students to rule out any possibilities at which his favorite state fails to



exhibit the shiny copper-colored metallic appearance. This explains why they rule out the possibility at which the Statue of Liberty is his favorite statue.

As we can see, in cases in which the interlocutors do not know how the individual or kind presents under the relevant conditions, allowing for a kind of reinterpretation on which the proper names and natural kind terms do not refer uniformly, but only to that which has the property, or satisfies the description, that the interlocutors associate with the expression, allows us to explain why the interlocutors mistakenly ruled out possibilities.

Lastly, in section 5.2, we will discuss the need for a means of modeling how an assertion changes the state of a conversation in cases in which the interlocutors have different expectations of the characteristics had by samples of that kind under the relevant conditions. Though two individuals may not know, or may have false beliefs about, how an individual or kind presents under the relevant conditions, this may be due to having different expectations of the characteristics had by the individual, or samples of the kind, under the relevant conditions, and this can have as a consequence that they may rule out different worlds in light of the same assertion. Furthermore, we will also want a means of being able to model the force of assertions in cases in which some but not all of the interlocutors do not know, or have false beliefs about, how the individual, or samples of the kind, present under the relevant conditions, for the same reason. In the approach that we will develop in the next section, we accommodate these differences in expectations, and so these differences in which worlds are ruled out in light of assertions.

### *Section 3: A Solution Employing a Three-Dimensional Framework*

In this section, I will introduce a basic general three-dimensional framework that, in chapters five and six, respectively, I will suggest as an extension of Stalnaker's, Jackson's, and Chalmers's two-dimensional approaches as a means of responding to the problem of informativity. I begin this section by discussing my diagnosis of why Stalnaker's, Jackson's, and Chalmers's views are unable to explain why Bruce's students ruled out Alpha in light of Bruce's assertion. Next, employing the notions of approximating counterfactual evaluation and approximating truth-conditions, as developed earlier in this chapter, I discuss how a three-dimensional framework can be developed as an extension of their views that can serve as a solution to this problem.

#### *Section 3.1: A Sketch of the Three-Dimensional Framework*

This three-dimensional proposal makes use of only resources to which Stalnaker, Jackson, and Chalmers are already committed in their current two-dimensional approaches, and so is a solution that is available to them without requiring additional semantic or ontological commitments. The three-dimensional framework, on my proposal, is so-called because it is represented within a three-dimensional matrix. The matrix is three-dimensional because it is constructed from more than one two-dimensional matrix lined up in such a way that the cells from each two-dimensional matrix match up, as explained in section 5. That there are multiple two-dimensional frameworks matched up in this way extends the matrix it along the z-axis in addition to the x-axis and y-axis.

The central motivating idea is that we can retain the attractiveness of Stalnaker's, Jackson's, and Chalmers's two-dimensional approaches to modeling the force of assertions made in contexts in which we can successfully appeal to the standard semantic interpretation while also responding to the challenging cases that cannot be accounted for by appealing to the standard interpretation of the assertion, and that we can do this in such a way as to only require the resources of their present accounts rather than needing to adopt additional commitments. We can think of this approach as an extension of their two-dimensional approaches, rather than a departure from them, because it *incorporates* the resources of the two-dimensional framework that they employ, modifying these resources slightly in a way that is parasitic on standard interpretation for the purposes of pragmatic reinterpretation, and then extends it into a three dimensional matrix by adding additional two-dimensional matrices. I will explain how these additional two-dimensional matrices are to be constructed, as well as how they, when combined into a three-dimensional matrix, allow for a solution to the problem of informativity, in what follows.

Furthermore, as I claimed at the beginning of this chapter, my solution of employing the three-dimensional framework makes use of the contingent extension intuition and the uniform extension intuition. However, they are employed in a slightly different way than they are on Stalnaker's, Jackson's, and Chalmers's accounts, for the purposes of constructing additional two-dimensional matrices that form a three-dimensional matrix. More specifically, we'll see that progress can be made on a solution to the problem of

informativity by applying the uniform extension intuition differently in ruling out possibilities for updating the context set in light of Bruce's assertion, as we will discuss in section 4.1. We will now turn our attention to the general procedure for constructing the three-dimensional matrix.

#### ***Section 4: Constructing the Three-Dimensional Matrix***

In this section, I will discuss the construction of the three-dimensional matrix from the resources already in use in Stalnaker's, Jackson's, and Chalmers's accounts. In section 4.1, I explain how the intuitions can be used in a slightly different way to both account for the problematic assertions in a way while still allowing Stalnaker, Jackson, and Chalmers to retain their commitment to these intuitions. In section 4.2, I illustrate how to use these approximating intensions to fill in each cell of the two-dimensional matrix for each interlocutor. In section 4.3, I explain how to construct the three-dimensional matrix from the two-dimensional matrices constructed in 4.2, and how to use the three-dimensional matrix to explain why the students ruled out the actual world in light of Bruce's assertion. In section 4.4, I use the three-dimensional matrix to explain the way in which the speaker takes the assertion to update the state of the conversation for the purposes of updating his beliefs about the presuppositions, and also for the purposes of determining what things he should (and should not) say next, based on those presuppositions.

##### ***Section 4.1: Preserving the Intuitions***

In chapter one, we discussed the uniform extension intuition and the contingent extension intuition. I argued that it is quite plausible that, at least tacitly, we have such intuitions, and that they play a role in explaining how we rule out possibilities in light of assertions. However, it seems that the uniform extension intuition is very much at odds with developing a solution to the cases that have been found to be problematic for Stalnaker, Jackson, and Chalmers. If ‘water’ uniformly picks out all and only samples of the same kind, and if, at our (H<sub>2</sub>O) world, ‘water’ picks out all and only H<sub>2</sub>O, then if we use this intuition in constructing the two-dimensional matrix, it looks like, on the face of it, any proposed solution that accommodates the uniform extension intuition is going to encounter the problem of informativity in just the same way that Stalnaker, Jackson and Chalmers did.

We have a few options for resolving the problem of informativity as it arises in this case, which I will use to illustrate my more general proposal for resolving the problem. First, we could attempt a resolution on which we hypothesize that Bruce’s students do not have the uniform extension intuition. That is to say, we could argue that, though it might have seemed plausible, while considering the anti-descriptivist examples, that individuals have these intuitions (if only tacitly), it turns out that, as we can see in looking at which possibilities they rule out in light of assertions, the interlocutors must not have this intuition (even tacitly) after all.

This would offer *an* explanation of why they ruled out Alpha in light of Bruce's assertion. If they ruled out Alpha because they weren't committed to taking 'water' to refer to all and only samples of the same kind, this is a potential explanation of why they ruled out the actual world in light of Bruce's assertion. However, it isn't a very *plausible* explanation. As we discussed when the example was first introduced, Bruce's students take 'water' to be a natural kind term, and they understand that natural kind terms uniformly refer to all and only samples or members of the same kind.

For this reason, I recommend a second, and, I think, more plausible, solution. Progress can be made by reflecting on how we should make use of the fact that the students have the uniform extension intuition for the purposes of modeling the way the state of the conversation changes in light of Bruce's assertion, and in other cases in which they don't know how the individual or kind presents under the relevant conditions. We can make sense of both the fact that the students take 'water' to be a natural kind term, and so to refer uniformly, *and* the fact that they rule out Alpha in light of Bruce's assertion, by using the uniform extension intuition in a slightly different way in modeling the way in which the state of the conversation changes in light of Bruce's assertion. Instead of modeling the change in the state of the conversation by excluding those worlds on which what is said at that world is false at that world (i.e., either on the horizontal or on the diagonal) using the standard matrix, we will, for these purposes, allow, for the purposes of a kind of pragmatic reinterpretation, the extension of 'water' to be such that it includes all and only those samples to which the students take it to apply—which they take to mean that the expression thereby applies uniformly—and so model the state of the

conversation to change in such a way that the worlds that the students take to be such that what is said at that world is false at that world are ruled out.

To better understand my proposal, it helps to think about it from the students' perspective. The students take 'water' to apply uniformly to all and only samples of the same kind. The problem is that they believe that samples of *ice* don't belong to that kind. That is, though they take 'water' to apply to all and only samples of the same kind, they believe that ice belongs to a distinct kind, and so they don't believe that ice is in the extension of 'water'.

What I am proposing here is a different way of accommodating the uniform extension intuition, for the purposes of a kind of pragmatic reinterpretation that I argue can be used to account for these problematic cases. For the purposes of engaging in this kind of pragmatic reinterpretation in these cases, we allow the extension of the proper name or natural kind term to be reassigned using either a property or a description that captures the qualitative and/or functional characteristics that the interlocutors expect the individual or kind to exhibit. In engaging in this pragmatic reinterpretation, the extension of that proper name or natural kind term is assigned as that individual, or those samples of that kind, that satisfies the description or has the property, respectively.

#### *Section 4.2: Constructing the Two-Dimensional Matrices*

The two intuitions, with the uniform extension intuition being applied in this new way, will be used to explain why the students ruled out Alpha in light of his assertion. In

constructing the three-dimensional matrix, we will first need to construct an individual matrix for each interlocutor—these individual two-dimensional matrices will combine to form the three-dimensional matrix for that assertion, as we will discuss in section 5.

The purpose of having a two-dimensional matrix for each interlocutor, on this pragmatic reinterpretation, is that it can accommodate differences between interlocutors regarding which worlds they take to be ruled out by an assertion. This is valuable because, in at least some cases, only some of the interlocutors don't know how the individual or kind presents under the relevant conditions, or the interlocutors differ in their false beliefs about how the individual or kind presents under the relevant conditions.

We will now look at how to construct an individual two-dimensional matrix for the purposes of this kind of reinterpretation, and then we will discuss how to combine them to form the three-dimensional matrix that we can use to explain both why the students took Bruce's assertion to rule out Alpha, and also to answer the question of how to model the way in which the assertion changes the state of conversation in cases in which some, but not all, of the interlocutors take the assertion to rule out a particular world.

First, let's look at the role that the contingent extension intuition plays in formulating a two-dimensional matrix on this approach for one of his students—call her Alice. Alice—perhaps only tacitly—understands that what 'water' refers to depends in part on the way the world turns out, in an epistemic sense of 'turns out': if it turns out that the stuff that fills the oceans, lakes, and streams, and falls as rain, and fills any other functional roles or



has any other observable characteristics that she associates with ‘water’, is H<sub>2</sub>O, then ‘water’ picks out H<sub>2</sub>O, but if that stuff instead turns out to be XYZ, then ‘water’ picks out XYZ. We can capture her grasp of this by, at each world in the context set, allowing ‘water’ to refer to those samples that either satisfy the description, or have the property, at that at that world. It’s important to note that there will be such a substance at each world in the context set since Alice (and her classmates) all know that there *is* a substance with these qualitative and/or functional characteristics, though they don’t know its underlying constitution.

As noted earlier, Alice, like her classmates, takes ‘water’ to be a natural kind term, and so takes it to refer uniformly to all samples of the same kind. However, in her present epistemic state, Alice isn’t in a position to accurately assess, in all situations, whether a given sample of a substance *is* in the extension of ‘water’ or not. While she does have a something of a grasp on the truth-conditions of Bruce’s assertion, at least to the extent that she understands that his assertion is true just in case there is *water* in the basin, she isn’t in a position to correctly assess, in all the relevant conditions, that the truth-conditions are satisfied—to the contrary, in some cases, such under the conditions in Alpha, she thinks they are not satisfied. That is, like her classmates, Alice expects water to be presented as a clear odorless *liquid* that constitutes oceans, rivers, lakes, streams, etc.; falls as rain; comes out of the taps in homes; quenches thirst; and so on. Ice fails to meet her expectations, and so she does not take ice to be in the extension of ‘water’. However, in reassigning the extension of ‘water’ to all and only those samples that have the qualitative and/or functional characteristics that she associates with water, we

accommodate the uniform extension intuition because this does include all and only those samples that she takes to be water.

This is why Alice's matrix succeeds where the standard matrix cannot: her matrix can explain why she ruled out the actual world in light of Bruce's assertion. Since she believes that what Bruce said is false at Alpha, she rules out that possibility (though she wouldn't describe it as ruling out the actual world—she'd likely describe it as ruling out the possibility that there is ice in the basin). The standard matrix doesn't allow this because what Bruce says, is *true* at Alpha.

Recall that on Stalnaker's account, the worlds at which what is said at that world (on reinterpretation at that world) is true at that world—i.e., those worlds at which what is said (on reinterpretation) is true on the diagonal—are in the posterior context set, and that, on Jackson's and Chalmers's approaches, those worlds which the primary intension of the sentence maps to true, are in the posterior context set. Just so that we have a neutral expression (rather than the diagonal, the primary intension, or the A-intension, as on Stalnaker's, Chalmers's and Jackson's accounts, respectively) that refers to those worlds at which what is said is true on the diagonal, call this the *diagonal intension*.

We will call any two-dimensional matrix constructed as described in what follows an *individual approximating matrix*, and any three-dimensional matrix constructed from approximating individual matrices an *approximating matrix for the assertion*. They are so called because they capture the way the interlocutor takes the state of the conversation

to change in light of the assertion, based on his individual approximating intensions for natural kind terms that are the constituent expressions of the assertion, and so by virtue of the assertion ruling out those worlds at which the approximating truth-conditions fail to be satisfied.

Recall that an approximating intension for an expression is a function from worlds to extensions that picks out all and only those samples that have the qualitative and functional characteristics that individual expects them to exhibit under the relevant conditions. For each natural kind term that has any significance at all to an individual, that individual will have expectations regarding how that which is in the extension of the term will be presented under various conditions. The individual's approximating intension for a proper names captures the (perhaps only tacit) criteria that she has for judging that an individual is in the extension of that proper name. The individual's approximating intension for a natural kind term captures the (perhaps only tacit) criteria that she has for judging that a given sample is in the extension of the kind term. In the next section, we will discuss the notion of *roles* as a means of capturing all and only those things that satisfy what a given interlocutor takes to be the qualitative and/or functional characteristics had by that which is in the extension of a proper name or natural kind term.

#### *Section 4.3: Roles*

Let us say that anything that has qualitative and/or functional properties that an interlocutor associates with a proper name, N, plays the N-role for that individual. If an

individual plays the N-role for that interlocutor, then that interlocutor judges that individual to be in the extension of ‘N’. Similarly, let us say that anything that has qualitative and functional properties (relativized to the appropriate conditions) that an individual associates with a natural kind term, K, plays the K-role for that individual. If a sample plays the K-role for that individual, then that individual judges that sample to be a sample of kind K, and to thereby be in the extension of ‘K’.

For an example of an N-role, consider the *Hesperus-role* for an individual that associates with ‘Hesperus’ the property of being the brightest celestial body visible just after dusk. This individual’s criterion for playing the Hesperus-role is being the brightest celestial body visible just after dusk. For the purposes of pragmatic reinterpretation of this kind, the extension of ‘Hesperus’ is reassigned at each world in the context set as that which is the brightest celestial body visible just after dusk at that world. This means that, at a world at which Jupiter is the brightest celestial body visible just after dusk, the extension of ‘Hesperus’ at that world is reassigned as Jupiter.

For an example of a K-role, consider the *water-role* for Alice. We will say that anything that has the property that she associates with ‘water’ plays the water-role for her, and so she judges it to be in the extension of ‘water’. Alice’s criteria for playing the water-role—that is, the qualitative and functional characteristics that she associates with ‘water’—may include being a substance that has certain qualitative characteristics, such as being a clear odorless liquid, as well as certain functional characteristics, such as being

found in rivers, lakes, streams, and oceans; quenches thirst; comes out of taps; being the sort of thing one uses to wash one's dog; and so on.

That which plays the water-role for an individual at a world is thereby in the extension of that individual's approximating intension for 'water' at that world. The criteria for playing the water-role for Alice is determined by the qualitative and functional characteristics that Alice expects water to have under various conditions. Furthermore, she understands that 'water' picks out all and only samples of the same kind uniformly. Since what plays the water-role is what she takes to be water, she takes that which plays the water-role to be in the extension of 'water'. What she takes to be in the extension of 'water' at a world is thereby the extension of her individual approximating intension of 'water' at that world.

Note that we need not take this to be a failure of epistemic modesty on her part. We need not take her to have a firm considered conviction about whether that which she takes to be in the extension of 'water' is, in fact, *all* that is in the extension of 'water'. She might well have not thought about it, and, if we asked her, she might be willing to say that perhaps there could be some things in the extension of 'water' that aren't presented in the way that she expects. But, pre-reflectively, she might not take herself to have a reason to suspect that 'water' presents itself in ways that outstrip her expectations, and so, at least tacitly, believe that since that is the way that water presents itself, then all and only those things so presented are in the extension of 'water'.

The more Alice learns about water, the better of an approximation of the extension of ‘water’ her individual approximating intension for ‘water’ becomes. This is why I have called it an approximating intension—it approximates the extension of the expression. However, not *all* changes to an individual’s approximating intension make the approximating intension a better approximation of the extension of the expression than it had been prior to the change. That is, though changes to our approximating intensions for an expression are generally such that they become better approximations of the extension of the expression, this isn’t *always* the case. Sometimes we acquire new *false* beliefs about that which is in the extension of the expression, and this will change our approximating intension for that expression for the worse.

For example, suppose that Alice is reading a fictional story, though she falsely believes that it is not fiction. In this fictional story, the author says that at 37 degrees Fahrenheit, liquid water turns into its solid state as a purple crystalline structure. Alice, believing that this is true, updates her expectations about the way in which water is presented in conditions in which the ambient temperature is 37 degrees Fahrenheit or less: she now expects that, under those conditions, water will not be presented as a clear odorless liquid, but instead as a purple crystalline structure.

Prior to reading this story, her expectations about how water would be presented under various conditions was that it is presented as a clear odorless liquid under all conditions, regardless of the ambient temperature. In other words, she believed that water only has one state, namely its liquid state. Her individual approximating intension for ‘water’

prior to reading the story is a better approximation of the extension of ‘water’ than her individual approximating intension for ‘water’ after reading the story. One way to see this is to compare the extension of ‘water’ at each world with the extension of her individual approximating intension before the story, and with the extension of her individual approximating intension after the story. Considered across worlds, there will be more overlap between the extension of ‘water’ at a given world and her individual approximating intension prior to the story at that world than there is between the extension of ‘water’ and the extension of her individual approximating intension for ‘water’ after the story.

We need to accommodate not only true beliefs about the way that water is presented under various conditions, but also false beliefs about it, as well. This will allow us to explain why the state of the conversation changes as it does in cases in which that which is in the extension of the natural kind term in question isn’t presented in the way the interlocutors expect. We will see this in action in section 5, after we’ve discussed the individual approximating intension for assertions, individual approximating matrices, and approximating matrices for an assertion.

From what I have said so far, we are in a position to see that there can be variation in the conditions for playing the water-role across individuals. If, like Bruce’s students, an individual only knows that water has a liquid state—that is, she doesn’t know that water has a solid or gaseous state—then her criteria for playing the water-role will only include that the substance is a clear odorless liquid. If another person knows that water has a

solid state under cold conditions in addition to a liquid state in warm conditions, her criteria for playing the water-role will include substances that are presented as a clear solid under cold conditions, and substances that are presented as a clear odorless liquid under warm conditions.

There is also variation in conditions for playing the water-role for the same individual across time, as she learns more about water, or at least acquires further beliefs about the nature of water. I say the latter because, as we've discussed, an individual's false beliefs also shape the individual's criteria for playing the water-role: the criteria for playing the water-role for that individual is determined by the ways in which the individual expects water to present itself under various conditions, even if some of these expectations are mistaken.

For example, at the time of Bruce's assertion, Alice expects water to present as a clear odorless liquid under all conditions (because she doesn't know that water has other states), and so her criteria for playing the water-role is that the substance is a clear odorless liquid. However, if Alice later learns that water also has a solid state, and that which state it is in is determined by the ambient conditions—liquid in warm conditions, and solid in cold conditions—the criteria for playing the water-role for her changes to include that the substance is presented as a clear odorless liquid under warm conditions, and is presented as a clear solid under cold conditions.



Similarly, recall our statues example: Bruce's students rule out the possibility that the Statue of Liberty is his favorite in light of his hint, 'My favorite state is made of copper'. Each student rules this out because his or her criteria for playing the copper-role is governed by that student's beliefs about how copper is (or would be) presented under various conditions. Suppose that Alice is familiar with copper pennies, copper kettles, copper wiring, and so on, and knows that they are made of copper, but doesn't know that copper can undergo any chemical changes that result in an appearance other than the familiar copper-colored metallic appearance, such as oxidation, so she believes that copper is always presented as a shiny copper-colored metal.

In her present epistemic state, the criteria for playing the copper-role for her include being presented as a shiny copper-colored metal—anything not presented in that way will fail to play the copper-role for her. As she learns more about copper, her beliefs about the way that copper is presented under various conditions will change, too. She'll learn that those old green statues in the park—and the Statue of Liberty—are made of copper, too, and that, more generally, oxidized copper presents as a green dull solid. This leads to a change in the criteria for playing the copper-role for her: the substance must be presented as a shiny copper-colored metal unless it's been exposed to the outdoors or other conditions under which it would become oxidized, in which case it must be presented as a dull green solid.

On a related note, samples of more than one kind can play a given K-role for an individual, too, if the individual's beliefs about how the substance is presented are such

that samples of more than one kind meet those criteria. For example, if an individual's gold-role is such that the substance must be a shiny golden metal, then samples of pyrite and samples of gold that are presented as having that appearance will both play the gold-role for that individual (provided that there aren't other criteria for playing that gold-role that aren't met by them).<sup>65</sup>

#### *Section 4.4: Default Reasons for Judging that a Sample is in the Extension of K*

The qualitative and functional characteristics of a substance that are the criteria for playing the relevant N-role or K-role for that individual serve as a *default* reason for that individual to take an individual to be in the extension of N, or a sample to belong to kind K. By a default reason, I mean that this serves as a reason to judge that individual to be in the extension of N, or that a sample is of kind K, unless further information (from a reliable source, such as observation, or an assertion made by someone who the individual takes to plausibly be informed about the matter) that is contrary to the expectations about how the individual named by N, or samples of kind K, is presented under certain

---

<sup>65</sup> I am developing this as though it is either the case that a given sample clearly does, or clearly not does not, play a given role, as though the individual had absolute confidence in their judgment of whether the sample was of the relevant kind or not. This is merely for the sake of simplicity. A way that I would like to develop it in the future is to add some sort of weighting based on the individual's credence about whether that sample belongs to that kind. For example, I might initially be highly confident that ice isn't water. However, as I acquire more experience with water, I begin to notice that ice melts into something that looks an awful lot like water. Perhaps I only notice it at first in a fairly minimal way: I notice that when I hold an ice cube in the palm of my hand, it appears to leave a clear odorless liquid. Initially, this might make me only slightly doubt my judgment, but the more evidence, the more I doubt my hypothesis. I'd like to model approximating intensions as functions from worlds to truth values, but without being such that these are made with either credence 1 or 0. It might be that I take it that the counterfactual truth-conditions for an assertion are satisfied at a given world, but only with credence .65. If I'm unsure, I should be taken to have credence .5. So, the values need not be just credence or 1 or 0: they could be values between 1 and 0, reflecting the individual's level of confidence. This is especially important since one of the kinds of conversations I am most interested in modeling is discussions of cutting-edge discoveries in research, such as the conversation in the physicist example in chapter four.

conditions is introduced<sup>66</sup>. When this further information is introduced, if it is accepted, then the default reason is overridden, and, often, this may require a new default reason that is updated in light of the new information<sup>67</sup>. That a sample has these qualitative and functional characteristics, and thereby meets the criteria for playing the K-role for her at that time, is the best reason she has (in that epistemic state) to judge that the sample is of kind K. If she acquires further information to the contrary from a reliable source, then her previous best reason is no longer the best reason that she has—her reasons to judge to the contrary are now stronger, and so override her default reason. Her criteria for playing the K-role updates to incorporate this new information, and becomes her new default reason for judging that a sample is of that kind, provided that it meets the new criteria.

Let's look at this more closely. Since these qualitative and functional characteristics that serve as the criteria for playing the K-role capture her prior expectations of how a sample of K would be presented under these conditions, and she has these expectations based on what she has observed of samples of K (or at least what she has observed of that which she has *judged* to be samples of K), and what she has been told about the nature of K, then, in the absence of further information about how samples of K are presented under these conditions, her prior expectations provide a default reason for her to judge that the sample is of kind K. Once she is introduced to further information (either through

---

<sup>66</sup> Compare with Peacocke's *primitively compelling inferences*, about which he writes: "To say that a thinker finds such transitions primitively compelling is to say this: (1) he finds them compelling; (2) he does not find them compelling because he has inferred them from other premises and/or principles; and (3) for the possession of the concept C in question...he does not need to take the correctness of the transitions as answerable to anything else" (1992, 6). These default reasons are like the primitively compelling inferences with respect to (1) and (2), but not (3). The default reasons *can* be overridden in light of some new information (e.g., further information from an authority on the topic, novel observations, etc.) Thanks to Paul Pietroski for suggesting consideration of this comparison.

<sup>67</sup> It appears that, on Chalmers's view, this this will count as an instance of conceptual change. See chapter five of his *Constructing the World*.

observation or testimony) that is contrary to her prior expectations, and she has reason to accept the observation (e.g., fails to have reason to think that she is hallucinating, or that she needs a new prescription for her glasses, etc.) or testimony (e.g., she has reason to think that the testifier knows what she is talking about), then this new information both overrides her previous reason, and gives her a reason to update her criteria for playing the K-role. The new information thereby provides a more stringent reason for judging that the sample is not of that kind because she, like the rest of us, updates her expectations about the way samples of K are presented in light of further (unexpected) observations of K and further informative assertions about K.

For example, given that I expect that water presents as a clear odorless solid below 32 degrees Fahrenheit, I might expect a sample of water at 0 degrees Kelvin (-459.67 degrees Fahrenheit) to present as a clear odorless solid, too. Suppose that, under very special laboratory conditions, I am able to view what appears to be a clear odorless solid at 0 degrees Kelvin. Given that the sample has both the qualitative (clear odorless solid) and functional (found in cold temperatures) characteristics that serve as my criteria for playing the water-role, this serves as a default reason for me to think that it *is* water in its frozen state. However, suppose that, unbeknownst to me, gold becomes a clear solid at that temperature, and water becomes a purple crystalline structure at that temperature. Upon being told this further information by, say, one of the experts working in the lab in which the observation takes place, I now have a reason for think that it is *not* water, and that new reason overrides my default reason. Furthermore, I update my criteria for playing the water-role to include how water is presented at 0 degrees Kelvin.

We have a few options regarding whether to think of these criteria for playing the K-role as *necessary* conditions, as *sufficient* conditions, or both, for being a default reason for the individual to judge that the substance belongs to K. By a necessary condition for being a default reason, I mean that if the substance doesn't have characteristic C, where C is among the criteria for playing the K-role, then the individual will not judge that the sample of the substance belongs to kind K. In other words, the substance must have C in order to be judged by that individual as belonging to kind K. By a sufficient condition for being a default reason, I mean that if a substance has C, that is sufficient for the individual to judge that the sample belongs to kind K, though having C isn't necessary for being judged to be a K by that individual—that individual could judge it to be a sample of kind K by virtue of having another sufficient condition for being a K, instead.

It may be that this varies by case, varying across individuals, or even for the same individual across times as her epistemic state changes. For example, prior to learning that there are other kinds of clear odorless liquids, that a sample of liquid is clear and odorless provides Alice with a default reason for thinking that the liquid is water, and behaves as both a necessary and sufficient condition for her judging that it is water: she judges anything that has these characteristics to be water, and does not judge to be water anything that does not have these characteristics

We can see that it is a sufficient condition to serve as a default reason for her to judge that it is water because, in having that appearance, she judges that sample of a substance

to be water. For example, even if she were presented with a clear odorless liquid under very cold conditions, she would judge it to be water because it has the appearance that she expects water to have, and she doesn't know that water changes to a solid under those ambient conditions. We can see that this is a necessary condition to serve as a default reason for her to judge that it is water by considering that she judges water in its solid state to not be water—she judges it to be of a different kind.

However, once Alice is aware that there are other kinds of clear odorless liquids, this appearance of being a clear odorless liquid no longer provides a *sufficient* condition for judging that it is water. Similarly, once she learns that water has other states, such as being a solid or a gas, these characteristics no longer serve as a *necessary* condition for her to judge that the sample is water.

I think that there is a case to be made that some combinations of criteria serve as *jointly sufficient conditions* for serving as a default reason to judge that the sample is of kind K.<sup>68</sup> For example, being a clear odorless liquid is insufficient, by itself, to serve as a default reason to judge that the sample of the substance is water, if one has reason to believe that there are other clear odorless liquids, or that water has states other than being a liquid. However, together with having some functional characteristics, such as exhibiting that appearance under conditions in which one thinks that it is exceedingly likely that any sample of a clear odorless liquid under those conditions is a sample of

---

<sup>68</sup> Jackson and Chalmers entertain roughly this idea in “Conceptual Analysis and Reductive Explanation”. See their discussion in section three. Chalmers also entertains a similar idea in *Constructing the World*.

water, or being found under warm ambient conditions in that case, serve as jointly sufficient for being a default reason to judge that the sample is water.

To illustrate, suppose that I see a clear odorless liquid *coming out of my kitchen tap*. The qualitative characteristic of being a clear odorless liquid, together with the functional characteristic of coming out of my tap, together are jointly sufficient to serve as a default reason to think that the liquid is water. From this, we can see that, at least in that epistemic state, being presented as a clear odorless liquid is insufficient to serve as a default reason to judge that the liquid is water, but being presented as a clear odorless liquid that also has the functional characteristic of coming out of my kitchen tap *is* jointly sufficient.

However, if I see a sample of a clear odorless liquid *in a test tube in a chemistry laboratory*, the qualitative characteristic of being a clear odorless liquid together with the functional characteristic of being found in a test tube in a chemistry laboratory don't constitute jointly sufficient conditions to serve as a default reason for me to believe that the liquid is water. This is because I both know that water isn't the only clear odorless liquid, and I don't have a reason to think that water is the *most likely* clear odorless liquid that could be found in a chemistry laboratory. From this, we can see that, at least in that epistemic state, being presented with the qualitative characteristics of being a clear odorless liquid, together with the functional characteristic of being in a test tube in a chemistry laboratory, is not jointly sufficient to serve as a default reason to judge that the liquid is water.

Similarly, as we've discussed, the ambient conditions under which a sample is presented make a difference to whether that sample is judged to be of a particular kind or not, based on the interlocutor's expectations of how a sample of that kind would be presented under those conditions. That a sample has the qualitative appearance of being a clear odorless liquid is insufficient to serve as a default reason to judge that the sample is water, if the interlocutor knows that water has more than one state. Suppose that the interlocutor knows that water is a liquid under warm conditions, and that water is a solid under cold conditions. If a sample of a clear odorless liquid is presented to her under warm ambient conditions, this is jointly sufficient to serve as a default reason to judge that the sample is water. If a sample of a clear odorless liquid is presented to her under cold ambient conditions, this is not jointly sufficient to serve as a default reason to judge that the sample is water. If a sample of a clear odorless solid is presented to her under cold ambient conditions, this is jointly sufficient to serve as a default reason to judge that the sample is water. If a sample of a clear odorless solid is presented to her under warm ambient conditions, this is not jointly sufficient to serve as a default reason to judge that the sample is water.

#### *Section 4.5: Credence Values in the Approximating Matrix*

We are not always highly confident in our judgments about natural kinds, especially when we are learning about a new kind, or learning something new about a kind with which we are previously familiar. The stakes play a role in how confident we are, too:



As the importance that one gets it right increases, one might reasonably suspect that one's confidence decreases.<sup>69</sup>

One might be less confident in one's judgment that a sample belongs to a particular kind in some contexts than in others. For example, I might be more confident in my judgment that it is water in the test tube if I'm told that we aren't going to do anything with the substance than if I'm told that, if I judge that it is water, then I'll be required to drink it, or I'm told that if it is anything other than water, when we combine it with a mixture in the next test tube, it will explode.

Note that this might also change depending on the context or the project in which I am currently engaged. For example, even in some cases in which I am considering the clear odorless liquid coming out of my kitchen tap, these characteristics may not serve as jointly sufficient to be a default reason for me to judge that it is water. Consider a case in which we are observing a clear odorless liquid coming out of my kitchen tap as you fill up my glass, and say, 'here's a glass of water'. In this context, it seems that the qualitative and functional characteristics serve as a default reason to judge that the liquid is water. In such a situation, I have high confidence that what you've said is true.

However, suppose that you inform me that, say, you heard something from your friend who works at the Health Department about the government quietly substituting as the

---

<sup>69</sup> This point bears a similarity to contextualism in epistemology. The higher the stakes for knowing something, the less likely we are to attribute knowledge to that individual. For example, see DeRose, "Contextualism: An Explanation and Defense", in J. Greco and E. Sosa, ed., *The Blackwell Guide to Epistemology*, Blackwell Publishers, 1999. Also, his "Contextualism and Knowledge Attributions", *Philosophy and Phenomenological Research*. Vol. LII., No. 4, Dec 1992.

liquid coming out of taps a man-made chemical compound that is qualitatively indistinguishable from water, though your friend is unsure whether the switch has, or will, take place, and, if so, when it will occur. This might reasonably serve as a defeater: that is, it might serve to weaken the strength of my reason for judging that there is water coming out of the tap. This further information ought to decrease my confidence that what you've said is true.

In section 5.2, I will discuss how to accommodate these credences. In the next section, we will discuss the relationship between default reasons and individual approximating intensions.

These default reasons for an interlocutor to judge that a sample is, or is not, in the extension of kind term K elucidate that individual's approximating intension for K. A sample is in the extension of that individual's approximating intension for K just in case that sample has the qualitative and/or functional characteristics that lead the interlocutor to judge that the sample belongs to that kind. The interlocutor's default reasons for judging that a sample is of the relevant kind capture the conditions under which the interlocutor judges that a sample is of that kind, and so capture the conditions under which that sample is in the extension of the individual's approximating intension for K.

#### *Section 4.6: Approximating Truth-Conditions*

In cases in which we appeal to my proposed pragmatic reinterpretation, we use an interlocutor's individual approximating intension for 'water' to play a role in determining

which possibilities are ruled out by an assertion in which ‘water’ is a constituent expression. This is because her approximating intension for ‘water’ plays a role in determining what she takes to be the truth-conditions of that assertion. As we noted earlier, there is a very important sense in which Alice is not in a position to know the precise truth-conditions for Bruce’s assertion. While she understands that Bruce’s assertion is true just in case there is water in the basin, she doesn’t know precisely what the world would have to be like, under the relevant ambient conditions, in order for that to be the case. This is because there are relevant facts about water that she doesn’t yet know: she doesn’t know how water presents itself under cold ambient conditions. It is in this sense that she doesn’t know what the world must be like in order for what Bruce said to be true in that situation.

For this reason, Alice cannot use the truth-conditions of Bruce’s assertion to rule out the possibilities under consideration at which what Bruce says isn’t true. (Alice, of course, doesn’t know this—she thinks that she does know the truth-conditions for Bruce’s assertion because she thinks she does know how the world must be in order for Bruce’s assertion to be true.) Instead, just as she has an approximating intension for ‘water’ which serves as her approximation of the intension for ‘water’, she also has *approximating truth-conditions* for Bruce’s assertion that serve as her approximation of the truth-conditions for his assertion.

Just as the truth-conditions for an assertion are the conditions the world must meet in order for that assertion to be true, Alice’s approximating truth-conditions are the

conditions the world must meet in order for Alice to judge that the assertion is true. Her approximating truth-conditions for Bruce's assertion are determined by her approximating intensions for the constituent expressions. Since she understands that what Bruce asserted is true just in case there is water in the basin, and she believes that if there is water in the basin then the world will be such that there is a clear odorless liquid in the basin, then that there is a clear odorless liquid (that also has any functional characteristics that she associates with water) in the basin serves as her approximating truth-conditions for his assertion. Any possibilities at which there is not a clear odorless liquid in the basin are ruled out as possibilities at which what Bruce asserted is false.

Using this, we can construct Alice's approximating matrix for Bruce's assertion, represented in figure 1, below. As with the standard matrix, the worlds in the context set are represented in both the top row and leftmost column. However, instead of determining the value of each cell in the matrix by whether what is said at the relevant context of utterance is true at the relevant circumstance of evaluation world, we instead determine the value of each matrix by whether Alice's approximating truth-conditions for Bruce's assertion, as said at the relevant context of utterance, are met at the relevant circumstance of evaluation.

Figure 1

(Alice's Matrix)

	>32°F H <sub>2</sub> O (liquid)	<32°F H <sub>2</sub> O (solid)	<32°F pseudo- water	Mud
>32°F H <sub>2</sub> O (liquid)	<b>T</b>			
<32°F H <sub>2</sub> O (solid)		<b>T</b>		
<32°F pseudo- water			<b>T</b>	
Mud				<b>F</b>

As we can see, Alice's approximating truth-conditions for Bruce's assertion are not met at Alpha, but they are met at the below-freezing-pseudo-water world, and at the above-freezing-H<sub>2</sub>O world. This explains why Alice ruled out Alpha in light of Bruce's assertion.

### *Section 5: Constructing the Three-Dimensional Approximating Matrix*

In order to construct the three-dimensional approximating matrix that we will use to model the way that Bruce's assertion changes the state of the conversation, we need to generate two-dimensional approximating matrices for each interlocutor—that is, for each of Bruce's students—for that assertion. We will first simplify matters in section 5.1 by stipulating that Bruce's students are all in the same epistemic state with respect to both the nature of water (i.e., they believe that it is always presented as a clear odorless liquid) and the ambient temperature (they believe that the temperature is around 32 degrees F,

though perhaps it is just above or below it), and we will stipulate that Bruce has only three students. In section 5.2, we will discuss how to accommodate differences in the epistemic states of the interlocutors.

### *Section 5.1: A Simplified Three-Dimensional Matrix*

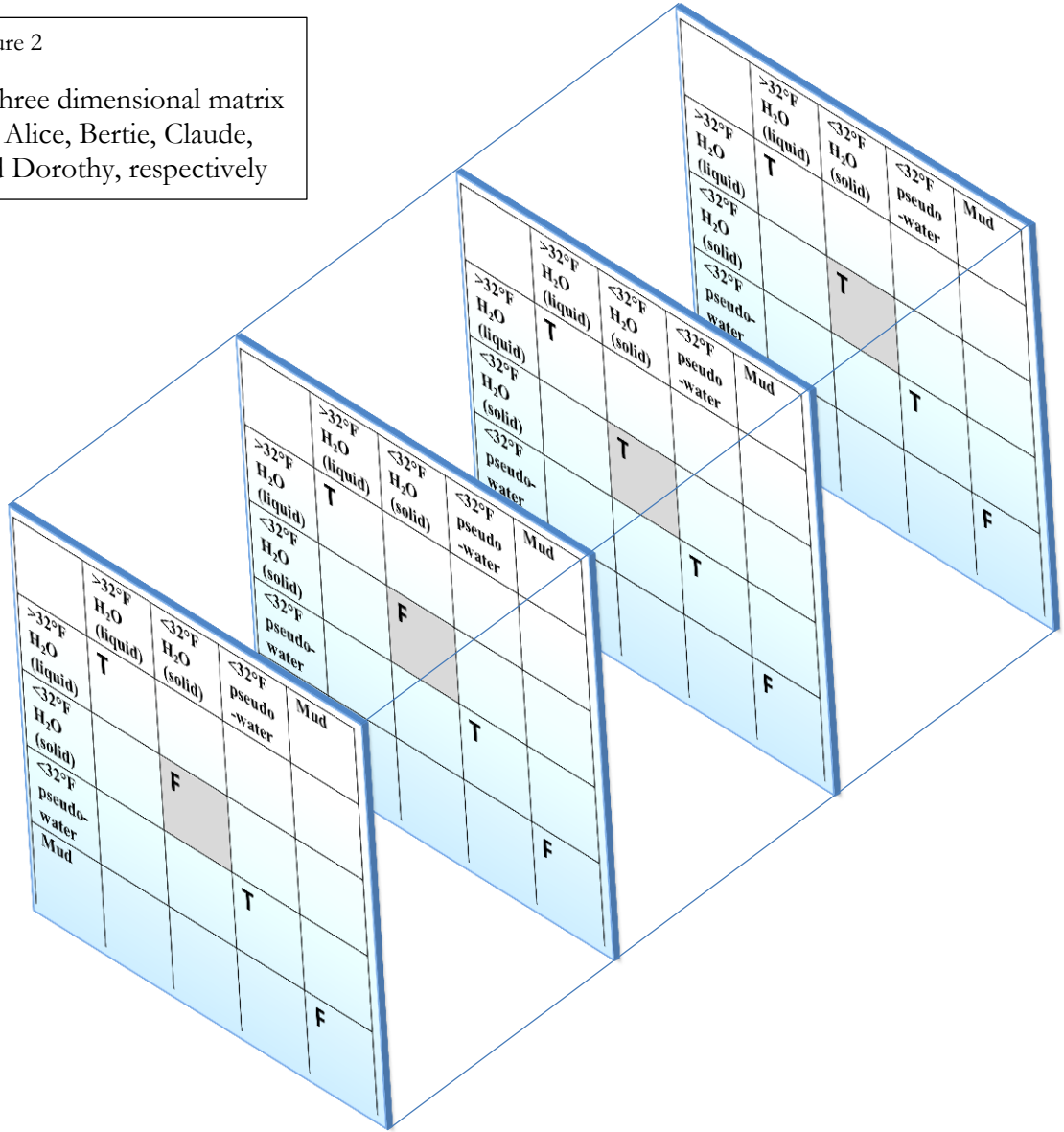
Consider, in addition to Alice, the two other students in Bruce's class—call them Bertie and Claude. For present purposes, as I noted above, we will be making the simplifying assumption that they are in precisely the same epistemic state as Alice at the time of Bruce's assertion. For this reason, the approximating two-dimensional matrix for Bruce's assertion is the same for each student.

To construct the three-dimensional matrix, we line up the two-dimensional matrices of each student (in no particular ordering of the matrices), such that the same cells line up for all of the matrices, as in figure 1, above. Since the students are in the same epistemic state with respect to water (and with respect to the ambient temperature), they are in agreement about whether each possibility is ruled out in light of Bruce's assertion.

To use the three-dimensional matrix to determine whether a world in the prior context set is ruled out by Bruce's assertion, we look at the *rank*—that is, the row that is constituted by the corresponding cell in each individual two-dimensional matrix—for the intersection of the relevant world column and world row. See figure 2, below.

Figure 2

A three dimensional matrix for Alice, Bertie, Claude, and Dorothy, respectively



Of interest to us is the rank for each world on the diagonal. This is what we will use to determine whether that world has been ruled out in light of the assertion (or, better, whether the conversation ought to proceed as though a given world has been ruled out in light of that assertion, for reasons to be discussed in the next section). The rank highlighted in figure 1 is the rank for Alpha on the diagonal. We can see that the rank for Alpha is such that, at each two-dimensional matrix, the interlocutor for that matrix has ruled out Alpha in light of Bruce's assertion. Thus, the matrix tells us that the state of the conversation has changed such that Alpha is not among the worlds in the posterior context set. Like Alpha, the mud world has been ruled out. We can see that both the below-freezing-pseudo-water world (at which the temperature is below 32 degrees F, and the substance in the basin is a clear odorless liquid) and the above-freezing-H<sub>2</sub>O (at which the temperature is above 32 degrees F, and the substance in the basin is a clear odorless liquid world have not been ruled out, and so both are in the posterior context set. All the worlds in the posterior context set are such that there is a clear odorless liquid in the basin.

More generally, to use the three-dimensional matrix for an assertion to determine the force of an assertion, we look at the rank for each world on the diagonal. Those worlds marked as false on the diagonal for all individual 2D matrices are ruled out by the assertion, and those worlds that are marked as true on the diagonal for all individual 2D matrices are in the posterior context set.



### *Section 5.2: A More Sophisticated Three-Dimensional Matrix*

It is quite plausible that often the interlocutors will *not* be in the same epistemic state in a given context. Whether the difference in epistemic state makes a difference to their individual matrices depends on whether the differences in epistemic state lead to different expectations of how the relevant sample(s) will be presented under the relevant conditions, namely those conditions—that is, the way the world must be—that are relevant to their approximating truth-conditions for the assertion. This kind of difference can also arise in contexts in which some individuals know how the individual or kind presents in the relevant conditions, but some do not—even though the former are able to use the truth-conditions, according to Stalnaker’s account, or the counterfactual truth-conditions, on Jackson’s and Chalmers’s accounts, the latter must rely on their approximating truth-conditions, and so we’ll need a way of modeling how the assertion changes the state of the conversation in those cases, too.

In some such cases, it may not make a difference to the question of whether a world is ruled out because the difference in what the interlocutors know isn’t relevant to their approximating truth-conditions. For example, Alice may expect that water presents as a steam under very hot conditions, and Bertie may expect water to remain a liquid under even those conditions, but these expectations aren’t relevant to how they think the world must be in order for Bruce’s assertion to be true under the conditions in which he made the assertion, given that the ambient temperature is cold.

It will make a difference if their different epistemic states lead to different beliefs about what the world must be like in order for Bruce's assertion to be true. That is, it will make a difference if they have different approximating truth-conditions for his assertion. For example, suppose that Claude, unlike Alice and Bertie, *was* paying attention when Bruce told the class that water has three states—a gas, a liquid, and a solid state—and that, under cold ambient temperatures, water presents as a clear solid. Claude knows that, if the ambient temperature is below freezing, there must be a clear odorless solid in the basin. Thus, we could appeal only to the standard interpretation in order to explain why Claude didn't take Alpha to be ruled out in light of Bruce's assertion. However, we are concerned with whether the *conversation* should proceed as though Alpha is ruled out, and so we need a way of modeling how each interlocutor took the assertion to change the state of the conversation, and to model this in a way that allows us to compare them. For this reason, we will use Claude's beliefs about how water presents under the relevant conditions to generate approximating truth-conditions for him, and so a two-dimensional approximating matrix for him. Claude's two-dimensional matrix will be such that, unlike

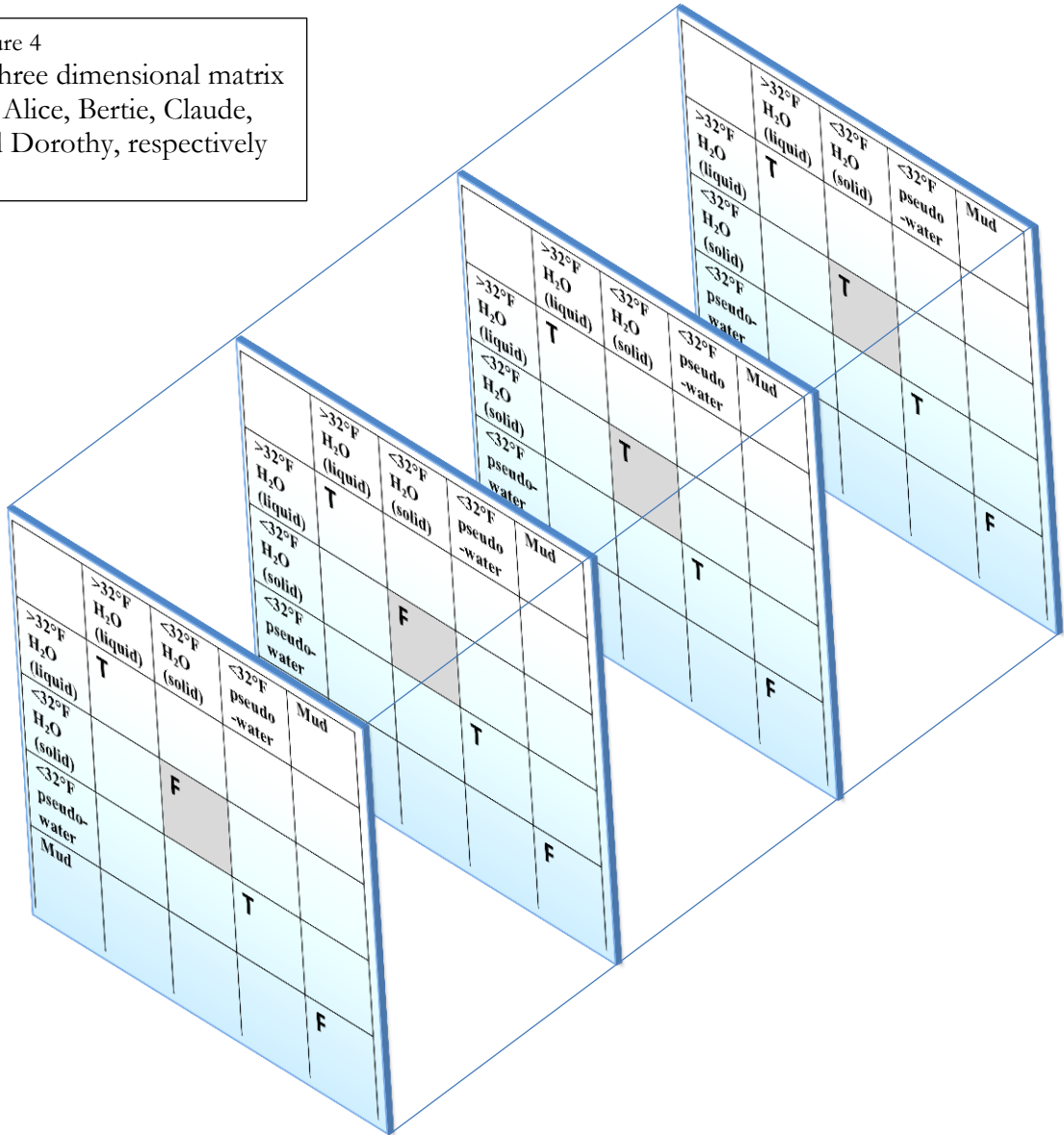
Alice's and Bertie's matrices, Alpha will not be ruled out on the diagonal. See figure 3, below.

Looking at the rank for Alpha on the diagonal in figure 4, below, we see that, on Alice's matrix, the cell is marked with an F, as well as on Bertie's matrix, but also that, on Claude's matrix, the cell is marked with a T. Since two of them ruled it out and one of them did not, we cannot say that it is ruled out on the grounds that all the interlocutors took his assertion to exclude that possibility. We need a way of determining how to update the state of the conversation in light of Bruce's assertion—is Alpha ruled out or not?

Figure 3  
(Claude's Matrix)

	>32°F H <sub>2</sub> O (liquid)	<32°F H <sub>2</sub> O (solid)	<32°F pseudo- -water	Mud
>32°F H <sub>2</sub> O (liquid)	<b>T</b>			
<32°F H <sub>2</sub> O (solid)		<b>T</b>		
<32°F pseudo- -water			<b>T</b>	
Mud				<b>F</b>

Figure 4  
 A three dimensional matrix for Alice, Bertie, Claude, and Dorothy, respectively



Here, I will discuss a general procedure for how to update the state of the conversation, and then in chapters five and six, I will discuss how to interpret the procedure as an extension of Stalnaker's account, and as an extension of Jackson's and Chalmers's accounts, respectively. As a first pass at working out a solution, one approach is to say that the world is ruled out if more interlocutors than not take it to be ruled out, and that the world isn't ruled out if more interlocutors than not take it to not be ruled out. This approach has some intuitive attractiveness: it says that we ought to go with the majority, which, if our aim is to help as many of our interlocutors as possible to get closer to determining the way the world in the relevant respect, then siding with the majority seems to help us to attain that aim.

Here's one kind of case that seems to support this way of thinking about it. If the number of interlocutors who mistakenly failed to rule out the possibility is greater than the number of those interlocutors who correctly did rule it out, it seems reasonable for either the speaker, or one of the other interlocutors who correctly ruled out the world, to express the same thing using a different assertion, if aware that at least some of their fellow interlocutors mistakenly failed to rule out the world in question. Making a second assertion with the same content would only be appropriate if the world in question had not been ruled out. For, otherwise, in making the second assertion, one is suggesting to one's interlocutors that at least one world that is not *in* the prior context set for your second assertion is nonetheless excluded by the second assertion.

For example, as part of an exercise to help reinforce their studies of the biological classifications of animals, suppose that Bruce is trying to help his students to guess his favorite animal. He gives them a hint: “my favorite animal belongs to the Class *Aves*”. Suppose that Bruce notices that though some of his students are nodding or saying “oh!” to indicate that they understand, the majority of his students look perplexed because they failed to study the biological classifications of animals, and so did not remember that the Class *Aves* is the class to which birds belong. Bruce says the same thing: “my favorite animal is a species of bird”.

Making this follow-up assertion is only appropriate if the prior context set for the follow-up assertion contains at least one world at which Bruce’s favorite animal is not a bird. Since at least some of his students had failed to rule out all possibilities at which his favorite animal is not a bird, these possibilities still need to be excluded. In making the follow-up assertion, these possibilities are successfully excluded. However, if the non-bird possibilities were not still in the context set after Bruce’s initial assertion, then his follow-up assertion would have been inappropriate: it would serve as a proposal to rule out possibilities that had already been excluded, and so were not currently under consideration. This would also leave unexplained why at least the majority of the students found the assertion informative. These considerations serve as motivation for modeling the state of the context set after the initial assertion, and just prior to the follow-up assertion, as containing at least one world at which Bruce’s favorite animal is not a bird.

However, if more interlocutors took a world to be ruled out than those who did not, this suggests that the interlocutors could help more of their fellow conversational participants learn more about the way the world is by moving on to the next assertion that will help them to learn more about the world, rather than spending that same time expressing the same content again. Furthermore, more specific to our immediate purposes, this gives us guidance for our present case: since Alice and Bertie ruled out Alpha, but Claude did not, then Alpha is not in the posterior context set. This gives us a means of explaining the appropriateness of Bruce's follow-up assertion that there might be ice in the basin since ice is water in its solid state, and it might be cold enough for water to be in its solid state.

However, this procedure won't give us guidance in cases in which there is an even number of interlocutors, and the same number of them take a world to be ruled out as those who take it that the world is not ruled out. For example, consider a fourth student—call her Dorothy—who, like Claude, was paying attention when Bruce informed the class that water presents as a clear solid under cold conditions, and so she also did not rule out Alpha in light of Bruce's assertion. According to Alice's and Bertie's matrices, Alpha is ruled out, and, according to Claude's and Dorothy's matrices, it is not. For this reason, our first attempt at a solution isn't as sophisticated as we need a solution to this problem to be.

I suggest adding weights to the matrices that serve to order them based on which interlocutors it is most important (relative to the conversation's purposes) to keep updated on the state of the conversation. My thought is that, in at least many cases, it is both the

case that we are aware that some interlocutors are understanding what we are saying better than others, and that it is more important (based on our projects) that some of the interlocutors understand than others.

To illustrate, consider a pedagogical example. In teaching, one is often aware that some students are following along quite well, others are following along to some extent, and still others are failing to follow along at all. One is also often aware that some students are putting in much more effort than others. Based on this information, after making an assertion that one has good reason to think that some students understood (and updated their beliefs appropriately), and other students did not understand (and so either didn't modify their beliefs at all, or modified them incorrectly), one can decide whether to make a second assertion with roughly the same content (in an effort to bring the latter students' matrices back in line with the matrices of the former students), or whether to continue without making the second assertion.

Whether one makes the second assertion often has to do with *which* students did, and did not, understand what one said. One might, for example, adjust one's discussion to give greater weight to whether the students who have been making an effort in the course understood than to those students who have not been making an effort. If the students who have not been making an effort in the course did not understand, but the students who have been studying, coming to office hours, etc., did understand, then even if there is an even number of both kinds of students—or, perhaps even a greater number of students who are not making the effort—the greater weight on how the students who are making



an effort means that one takes the state of the conversation to change in light of one's assertion in the way that those students take it to change.

For example, suppose, as before, that Alice and Bertie haven't been making an effort in Bruce's class, and that Claude and Dorothy have been making a significant effort. While Claude and Dorothy sit in the front of class, take notes, ask good questions, always turn in quality work on time, etc., Alice and Bertie always spend their days in the back of class, obviously not attending to Bruce's lecture, not knowing the answer when called upon in class, turning in poor quality work when they bother to submit an assignments at all, and so on. From this behavior, Bruce has good reason to think that Claude and Dorothy are making a significant effort in the class, and that Alice and Bertie are not.

Suppose that Bruce makes an assertion that Claude and Dorothy understand, but that appears to puzzle Alice and Bertie. He wonders whether he ought to say more to help clarify matters for the latter two, but concludes that it would be an unfair use of Claude and Dorothy's time. Bruce has assigned greater importance, or greater weight, to the way in which Claude and Dorothy take his assertions to change the state of the conversation than the way in which Alice and Bertie do. In term of assigning weights to the matrices, he puts a greater weight on Claude's and Dorothy's matrices than on Alice's and Bertie's matrices. The greater weight given to Claude's and Dorothy's matrices tips the scales in favor of taking the conversation to change in the way they think it does, rather than the way Alice and Bertie think it does.

In different contexts, different individuals may assign the weights differently, based on how important it is to keep all the interlocutors up to speed, so to speak. Consider two instructors. Suppose that one is told by the principle that what is most important is to advance the class over the material as far as possible over the course of the semester, and so will assign a much higher weight to those students making an effort. Suppose that the other instructor, at a different school, is told by the principle that what is most important is that all the students learn whatever material is covered, even if that means moving at a slower pace and, at least some of the time, being redundant (at least by the more attentive students' lights). The latter instructor thereby ought to assign a higher weight to the students who *didn't* understand, such that she ought to make a follow-up assertion to rule out the possibility that wasn't ruled out by the previous assertion. These differences in pedagogical philosophies lead to different assignments of weights to the matrices.

This is an important way in which this 3D reinterpretation belongs to the pragmatics. It isn't a matter of *whether a given world is, in fact, ruled out by an assertion*, but rather whether, given certain practical considerations, including aims of the project in which one engages in participating in the conversation, the practical considerations would be best served by the conversation proceeding as though a given world had been ruled out or not.

If one adopts credence values in the individual approximating matrices, this can also play a role in whether the conversation ought to take a world to be ruled out in light of an assertion. To see this, consider a three-dimensional matrix constructed from two-

dimensional individual approximating matrices in which, on the rank for a particular world on the diagonal, each interlocutor did judge that the world in question satisfied their approximating truth-conditions, but only with a credence of, say, .6. On the one hand, they *did* take the assertion to rule out the world, but on the other hand, they did so with a fairly low level of confidence. For this reason, it seems that, practically speaking, if anything else one says turns on their acceptance of the assertion in question, it would be advisable to express the same content with a follow-up assertion, or to say something else that might otherwise make the assertion in question more plausible.

Because we are also assigning weights to the importance of each interlocutor's approximating matrix, simply taking an average of the credence values in that rank alone does not determine whether one should take it that the world is ruled out from the context set. It is a calculation that is done first for each individual matrix by multiplying the score of the importance of that matrix by the credence value (assigned as a number *between zero*<sup>70</sup> and one) for the cell, and then adding up the score for each matrix, dividing that number by the number of matrices, and then dividing that number by 10. If that number is .5 or greater, then you should proceed as though the world has not been ruled out, and, otherwise, you should proceed as though the world has been ruled out.

---

<sup>70</sup> Allowing for a value of zero would cause calculations in which a zero is assigned as a credence value, the overall calculation would be zero, even if the other matrices had a higher weight, and high credence values.

***Section 6: Evaluating the General Three-Dimensional Reinterpretation as Proposed as an Extension to the Two-Dimensional Approaches***

*Section 6.1: What Can Approximating Intensions Do?*

Approximating intensions can be used to capture the presuppositions being made about that which is in the extension of the relevant proper name or natural kind terms in conversations in which at least some interlocutors are not in a position to judge whether the relevant samples belong to the kind in question, such as the conversation in which Bruce makes his assertion. Consider the conversation that is being had by his students prior to Bruce entering the conversation and making the assertion. Since his students (perhaps with the exception of Claude and Dorothy) weren't paying attention when he told them that ice is water in its frozen state, they believe that water is always presented as a clear odorless liquid. Furthermore, each student (again, perhaps with the exception of Claude and Dorothy, depending on whether they noticed that their classmates weren't paying attention) can also reasonably presuppose that their classmates think so, too. To simplify, we'll ignore Claude and Dorothy for present purposes, and reintroduce them again later. Thus, each student is in a position to presuppose that, when someone uses the word 'water', that person is talking about a clear odorless liquid that also has the functional characteristics that the student associates with water.

This idea is already familiar from our discussion in chapter one: in having this sort of conversation, among the facts that the interlocutors are trying to learn about the world are facts about what is in the extension of a given proper name or natural kind term. Unlike a case in which an individual knows how the individual or kind presents in the relevant

conditions, and so can reliably pick it out under those conditions, when the individual does not know how the individual or kind presents under the relevant conditions, part of what the interlocutor learns about the world while engaging in these sorts of conversations is which things *are* in the extension of that term, and how that which is in the extension of that expression is presented under those conditions.

Applying this to our water example, among the presuppositions of the conversation may well be a presupposition about what ‘water’ picks out, where the conditions for being picked out by ‘water’ are taken to be the student’s conditions for playing the water-role. If, as we’ve been discussing, the students have (if only tacitly) the uniform extension intuition, then they take ‘water’ to apply to all and only samples of water. However, their judgments about whether a given sample *is* water aren’t always correct. Therefore, it isn’t appropriate to model the presupposition about the extension of ‘water’ by picking out all and only samples of that natural kind that happens to play the water role at that world, including those samples of the kind that don’t meet the conditions for playing the water-role at that time. The unifying feature—that which makes it the case that the kind term refers to all and only samples of the kind, regardless of how they are presented—allows for samples to be in the extension of ‘water’ that the students do not believe to be in the extension of ‘water’.

### *Section 6.2: Ruling out the actual world*

A very tempting challenge to this approach is to argue that this is unnecessary, on the grounds that Alpha isn’t among the worlds in the context set for the students. This is a

variation on the challenge that we discussed in chapter two. If the worlds in the context set are the ways that this world could still turn out to be that are consistent with the student's beliefs about the nature of water, it seems very plausible to think that no worlds in the context set are a way that the students believe that the world could not turn out, given what they know (or at least what they believe). Since the students believe that the world could not turn out to be such that liquid water and ice could turn out to be of the same natural kind, this seems to give us good reason for thinking that there must not be any world in the context at which it turns out that that which plays the water-role is of the same natural kind (shares its underlying constitution with) that which plays the ice-role. Unbeknownst to the students,  $\alpha$ , that which plays the water-role is of the same kind as that which plays the ice-role. This suggests that  $\alpha$ , then, is not among the worlds in the prior context set for Bruce's assertion.

I think that there are reasons to resist this admittedly very attractive response, however. What possibilities do the students take themselves to be excluding? Among the worlds they take to not be the way that Bruce said the world is are any worlds at which there is ice in the basin, and so any worlds at which there is a clear odorless solid in the basin, which they take to be the way in which ice is presented. That is, any worlds at which a sample of that kind which, at that world, plays the ice-role is in the basin are ruled out by the students in light of Bruce's assertion.

This is a rather coarse-grained way of ruling out worlds in light of assertions: all worlds presented in this way are ruled out. It doesn't matter whether, at that world, that which is

so presented happens to have the same underlying composition as that which plays the water-role. What matters for the purposes of ruling out worlds is the students' beliefs about that which is presented in that way, and not the underlying nature of that which is so presented. My reason for saying this is that they aren't sensitive to the underlying details—they don't know them, and might not even be thinking about them as being that which differentiates kinds.

This suggests to me that the constraint that the worlds in the context set must each be a way the world could turn out to be, for all the interlocutors know, is in need of further explication. I think that, for the reasons we've been discussing, there is reason to not exclude Alpha from the prior context set even though, at Alpha, that which plays the water-role is of the same kind as that which plays the ice-role, and that the students believe that which plays the ice-role is distinct in kind from that which plays the water-role. From the fact that the students believe that water and ice are of distinct kinds, it does not follow that the way to represent this belief is to only allow in the context set those worlds at which that substance which plays the water-role is chemically distinct from that substance which plays the ice-role. This is because even though what makes it the case that two samples of substances are samples of distinct kinds is that they differ in their underlying constitution, this may not be the grounds on which the students take these two substances to be distinct. Perhaps they hadn't thought at all about the underlying constitution, or the notion of things having underlying constitutions has not been introduced to them.

Not uncommonly, we categorize things by appearance, and a significant difference in appearance provides a default reason for thinking that two things are of distinct kinds. Note that I am not making a claim that this is the *only* reason that we categorize things as members or samples of distinct kinds—we might do so instead *if* we have reason to think that they differ in their underlying composition. Furthermore, information regarding the underlying composition overrides the default reasons to think that two samples are of the same kind, or distinct kinds, based on appearances.

For example, though the difference in appearance of ice and water provides a default reason for thinking that they are members of distinct natural kinds, discovering that they have a common underlying constitution overrides that default reason, and gives us reason to think that they are not of distinct kinds. Also, though the very strong similarity of appearance of jadeite and nephrite provide a default reason for thinking that they are of the same kind, information about the differences in their underlying composition serves as an overriding reason to think that they are of distinct kinds.

What is particularly important for our purposes here is how the students think about (what they take to be) the difference in kind between that which plays the water role and that which plays the ice role. Their thoughts about what makes two samples of substances belong to distinct kinds may not involve thinking at all about the underlying constitution of these substances. Rather, it may just be a matter of expecting that the sample of the substance that plays the ice-role (at least while playing the ice-role) is presented in a way that is different from the way that the sample of substance plays the



water-role (at least while playing the water-role) is presented. This is satisfied by Alpha. While playing the ice role, H<sub>2</sub>O is presented differently than it is presented while playing the water role.

This gives us a much weaker constraint on which worlds are in the prior context set than that the substance that plays the water role is of a distinct kind than the substance that plays the ice role. This weaker constraint is that the worlds in the prior context set are those that are presented in the ways that the interlocutors expect, where these expectations are governed by the roles that they associate with the relevant proper names or natural kind terms. This weaker constraint allows for the actual world to be in the prior context set, unlike the original constraint.

Having introduced the three-dimensional framework in a general way, we are now ready to discuss the proposals on which it serves as an extension to Stalnaker's, Jackson's, and Chalmers's accounts. We will discuss my proposed extension to Stalnaker's account in the next chapter. In the following chapter, we will discuss my proposed extensions to Jackson's and Chalmers's accounts.

## **Chapter Five: Three-Dimensional Approach as Extension to Stalnaker's Account**

In this chapter, I discuss an interpretation of the three-dimensional framework that I propose as an extension to Stalnaker's two-dimensional metasemantic view for the purposes of serving as a solution to the problem of informativity. Since it only makes use of resources to which Stalnaker is already committed, accepting it would not constitute a significant change to his view. Furthermore, I propose that it is understood as a solution to the problem of informativity that resembles his metalinguistic approach to diagonalization.

### ***Section 1: Pragmatic Reinterpretation***

As we discussed in chapter two, on Stalnaker's view, he thinks that direct reference theory gives the right account of the semantics of natural kind terms, in order to explain why some assertions, such as the necessary a posteriori, are informative, we need to engage in a kind of pragmatic reinterpretation that he calls diagonalization.<sup>71</sup> We will only revisit the metalinguistic approach to diagonalization here because, as we've discussed, it avoids worries that can be raised for the fixed-metasemantic approach, and doesn't appear to suffer from additional worries to which the fixed-metasemantic approach is not subject.

Remember that, on the fixed metasemantic approach to diagonalization, we use what the interlocutors know about the nature of that which is in the extension of the expression, or

---

<sup>71</sup> See chapter two for our earlier discussion of this.

the ways in which the individuals are uncertain about the nature of that which is in the extension of the expression, in order to engage in a kind of pragmatic reinterpretation. As we saw in the case of Bruce's assertion, metalinguistic diagonalization makes use of what the students believe about water—namely, that it is a clear odorless liquid—and uses either the description, 'clear odorless liquid', or the property, *being a clear odorless liquid*, as the metasemantic facts, together with the unifying feature (since they understand that natural kind terms refer to all and only samples of the same kind), for the reinterpretation of 'water' as part of the reinterpretation of Bruce's assertion at each world in the context set.

We encountered difficulty in explaining why the students ruled out Alpha in light of Bruce's assertion because, both on the standard interpretation, and on metalinguistic diagonalization, natural kind terms refer uniformly to all and only samples of that kind. This is so because among the metasemantic facts for natural kind terms used for reinterpretation of those terms—provided that the interlocutors believe that the natural kind term *is* a natural kind term, and that it refers to all and only samples of that kind—is that the natural kind term refers to all and only samples of the same kind as that which is picked out by the reference-fixer for that term, which we have been calling the unifying feature. The unifying feature of the metasemantics, on both the standard interpretation, and for the reinterpretation of 'water', ensures that, even on metalinguistic diagonalization, what is said at Alpha is true at Alpha.

In this chapter, we will discuss how to apply the uniform extension intuition in the way that yields a pragmatic reinterpretation that is very much like metalinguistic diagonalization, save for the unifying feature being absent. This has the benefit, for the purposes of the pragmatic reinterpretation, of allowing ‘water’ to pick out all and only those samples of substances that are presented in the way that his students expect for water to be presented, rather than to those samples *and* all other samples that share the same underlying constitution.

### ***Section 2: Pragmatic Reinterpretation and the Three-Dimensional Matrix***

In this section, we will discuss how to interpret the three-dimensional matrix as an extension of diagonalization within Stalnaker’s metasemantic approach. Just as diagonalization is a means of pragmatic reinterpretation that can be used to account for the fact that necessary a posteriori sentences having as constituents proper names or natural kind terms are informative within an account that accepts direct reference as giving the correct semantics for proper names and natural kind terms, we can also use the three-dimensional matrix approach as a kind of pragmatic reinterpretation that serves as a means of accounting for the way that assertions change the state of conversations that would not be explained by appeal to either the standard interpretation, or to pragmatic reinterpretation.

I propose the following procedure for accounting for how assertions change the state of the conversation, as an extension of diagonalization on Stalnaker’s account. As discussed

in chapter two, if we cannot explain the way that an assertion changes the state of the conversation by appealing to the standard interpretation, we appeal to metalinguistic diagonalization. I propose that, if we cannot explain the way in which the assertion changes the state of the conversation by appealing to either the standard interpretation or metalinguistic diagonalization, we appeal to the three-dimensional matrix. As I noted in chapter four, the cases in which we'll need to appeal to the three-dimensional matrix are those cases in which at least some of the interlocutors do not know how the individual or natural kind presents under the relevant conditions, such as the fact that Bruce's students do not know how water presents itself in the conditions relevant to the truth of Bruce's assertion, 'There is water in the basin'.

The procedure for constructing the two-dimensional matrices for each student is the same as in chapter four, as is the procedure for combining them to form the three-dimensional matrix. It is how we *interpret* the three-dimensional matrix, so constructed, on Stalnaker's account that differentiates this three-dimensional approach from the one that I propose as an extension of Jackson's and Chalmers's views in the next chapter. On my proposal for the extension to Stalnaker's view, the approximating intensions for proper names and natural kind terms, and assertions in which they are constituents, are part of the pragmatic reinterpretation, and so are *not* part of the semantics for these expressions. Just as the reference-fixer that we use for diagonalization is not part of the semantics, and is only used for the purposes of that instance of pragmatic reinterpretation, the individual approximating intensions for that term are not part of the semantics, and are only used for the purposes of that instance of pragmatic reinterpretation. Furthermore, the three-

dimensional extension itself is also only part of the pragmatic reinterpretation. These are all resources to which we may appeal in cases in which the standard interpretation (i.e., employing the standard one-dimensional matrix, namely the horizontal) and, failing that, the metalinguistic reinterpretation (i.e., employing the standard two-dimensional matrix), cannot explain the way in which the assertion changes the state of the conversation.

*Section 2.1: Metalinguistic Diagonalization vs. Three-Dimensional Matrix*

Let's compare metalinguistic diagonalization to the three-dimensional approach that I am proposing, to see why the latter, but not the former, can offer an explanation of why the students ruled out Alpha. Recall that, unlike on the three-dimensional approach, on metalinguistic diagonalization, the reinterpretation of 'water' included the unifying feature. On the metalinguistic diagonalization approach, we are to use the students' (perhaps only tacit) beliefs both that 'water' is a natural kind term, and so picks out all and only samples of the same kind, and about the way that water presents under the relevant conditions, for the purposes of reinterpreting Bruce's assertion at each world in the context set. Since the students believe that 'water' is a natural kind term, and take natural kind terms to refer uniformly to all and only samples of that kind, then on metalinguistic diagonalization, making use of the property or description that they associate with 'water', 'water' picks out not only that which is presented as having the qualitative and functional characteristics that they associate with water, such as falling as rain, coming out of taps, etc.), but also all and only samples of the same kind, including

those samples that aren't so presented.<sup>72</sup> This means that if the sample of substance in the basin is of the same kind as the samples that satisfy the relevant description (e.g., 'the predominant clear odorless liquid'), or have the relevant property (e.g., being the predominant clear odorless liquid), used for the purposes of reinterpretation, then, on reinterpretation, what Bruce said at that world is true at that world.

Constrast this with how the uniform extension intuition is employed on what I will call the three-dimensional pragmatic reinterpretation approach (3DPR). First, in order to reinterpret what is said at each world within a given student's two-dimensional matrix, we use that student's approximating intension to reinterpret the extension of 'water' at each world in the context set. We then use this to reinterpret what Bruce said at each world in the context set by determining that individual's approximating truth-conditions for his assertion. Lastly, we use the individual's truth-conditions to evaluate whether what is said by that assertion, on reinterpretation at that world, is true at that world by determining whether the truth-conditions are satisfied at that world.

Since the approximating intension does *not* have the unifying feature, it only picks out those samples that are presented the way that the individual expects samples of that kind to be presented under those conditions, rather than *all* samples of that same kind. This means that, on reinterpretation, what is said is true only at those worlds at which the substance in the basin is presented in the way that the students expect water to be presented (namely, those worlds at which the sample in the basin has the qualitative and functional characteristics that they associate with 'water'). From this, we can see that,

---

<sup>72</sup> For our earlier detailed discussion of this, see chapter two.

reinterpreted at Alpha, what Bruce said is false at Alpha, since the contents of the basin fail to have the qualitative characteristics that the students associate with ‘water’, and so this explains why the student in question—say, Alice—took his assertion to rule out Alpha.

So far, this only explains why Alice ruled out Alpha in light of what he said. In order to explain why *all* the students—here, we’ll exclude Claude and Dorothy—took what he said to rule out Alpha, we’ll need to construct a two-dimensional approximating matrix for each student, and then combine them into a three-dimensional matrix. Once we have our three-dimensional matrix, we look at the rank generated by the intersection of the Alpha row and Alpha column in each student’s matrix. Since they all expect water to only present as a clear odorless liquid with the functional characteristics that they associate with water, their approximating intensions for ‘water’ will not have ice in the extension, and so their approximating truth-conditions for Bruce’s assertion are such that, on the 3DPR approach, what he said is false at Alpha. For this reason, in each of their matrices, the cell that is the intersection of the Alpha row and the Alpha column is marked as false. That all cells in this rank are marked as false explains why the students ruled out Alpha in light of his assertion.

However, as with diagonalization, the 3DPR approach is merely a matter of pragmatic reinterpretation. Neither the approximating intensions, nor the approximating counterfactual truth-conditions, are part of the semantics. The strong similarity between the metasemantic facts used for the purposes of diagonalization, and the approximating



intensions, and between the reinterpretation of the assertion at each world, and the counterfactual truth-conditions of the assertion, serve as two reasons that it is plausible to think of 3DPR as an *extension* of diagonalization.

Another similarity between diagonalization and 3DPR is that they are both approaches of a strongly pragmatic nature. To see this, recall that Stalnaker says that a world is ruled out by virtue of that world not being the way the assertion says the world is, where the way the assertion says the world is is to be understood as the assertion's content, i.e., what the assertion says on standard interpretation. However, in cases in which we cannot appeal to standard interpretation to explain the way an assertion changes the state of the conversation, we instead appeal to a kind of pragmatic reinterpretation. When we cannot explain the way the assertion changes the state of the conversation by appealing to either the standard interpretation or diagonalization, we appeal to 3DPR. These two approaches—diagonalization and 3DPR—offer means to explain the way the assertion changes the state of the conversation when appealing to the standard interpretation cannot account for it, thereby allowing Stalnaker to be able to account for the informativeness of an assertion that would otherwise go unexplained.

A further reason to think that 3DPR is an extension of diagonalization is that whether or not a world is ruled out by an assertion is determined by what is going on in the diagonal. Even though, on 3DPR, we look at the rank for that world on the diagonal in order to determine whether a given world is ruled out (or, at least, whether we ought to proceed as though it has been ruled out—see my discussion of this in section 2.2 of this chapter),

rather than merely the intersection of that world's column with that world's row, the rank for the each world on the diagonal intersection plays very much the same role as the cell for the intersection of that world's column with that world's row: just as, on diagonalization, the world is ruled out if the assertion is false on the diagonal, the world is ruled out, on 3DPR, the conversation is to proceed as though the world is ruled out if it is ruled out on the rank for that cell, at least in cases in which the interlocutors are in agreement about whether the world is excluded by the assertion. In the next section, we will discuss how to determine whether the conversation should proceed as though a given world has been ruled out in light of an assertion, in cases in which some but not all of the interlocutors take the world to be ruled out.

### *Section 2.2: What to Do in Cases in Which Some But Not All Rule Out a World*

As in the case in which the interlocutors are in agreement about whether the world is ruled out, we look at the rank of the three-dimensional matrix on the diagonal for that world. The corresponding cell in each interlocutor's two-dimensional matrix tells us whether that interlocutor ruled out the world in light of the assertion in question. If all the cells in that world's rank on the diagonal are marked as false then it is clear that the conversation should proceed as though the world has been ruled out, and if all the cells in that world's row on the diagonal are true then it is clear that the conversation should proceed as though the world has not been taken to be ruled out. Sometimes, however, some cells in the row will be marked as true and others as false, telling us that some but not all of the interlocutors took the assertion to rule out that world. In such a case, should

the interlocutors proceed as though the world has been ruled out such that it is not in the posterior context set?

Since some interlocutors take the world to be ruled out, and others do not, we cannot appeal to the agreement across interlocutors to determine whether the world is ruled out. As we saw in chapter four, we should not appeal to whether the majority of interlocutors took it to be ruled out, either, for at least two reasons. First, it is not always the case that there *is* a majority. Sometimes, there are an even number of interlocutors, half of which took the world to be ruled out, and the other half that did not. Secondly, as we discussed in chapter five, whether some interlocutors ruled it out is more important than whether others did, and so, even in cases in which there is a majority, that may not be the most important factor.

Here, we will discuss how to interpret the general way that we addressed this question in chapter four within Stalnaker's account. Recall that, in these cases, we assigned weights to the students' matrices, such that the fact that a world was, or was not, ruled out on one matrix carried more weight than if that world was, or was not, ruled out on another matrix. How we go about assigning those weights will depend on various factors, which may include our projects, in what relation we stand to our fellow interlocutors, how strongly we hold others accountable for what we think they ought to know, how important it is that various interlocutors understand what is said, and so on.

As discussed in section 5 of chapter four, we also have the option of assigning credence values to whether a world has been ruled out. That is, rather than merely saying whether the approximating truth-conditions for an individual have been satisfied by a given world, and, so, whether that world is ruled out according to that individual's matrix, we can use that interlocutor's credence regarding whether the world in question satisfies their approximating truth-conditions for that assertion. The credence for whether the world has been excluded by the assertion, along with the weights for that matrix, determines a value for that individual two-dimensional matrix that can then be used, together with the values for the other individual two-dimensional matrices, to determine whether the world in question is ruled out.

### *Section 2.3: Weights and Credence Values, on Stalnaker's Account*

Given that this is intended to be an extension of Stalnaker's account, how should we think about the assignment of weights to the matrices and credences? Importantly, what reason(s) do we have to think that the interlocutors will be able to come to the same conclusion that a given world is, or is not, ruled out? One might reasonably be concerned about this given that the interlocutors may assign different weights to the individual matrices, or different credences for the individuals. In such a case, it seems that we'd need a 3D matrix for each interlocutor to explain why each interlocutor took the world to be ruled out, according to her fellow interlocutors, such that she should proceed as though it has (or has not, as the case may be) been ruled out. But then it seems that we'd need a way of aggregating the results from each interlocutor's 3D matrix, but in such a way that

it could explain how the interlocutors generally seem to be in agreement about which worlds were ruled out in light of the assertion (as evident from the fact that they tend to make informative subsequent assertions). One can see how this would lead to a regress.

However, I think that there are some social norms that can be used to make a case that interlocutors will assign similar weights to the same interlocutor's matrix, for each matrix. Consider the social norm to not allow someone who doesn't know what they are talking about not derail a conversation. Imagine a conversation during which it becomes apparent that one of the interlocutors not only is significantly less informed about the topic than the others, but is also repeatedly saying things that the others have good reason to be false, or at least to take to be quite dubious. They will ignore that person's assertions, and will no longer be concerned with whether he understands further assertions.<sup>73</sup> For example, consider a group of students who have formed a study group for an exam, and are sharing what they know about the topics on the review sheet. If one of the students regularly says things that are false, and seems to lag behind the others in a significant way in his understanding of the material, it seems reasonable to think that these students will no longer be concerned with whether he is ruling out possibilities correctly. Time spent helping him to catch up is time that isn't spent making further progress for the others in preparing for the exam. In such a case, it seems that his matrix has been given a very low weight.

---

<sup>73</sup> Unless they have further practical reasons for being concerned with whether he keeps up with the rest of the group. For example, suppose that the conversation takes place during a meeting for a group project for which it is a requirement that all members very well understand the relevant material, or at a professional conference. In such a case, there is a strong practical reason for the others to not discount whether he takes an assertion to rule out a possibility or not.

Another such example might be in a debate about, say, policy. If it becomes apparent that one of the interlocutors hasn't been keeping up with current affairs with respect to that policy, this seems like a case in which the interlocutors would carry on, rather than trying to help him to catch up on the issues.<sup>74</sup> In such a case, a much higher weight is assigned to the matrices of the other interlocutors than to the one who is uninformed on the issues.

Another norm might be to assign a higher weight to the matrices of those individuals to which one either has independent reason(s) for taking them to be well-informed on the topic, or if they have said things earlier in the conversation which has given one (very recent) reason to think that they are informed. This is a practical consideration: if the aim of the conversation is a mutual interest in learning more about how the world is in a particular respect, and an interlocutor seems particularly well-suited to help in achieving that aim based on what she knows, and if making assertions that are informative to her will keep her involved in the conversation for a longer duration (because it satisfies her interest in learning about the topic, too), then it seems advisable to assign higher weight to her matrix.

How would we apply this in the case of Bruce's assertion? His situation is a familiar one, especially to one with teaching experience: One says something in conversation with one's students, and must assess which students, if any, understood what one said. The situation I have in mind arises when some but not all of them understood what one said.

---

<sup>74</sup> Again, unless there is some important reason to help him to catch up. Perhaps they are waiting in line to vote, and want him to make an informed decision, or he is their friend, and so they want to help him to better understand, etc.

Furthermore, there is a social pressure among the students to not seem to know less than the others, and so a pressure to accurately anticipate what the others know, such that each student has an interest in accurately assessing the state of the conversation. This allows them to avoid, for example, asking a question to which the others already know the answer.

In such a situation, one might assign different weight, or importance, to whether a given student understood what one said, based on, for example, how much of an effort one thinks that the student is making in class (e.g., whether the student attends lecture, seems to do the reading, is an active participant, performs well, etc.). The more of an effort the student makes, the greater the importance one might assign to the student understanding what is said. How little weight one assigns to students who aren't making an effort, or are making a very poor effort may depend on how far their knowledge lags behind the others (and so how much it might slow down the class to keep repeating material that the others have already learned through doing the readings, listening in lecture, and so on). Notice that the relevant notion here is whether or not the student is to be held accountable for not understanding what was said. If a student, despite making a significant effort in the course, is still struggling to learn the material, one might want to assign a high weight to that student's matrix—unlike the student who isn't doing the reading, and isn't paying attention in lecture, she is making a significant effort to learn the material.

As I noted earlier, this decision is based on matters such as our present projects, how important it is that all interlocutors are “on the same page” with respect to the

information being conveyed in the conversation, how accountable the interlocutors ought to be held for knowing the relevant information, and so on. These matters give us guidance on how to assign the relative weights of the individual matrices. Let's turn our attention to each.

Our present projects make a difference to the way in which one ought to assign the weights. Since Bruce is the instructor, and they are on the present hike during their class field trip to learn about the natural environment in Alaska, among his projects at the time of his assertion is to correctly inform them about the surrounding area (including about the nature of the ice). This suggests that he should assign greater weight to the matrices of those students who took his assertion to rule out Alpha than to the matrices of those students who did not.

However, suppose that, instead, the purpose of the current hike is to take a nice class picture in the wilderness of Alaska. In order to look their best, everyone is dressed in their nicest clothes and shoes. For this reason, it is imperative to avoid water so that their shoes aren't ruined. When Bruce asserts, 'there is water in the basin', it doesn't really matter whether they mistakenly rule out the possibility that there is ice in the basin. His purpose in making the assertion is to let them know to steer clear of the basin while wearing their nicest clothes. In this case, it seems that he shouldn't assign greater weight to the matrices of the students who ruled out Alpha in light of his assertion because any follow-up assertion would not better serve the present purpose. More generally, one



should assign greater weight to the matrices of those interlocutors who didn't understand when further clarification through a follow-up assertion will be useful to one's project.

Another consideration is how important it is that all the interlocutors are on the same page. Suppose that the students are considering whether the basin might be a safe place to deposit some caesium. Being an alkaline metal, caesium reacts violently with water, causing an explosion. Suppose that the students understand that it is imperative not to put caesium in water, but they don't understand that ice is water in its frozen state. This isn't a situation in which it is important that they understand that the water in the basin is in its frozen state—since they know that there is water in the basin, they won't deposit the caesium.

However, suppose that some of the students have heard about the activity of winter swimming—in which one swims in *very* cold water—and get the idea to do just that. Suppose that, as we know, from their present view, they cannot see the contents of the basin. However, they can see that the basin is very large, and so they suspect that it must contain a good deal of water. After making his assertion, 'there is water in the basin', Bruce hears some of them formulating the idea of running and diving into the basin for a winter swim. This is a situation in which it is important to give higher weight to those who mistakenly ruled out Alpha: if they run and jump onto the ice in the basin, they might likely get very hurt. For this reason, it seems very appropriate for Bruce to make the follow-up assertion.

Differences in pedagogical strategy will also make a difference to how the weights ought to be assigned. If the strategy is to not proceed until all of the students are clear on the material, then one will want to assign a much higher weight to the matrices of those students who do not understand than to those who do. This means that one will proceed as though the relevant world has not been ruled out (though the assertion says the world isn't that way) or that it has been ruled out (though the assertion says the world is that way). However, if the course structure requires covering as much material as quickly as possible, and puts a heavy requirement on each student to keep up with the reading, then one will want to assign higher weight to those students who understand, or at least those students along with the ones who are making the required effort.

These are just some examples of social norms to which we are sensitive. That there are generally adopted social norms that govern these matters suggests that their fairly widespread adoption ensures that, while the interlocutors might not assign precisely the same weight to each individual matrix, they will likely assign similar relative weights to the matrices.

Similarly, though I'm not prepared to tell a story here about *how* we do it, we do seem to often be in agreement about how well others did, or did not, understand an assertion. While we may not assign the *same* credence to a particular interlocutor, we seem to be able to assign similar *relative* credences. By this, I mean that though one interlocutor may judge that fellow interlocutor A has credence .6 for the assertion, and that fellow

interlocutor B has credence .8, and that another interlocutor may judge that A has credence .7 and B has credence 1, both judge that B has a higher credence than A.

I think that these considerations should increase our confidence that we won't run into the calculation problem that raised at the beginning of this section. In the next section, I motivate the idea that Stalnaker would want to adopt this proposed extension of his account.

### ***Section 3: Why Adopt the Extension?***

The aim of this chapter is to motivate the importance of being able to model communicative exchanges between individuals who don't know how the individual or kind presents itself under the relevant conditions. This becomes a particularly pressing problem when one considers, as we did in chapter four, that having such conversations, such as in the context of discussing groundbreaking scientific research, is often the means *of coming to know how the individual or kind presents under the relevant conditions*. In the course of such a conversation, one works out facts about the way our world is, *including* facts about the way that which is in the extension of our expressions is presented under various conditions.

We can think of modeling communicative exchanges in the following way. Knowing the state of the conversation—that is, knowing which (relevant) possibilities are still open to the interlocutors—tells us what would, and would not, be informative to assert in that

state of the conversation. Knowing what would be informative helps us to achieve our project of getting closer to figuring out the way the world is, since, as we've discussed, informative assertions rule out possibilities that aren't the way the world is (at least if the assertion is true), and introduce possibilities that ought to be considered (if the assertion is true) but weren't under consideration prior to the assertion.

Because this is so central to our project of getting closer to determining the way the world is, if there is a way to have such guidance even in the cases which require resources that outstrip the means of Stalnaker's present account, such an account seems worth having. As interlocutors in such a conversation, we clearly do have the capacity to work out what is informative in these cases, and we do so in a way that seems (to us) to be ruling out those possibilities that aren't the way the assertion says the world is. I point this out because these two features strongly suggest that, in important respects, what we are doing in such conversations is continuous with what we are doing in contexts in which we do know how the individual or kind presents under the relevant conditions.

Furthermore, in adopting such an approach, it allows us to model *conversations*, rather than isolated assertions, in cases in which the interlocutors don't know, for at least some assertions, how the individual or kind must present under the relevant conditions in order for that assertion to be true. For example, this allows us to model a conversation in which the interlocutors do not know how water presents under the relevant conditions, in order to know how the world must be in order for the assertions in the conversation to be true. It also allows us to model a conversation in which the interlocutors know how

water presents in some, but not all, of the relevant conditions. Being able to model whole conversations allows us to explain why a later assertion is, or is not, informative, given the force of an earlier assertion, as we discussed in introducing the problem of informativity in chapter one.

## Chapter Six: Proposed Extension to Jackson's and Chalmers's Accounts

In this chapter, I discuss my proposal for how to interpret the three-dimensional framework as an extension of Jackson's and Chalmers's epistemic two-dimensional semantics to serve as a solution to the problem of informativity. As on Stalnaker's account, the approximating intensions are not, themselves, part of the *semantics*, though they stand in an important relation to a semantic value—A-intensions on Jackson's view, and primary intensions on Chalmers's view—and play an important role in the notion of the a priori.

### *Section 1: The Proposed Extension as Pragmatic Reinterpretation*

Recall that Jackson and Chalmers adopt epistemic interpretations of the two-dimensional framework. Though their interpretations differ slightly, with the former being grounded in the conventions of the language, and the latter being grounded in the cognitive significance of the expressions for speakers, as discussed in sections 1.1 and 1.2 of chapter three, both approaches take the A-intensions or primary intensions, respectively, to be part of the *semantics* of the expression, and that they are to be used to explain the role that the expression plays in reasoning. In section 2, I will discuss my proposal that Jackson should not take the approximating intensions and the approximating truth-conditions to be part of the semantics, though he should take them to stand in an important relation to the conventions for the expression, and to the truth-conditions for the sentence, respectively. In section 3, I will discuss two ways that Chalmers could

think about the relation between the primary intension of an expression and the approximating intension for the expression, and between the truth-conditions of the sentence and the approximating truth-conditions, respectively.

### ***Section 2: A Proposed Extension to Jackson's Account***

We will begin by discussing my proposal for situating approximating intensions within Jackson's account. Recall that, on his account, the A-intension for an expression captures the conventions of use for that expression. This means that the A-intension is the same for all users—for the sake of charity, we will ignore the statements that he makes that are suggestive of individuals having their own primary intensions because that would lead to the possibility of ambiguity on his account. To avoid ambiguity, we'll need an A-intension that captures the conventions of use for expressions, rather than the A-intension for individuals.

To see this, consider the following. If the C-intensions contribute to the truth-conditions of an assertion, and the A-intensions and C-intension are, as he says, co-extensive at the actual world, and given that the same property determines both the A-intension and the C-intension, then if A-intensions and C-intensions may vary across speakers, and two speakers associate different properties with the expression, the expression is rendered ambiguous. For example, recall Jack, who, while engaging in conceptual analysis, only considered samples of water in the UK. Suppose that there is another individual—call him Mac—that has spent his entire life in the United States, and so when engaging in

conceptual analysis for ‘water’, he only thinks of the bodies water that he’s seen before: Lake Michigan, Lake Erie, the Mississippi, and his local fishing hole. But then the property in common across the samples that Jack considers is *being the predominant clear odorless liquid in the UK*, and the property in common across the samples that Mac considers is *being the predominant clear odorless liquid in the United States*. These will determine a different A-intension and C-intension for Jack and for Mac. But, if C-intensions of expressions contribute to the truth-conditions of assertions in which they are a constituent, and if it is the same property for both the A-intension and C-intension (save for the fact that the A-intension is actually-involving) then assertions in which ‘water’ is a constituent will be ambiguous. For this reason, it is better for a C-intension to be determined that captures the convention of the language.<sup>75</sup> Provided that the conventions of use for a given expression do not render that expression ambiguous, then this avoids that worry. We will discuss some further concerns about determining the conventions of use for an expression in section 2.1 of this chapter

However, we can use the approximating intensions for the purposes of capturing what Jackson might have had in mind in saying things that suggest that individuals have their own A-intensions for expressions, such that A-intensions for expressions could vary across individuals, or for the same individual across times. First, it must be noted that, in order to avoid ambiguity, the approximating intensions cannot be part of the semantics for the expression. We can think of the role of approximating extensions as capturing the individual’s grasp, at the time of the utterance, of the conditions under which the

---

<sup>75</sup> For more on determining the C-intension as a convention of use, see my discussion of this in chapter three, section 1.1.



expression applies. According to that individual's grasp of the conventions of use for that term, the conditions under which the term applies are met when something is presented as having the qualitative and functional characteristics that the individual expects that which is in the extension of the term to have.

To situate approximating intensions within his view, we can think of these conditions for being in the extension of that individual's approximating intension for an expression as the individual's current hypothesis about the conventions for that expression, at least as it applies to the present context. In contexts in which the hearer knows how the individual or kind that is in the extension of the proper name or natural kind term, respectively, presents under the relevant conditions, the hearer has true beliefs about the conventions for that expression (at least as relevant to that context). However, in contexts in which the hearer does not know how the individual or kind presents under the relevant conditions, there is something that she doesn't know about the conventions of the expression as used in that context.<sup>76</sup> This explains why, on my proposed extension of his account, we need only appeal to approximating intensions in modeling the force of assertions made in contexts in which (at least some of) the interlocutors don't know how the individual or kind presents under the relevant conditions. In cases in which interlocutors do know how the individual or kind presents in the conditions that are relevant to the truth-conditions for the assertion, their hypothesis is correct about the conditions under which the expression applies in that context, and so we do not need to

---

<sup>76</sup> The hearer need only know how the individual or kind presents under the conditions that are relevant to the truth of the assertion. For example, if you hand me a glass of water, and say, 'Here's a glass of water', in order to know that what you said is true is, all I need to know about how water presents is that, under room temperature conditions, water is presented as a clear odorless liquid. I needn't know how water presents under other conditions that do not obtain in order to know whether what you said is true.

appeal to approximating intensions because we can model how they take the assertion to change the state of the conversation appealing only to the standard interpretation. In cases in which the interlocutors do not know how the individual or kind is presented under the relevant conditions, we are unable to accurately model the force of assertions, as we've seen with the case of Bruce's assertion, 'there is water in the basin'.

Consider the conversation in which Bruce's assertion takes place. At the time of his assertion, Alice doesn't know how water presents under the relevant conditions (i.e., conditions in which the ambient temperature is below 32 degrees Fahrenheit). That conversation serves as an opportunity for her approximating intension to become a better approximation of the A-intension for 'water'. When Bruce makes the follow-up assertion to inform the students that water has a solid state, which we call ice, and that it changes into ice at temperatures below 32 degrees, Alice is able to update her beliefs about the way that water presents under cold conditions, and to thereby make her approximating intension a better approximation of the A-intension for 'water'. More generally, the more she learns about the way in which that which is in the extension of the expression presents itself under various conditions, the more her beliefs about the conventions for that expression approximate the A-intension for that expression.

### *Section 2.1: Questions about Conventions*

One presumption here is that the conventions of the language themselves accurately capture the way that which is in the extension of an expression is presented. This raises a

few questions. First, *what* qualitative and/or functional characteristics are included in the conditions for applying the expression, and what grounds them as the conventions for that expression? Starting with the second part of the question, there are at least the following candidates for what grounds the conventions for an expression.

The first candidate for grounding the conventions of use for expressions is that which is used to fix the reference of the expression, whatever it was. The intuitive attraction of this approach can be seen in considering paradigmatic examples used to capture the relationship between the natural kind term and the reference-fixer. For example, suppose that the reference of ‘water’ was fixed by the description, ‘the predominant clear odorless liquid that falls as rain; comes out of taps; constitutes rivers, lakes, and oceans, is used to bathe’. On this approach, the conditions that must be met in order to satisfy the description also serve as the conditions of use for the convention. The thought seems to be that the same description both introduces the term and grounds the conventions for its use. Note that we could have just as easily used an example in which it is stipulated that a reference-fixing property—say, the property of being the predominant clear odorless liquid that falls as rain; comes out of taps; constitutes rivers, lakes, and oceans; is used to bathe—to determine the convention.

There are at least two attractions of using the reference-fixing description or property to determine the conventions of the expression. First, since the description or property was used to fix the reference, in also using it to ground the convention for that expression, it ensures that, in using the expression in accordance with the conventions, one is

successfully using that token of the expression to refer to that to which the expression has been fixed. A second attraction, which also explains why Jackson takes his approach to be an epistemic interpretation of the two-dimensional framework, is that the description or property used to fix the reference of, and to determine the conventions for, the expression also captures the cognitive significance of the expression. The description or property makes use of the qualitative and/or functional characteristics that individuals associate with the expression.

One concern about that approach is that it fails to capture the uniform extension intuition: if we understand that ‘water’ is a natural kind term, and that natural kind terms refer to all and only samples of the same kind, then, in addition to those samples that satisfy the description, we will also want all samples of the same kind to be in the extension of ‘water’. Thus, a second approach, modifying the first, says that the convention is determined by the reference-fixer *plus* all and only samples of the same kind as that which is picked out by the reference-fixer. This is very much like our discussion of second-order properties in chapter three, and will ensure that, in using ‘water’, one thereby refers to all and only samples of H<sub>2</sub>O.

However, as we saw in chapter two, the reference of an expression may well have been fixed by ostension, or some other means that did not make use of some characteristics that individuals commonly associate with the expression. Consider again our example in which ‘water’ was introduced using the description, ‘the liquid that pools in St. Nectan’s Kieve’: that made use of a description that no one (except, perhaps, the members of that

very small community) associates with 'water'. Similarly, consider if the reference of 'water' was fixed by ostension. Suppose that the reference had been fixed by someone pointing at the liquid in St. Nectan's Kieve, and saying, "call *that* water'. Neither of these cases make use of the qualitative and functional characteristics commonly associated with 'water'. Thus, in these cases, the reference-fixer cannot be used to ground the conventions for the expression, if the conventions are intended to serve as the conditions under which we are to apply the expression in such a way as to thereby capture the cognitive significance of the expression.

I take these to be serious challenges to the view that the conventions are grounded in the reference-fixer, if the conventions also capture the cognitive significance of expressions. That there are serious challenges is worrisome because such a relationship between the conventions of use and the reference-fixer is an attractive way of motivating the claim that the truth-conditions of at least some sentences are knowable a priori. If one is competent with the constituent expressions of an assertion because one (at least tacitly) knows the conventions for those expressions,<sup>77</sup> and if the conventions have this relationship to the reference-fixer, then in (at least tacitly) knowing the conventions, one (at least tacitly) knows the truth-conditions for that assertion. If we'd like to claim that competent speakers have at least a tacit understanding of the truth-conditions for an assertion, such that they can use them to determine which worlds are excluded in light of that assertion, we will need to find another way of motivating this within the proposed extension of his view.

---

<sup>77</sup> See pages 33 and 64 of *From Metaphysics to Ethics* for discussion that is suggestive that Jackson thinks that this is a matter of competence.

A third approach to grounding the conventions has to do, not with reference-fixing, but with use. On this approach, it doesn't matter how the terms were introduced. What is important is how individuals use them. Thus, if the convention for 'water' is to use it to refer to that which is the predominant clear odorless liquid that falls as rain, comes out of taps, and so on, that convention is in place *because* competent speakers use 'water' in that way.

This has the attraction of not holding the theory hostage to a posteriori facts about the way the reference was fixed for expressions. It also has the attraction of making it fairly straightforward to work out the conventions for an expression—we just have to investigate how it is used. In fact, interestingly, this approach fits nicely with the conceptual analysis defended by Jackson (and Chalmers): in engaging in conceptual analysis, we learn how that individual does, or would, use that expression under various conditions.

This raises a question: is it *actual* use that determines the convention, or is it *actual and counterfactual* (i.e., the ways that she would use the expression in counterfactual situations) usage? Here's what I mean. One way of filling in the details of this approach is to look at the actual usage—the way that competent speakers *actually* use the expression. Thus, in order to work out the convention of 'water', we look at how competent speakers use the expression in practice.

One reason that this approach is attractive is that it is much simpler than the approach on which the convention is determined by actual *and* counterfactual usage. In order to determine the convention, we need look no further than the individual's actual use of the expression. We needn't ask them to engage in conceptual analysis in addition to reviewing their actual use of the term. Another reason that this approach is attractive is that competent users tend to not be uncertain of their usage. For this reason, this approach gives us clear, determinate conventions: either the individual applied the term under such-and-such conditions, or she did not.

However, there is at least one concern for this approach. For each competent speaker, the conditions under which she *would* apply a term may well outstrip the conditions under which she *actually* applies the term. On the approach in which we only look at actual use, what should we make of her dispositions to apply the term in conditions that happen to not arise? Suppose that a new set of conditions is introduced—what are the conventions for that case, and what determines them? It cannot be the conventions already in place because this is a new circumstance to which they do not apply. Furthermore, given that her actual use is a function of her disposition to apply the term in that way under those conditions, on what grounds do we privilege the ways in which she does act on her dispositions from over the ways that she would act on her dispositions, if only the right situation presented itself, though it never happens to do so.

Adopting the approach on which conventions are based on both actual and counterfactual usage avoids these concerns, and it seems to fit better with Jackson's prescription for

determining the conventions for an expression through conceptual analysis, which we will adopt here. To determine the conventions for an expression, we ask competent speakers to engage in conceptual analysis for that expression: the conditions under which they would apply the term determine the convention for that expression.

There are concerns about this approach as well. First, this assumes that we have transparent access to our dispositions regarding the conditions under which we would apply these expressions. It might well be that we might be a poor judge, in some cases, of what we would say in at least some cases—what we might say in imagined scenarios might differ from what we would say if we found ourselves in in such an environment. That is, while imagining the scenario, we might judge that we would say that there is water in the scenario, but if we were actually in situation qualitatively indistinguishable from that scenario, we might judge that there wasn't water in that case after all. It may also be that we do not have a determinate answer for every case. If asked about a scenario that seems far removed from our everyday experience, we may find that we aren't sure what we would say about that case.

A second concern arises from differences in how individuals would apply the term in counterfactual situations that never arise. Consider two individuals who use 'water' correctly on all occasions, and so seem to have a very good handle on the conventions that govern the use of 'water'. According to what Jackson says about competence, both of these individuals count as competent. they are competent in their use of the expression in the contexts in which they have heard it used, or used it themselves, for a very long



time. Suppose that we ask these two competent speakers how they would use the term under various counterfactual conditions, and they give different answers for the same conditions. *Which* of these counts as the ones that ground the convention for that condition? Does one speaker count as competent in this case while the other does not, or do they both count as incompetent? Note that they cannot both count as competent in this context, on pains of ambiguity for that expression in that context.

Another worry arises in cases in which all speakers always use ‘water’ correctly in all the contexts in which they find themselves, but they all have a mistaken belief that ice is of a distinct natural kind than liquid water. Imagine that the world’s population all lives between the Tropic of Cancer and the Tropic of Capricorn, and so the weather is always too warm for ice to form, and, also, that they don’t have freezers. Suppose that a world-wide poll will be taken about each competent speaker’s judgments about a series of images that they will be shown. (Perhaps they are emailed to everyone, or perhaps they are sent the series of pictures in the mail—it doesn’t really matter.) Each competent speaker is asked to fill out a questionnaire that asks, for each image, whether if that situation obtains, is there any water in that situation, and, if so, what in the image is water. Each competent speaker says that, for each image depicting a clear odorless liquid, if that situation obtains then there is water in that situation, and says that, if it obtains then the clear odorless liquid is water. However, when they are shown pictures of icebergs, ice cubes, glaciers, etc., they say that if those situations obtain, there isn’t water in those situations.

One might be tempted to think that there isn't a worry here. Since whatever property determines the convention has the unifying feature, such that it picks out all and only samples of the same kind, then so long as the property that is common to all the things that each judges would be in the extension of 'water', provided that the situation obtains, is such that it is had by at least one sample of water, then 'water' picks out all and only water. The problem isn't that it would fail to pick out water uniformly, but rather than the convention governing the use of 'water' would fail to capture the cognitive significance for competent users: *all* competent users judge that ice is not in the extension of water, and yet the conventions of use for 'water' determine that 'water' is to be used to refer to ice.

It seems that we have come to a decision point: the convention can either capture the cognitive significance of the expression, or it can uniformly pick out the extension of the expression. It can only do the former if it lacks the unifying feature, and it can only do the latter if it has the unifying feature. I think that, given that the convention also serves as the expression's contribution to the truth-conditions of assertions in which it is a constituent, then the property that determines the convention should have the unifying feature. This fits best with both the uniform extension intuition and the contingent extension intuition.

What should be said about the relationship between the convention and cognitive significance of the expression? I suggest that the first order property discovered through an individual engaging in conceptual analysis for an expression captures only the

approximating intension for that expression for that individual. That approximating intension captures that individual's grasp of the conventions of use for that expression. It also captures the cognitive significance of that expression for her: it captures what she understands when she hears assertions in which that expression is a constituent, what she means to convey when using that expression, and so on.

We will adopt this proposal for how to think about the conventions of use of expressions. In adopting this way of thinking about conventions, however, it may sound odd to call them *conventions*. After all, this way of thinking about conventions allows for the case in which none of the competent speakers take 'water' to apply to ice, and yet the convention for 'water' is to use it to pick out all and only samples of the same kind, H<sub>2</sub>O. For this reason, we cannot think of this as a *descriptive* account of conventions: the conventions aren't determined by the ways in which the speakers would use the term in both actual and counterfactual situations. Instead, it is a *prescriptive* account: this is how we *should* use these expressions. This seems reasonable when considering that, in using a proper name, we intend to refer to the same individual uniformly, and, in using a natural kind term, we intend to refer to all and only samples of the same kind uniformly. The view of the conventions on which the property that determines the convention has the unifying feature ensures that, in following the convention, we refer to the same individual or kind uniformly. This fits well with both the uniform extension intuition and the contingent extension intuition, since both involve referring to the same individual or kind uniformly. In the next section, we will discuss the relationship between the A-intension of an expression and an individual's approximating intension for that expression.

*Section 2.2: Approximating Intentions as a Means of Acquiring Competence with an Expression, and Extending One's Competence to New Conditions.*

As we discussed in chapter three,<sup>78</sup> Jackson's A-intensions capture the conventions of use for an expression. For example, the A-intension of 'water' captures our (perhaps only tacitly grasped) convention of using 'water' to refer to all and only samples of that kind of which some samples have the property, for example, of being the predominant clear odorless liquid.

Let us suppose that Jackson is right that there are such conventions, and that they are roughly as developed within the second approach, above, setting aside (as we decided to do at the end of the preceding section) questions of what grounds those conventions. We are provisionally adopting this approach to conventions because it seems to best fit with the spirit of his account.

It seems exceedingly likely that different individuals will have different takes on the conventions of use for an expression: some speakers will have a better handle on the convention than others, and some may even have false beliefs about the convention. For example, someone who knows that water has a solid state in addition to a liquid state will know that we use 'water' to apply to water in its solid state, whereas an individual who doesn't know that water has a solid state will not be aware of this aspect of the conventional use of 'water'. Furthermore, someone could have a mistaken belief about the convention for 'water' such that she believes that 'water' is applied to something

---

<sup>78</sup> See chapter three.

with, say, a purple crystalline appearance because she mistakenly thinks that this is one of the states that water can be in under certain conditions.

I propose that we understand the individual's approximating intensions, or, to make it more in line with Jackson's terminology, the individual's approximating A-intensions, to capture that individual's grasp on the convention. When an individual has a sufficiently good grasp of the expression's A-intension regarding how the individual or kind is presented under the relevant conditions, such that the individual's approximating A-intension for the expression includes the relevant samples, then she will understand how the assertion says the world is. Provided that the other interlocutors have a similarly good grasp on the convention (or at least the majority of them do, or that the interlocutors to which the most weight have been assigned do), then we can model the force of the assertion using the standard interpretation, as understood on Jackson's account.

For example, suppose that Bruce's assertion had taken place on a warm day instead, such that the water in the basin was in its liquid state. In asserting, 'there is water in the basin', the water in the basin is in the A-intension of that token of 'water' in that assertion, and the students' approximating A-intensions for 'water' are such that the sample of liquid water in the basin is also in the extension. From this, we can see that they are in a position, epistemically speaking, to understand how the contents of the basin must be, given the current warm ambient temperature, in order for Bruce's assertion to be true in that circumstance, and so we can model the force of that assertion using the standard semantic interpretation.

However, if the interlocutors do not know how water will be presented under the relevant conditions, we need to appeal to their approximating A-intensions in order to explain how the assertion changes the state of the conversation. On my proposed extension of Jackson's account, an individual's approximating A-intension for an expression serves as the contribution of that expression to her approximating truth-conditions for assertions in which the expression is a constituent. In cases in which we need to appeal to the interlocutors' approximating A-intensions for at least one of the constituent expressions, the extension of their approximating A-intensions differ from the extension of the A-intension for that expression in such a way as to make the interlocutors' approximating A-intension (i.e., the set of worlds that are such that the interlocutor judges that what is said, according to her approximating truth-conditions, is true) for the assertion differ in truth-value from the A-intension for the assertion. For this reason, we cannot use the A-intension to explain the way that the assertion changes the state of the conversation.

On this proposal, we can think of the interlocutors as engaged in two very closely related projects. The first project is investigating the nature of that which is in the extension of the expression (e.g., learning about water, including how it is presented under various conditions, what one can do with it, where it can be found, and so on). The second project is investigating the conventions governing the use of that expression (e.g., learning about the conventions governing the use of 'water'). These two projects progress in tandem: in learning more about how water presents under various conditions, where it can be found, what one can do with it, and so on, one learns more about the

conventions of the use of ‘water’, and so about the A-intension for ‘water’. In learning more about the A-intension of ‘water’, one comes to have new expectations about how water presents itself under the relevant conditions.

When the interlocutors don’t know how the individual or kind will present under the relevant conditions, we use their approximating A-intensions for the proper name or natural kind term to construct their individual approximating matrices, and then combine them to construct the three-dimensional matrix. This proceeds in the same way as discussed in chapter four. We can then use the three-dimensional framework to model the force of assertions in cases in which the interlocutors do not know how the individual or kind presents under the relevant conditions.

### ***Section 3: A Proposed Extension to Chalmers’s Account***

We begin by discussing a worry that is commonly raised for Chalmers’s view, to which we will offer a solution later in this section. Sometimes, Chalmers talks as though *individuals* have primary intensions for expressions, such that these primary intensions vary across individuals, and even vary for the same individual across time. This means that two individuals could have different primary intensions for ‘water’, such that the two intensions are two distinct functions that pick out two different sets of worlds. This also means that the same individual could have a different primary intension for ‘water’ at one time than at another.

This approach to thinking about primary intensions ties the primary intension closely to the epistemic state of the individual at that time, and so to the cognitive significance of the expression for that individual at that time. The qualitative and functional characteristics that the individual associates with that natural kind term fix the reference of the term when evaluated counterfactually, and, at each world in the context set, the primary intension of that term refers at that world to the individual that has those characteristics, or all and only samples of the kind that is such that at least some samples that kind have those qualitative and functional characteristics. In other words, the individual primary intension has the unifying feature.

At other times, Chalmers talks as though *expressions* have primary intensions which would, presumably, be invariant across individuals. On this way of thinking about primary intensions, ‘water’ has a primary intension, rather than individuals each having their own primary intension for ‘water’. On this way of thinking about primary intensions, there isn’t any variation in the primary intension of ‘water’ across individuals: the primary intension for ‘water’ is a function from worlds to extensions that, at each world, picks out the same extension for each individual. The reference-fixer for counterfactual evaluation in this case is a set of qualitative and functional characteristics that are *commonly associated* with the expression. Even though precisely which qualitative and functional characteristics are associated with an expression may vary by individual, the ones that serve as the reference-fixer for the expression are those that are widely shared across users.



Some very reasonable questions about primary intensions include the following. Are there *both* individual primary intensions and a primary intension for expressions (that is invariant across individuals)? If so—and at least some of their work is quite suggestive that he is committed to both—then what is the relationship between the two kinds of primary intensions? I think that the three-dimensional framework gives an attractive account of the relationship between these two intensions.

I propose, rather than adopting the approximating intensions in addition to both kinds of primary intensions, they are adopted instead of individual's primary intensions. On this proposed view, there are two semantic values: the primary intension of the expression and the secondary intension of the expression. There is also, solely for the purposes of pragmatic reinterpretation, the individual's approximating intension, which we will call the individual's *approximating primary intension*. It is called the individual's approximating primary intension because it approximates the primary intension of the expression. The more the individual knows about how the individual or kind presents under various conditions, the more closely her approximating primary intension approximates the primary intension of the expression.

If the individual primary intension lacks the unifying feature, then the individual primary intension is co-intensive with that individual's approximating primary intension. If the individual primary intension has the unifying feature, then the individual primary intension is co-intensive with the primary intension for that expression. In either case, the individual primary intension isn't doing work that isn't already being done by another

feature of the view. Let's look at each way of thinking about the individual primary intension, starting with the approach on which it has the unifying feature.

If the individual primary intension has the unifying feature then, at each world, it picks out the same extension as the primary intension for the expression. For example, suppose that the individual expects water to always present as a clear odorless liquid—she doesn't know that water has a solid state, for example. She associates the property of being the predominant clear odorless liquid with 'water', and so this property determines her individual primary intension for 'water'. Since the individual primary intension has the unifying feature, it picks out, at each world, all and only samples of the kind at that world that has the property of being the most common clear odorless liquid, including those samples of that kind that aren't in the liquid state, and so aren't presented in the ways she expects water to be presented. The primary intension for 'water' also picks out, at each world, all and only samples of that kind that has the property of being the most common clear odorless liquid.

There are two concerns with this way of thinking about individual primary intension. First, the individual primary intension is doing the same work that the primary intension for the expression is doing. The two intensions are co-intensive at all worlds. This makes the individual primary intension superfluous. The second concern is that, in picking out water presented in ways that the individual does not expect, it does not capture the cognitive significance of 'water' for her, which is what the individual primary

intension was intended to do. For this reason, it is unattractive to adopt the individual primary intension as having the unifying feature.

If the individual primary intension lacks the unifying feature then it is co-intensive with the individual's approximating primary intension: both pick out, at each world, all and only those samples that are presented in the ways that the individual expects samples of that kind to be presented. If the individual expects that water is always presented as a clear odorless liquid, then both intensions pick out all and only samples of clear odorless liquids.

The attractive feature of this approach is that it captures the cognitive significance of the expression for the individual. However, if it is a semantic value, and Chalmers sometimes says things that suggest that it is, then since it won't be co-intensive with the primary intension for the expression (which is also a semantic value) then this risks rendering expressions ambiguous. Another consideration is that, since it is co-intensive with the individual's approximating primary intension for the expression, the individual primary intension isn't doing anything over and above that which is done by the approximating primary intension. Crucially, since the approximating primary intension is *not* a semantic value for the expression, in adopting the approximating primary intension as a replacement for the individual primary intension, the approximating primary intension captures the cognitive significance of the expression for the individual while avoiding the ambiguity worry. For these reasons, I recommend replacing the individual primary intensions with the approximating primary intensions.

Now that we have situated approximating primary intensions within the view, we can discuss the interpretation of the three-dimensional approach that I recommend as an extension of Chalmers's view. The default approach to explaining the way in which assertions change the state of conversations is to use the two-dimensional matrix: the worlds that are false on the diagonal are excluded. The primary intensions of the constituent expressions are those expressions' contribution to the epistemic truth-conditions,<sup>79</sup> as Chalmers sometimes calls them. In cases in which the interlocutors know how the individual or kind is presented under the relevant conditions, we can appeal to the default approach of modeling how the assertion changes the state of the conversation. If a world satisfies the epistemic truth-conditions, what is said, according to the epistemic truth-conditions, is true, and so the cell on the diagonal for that world is marked as true. If not, it is marked as false. The interlocutors exclude those worlds that fail to satisfy the epistemic truth-conditions, i.e., those marked as false on the diagonal.

However, in cases in which at least some of the interlocutors do not know how the individual or kind is presented under the relevant conditions, we must appeal to the three-dimensional pragmatic reinterpretation in order to account for how the assertion changes the state of the conversation. The reinterpretation on the recommended extension to Chalmers's view proceeds in the same way as the general explanation of the reinterpretation in chapter four. First, an approximating matrix is generated for each interlocutor, using his or her approximating primary intensions for expressions to determine her approximating truth-conditions for the assertion. For each world in the

---

<sup>79</sup> See chapter three, p. 23.

context set, if the approximating truth-conditions are satisfied for the assertion at that world, then that world is marked as true on the diagonal, and otherwise false. Alternatively, credences can be assigned to the cells on the diagonal instead of true or false, as explained in section 4.5 of chapter four. Once an approximating matrix has been generated for each interlocutor, they are combined to generate the three-dimensional matrix. Weights can then be assigned to the individual matrices, and then the calculation may be performed as explained in chapter four.<sup>80</sup>

Adding the extension to his account will allow for his view to accommodate the problematic cases. Furthermore, on the proposed account of primary intensions of expressions and secondary intensions, his account avoids the charge of ambiguity of expressions. Since the primary intension and secondary intension of expressions are co-extensive at the actual world, this avoids ambiguity of assertions at the actual world. Furthermore, since we evaluate the modal profile of assertions either counterfactually (e.g., for determining how that assertion changes the state of a conversation) *or* counterfactually (e.g., for the purposes of capturing the truth-conditions for the assertion, and the first proceeds by using only the primary intensions of the constituent expressions while the second proceeds by using only the secondary intensions of the constituent expressions, then this does not render the assertion ambiguous in its evaluation across worlds. In other words, if we are engaged in counterfactual evaluation, then we are only using the primary intensions, and if we are engaged in counterfactual evaluation, then we are only using the secondary intensions—for no kind of evaluation of assertions across

---

<sup>80</sup> P. 53.

worlds are we using *both* intensions for that evaluation—and so this avoids rendering the assertion ambiguous when evaluated at other worlds.

However, one might wonder whether this way of thinking about primary intensions of expressions, which we will henceforth refer to as the primary intensions, is in the spirit of Chalmers's account in one very important respect: in picking out the same individual, or all and only samples of the same kind, primary intensions fail to capture the cognitive significance of a proper name or natural kind term for an individual. For example, as we've discussed, the primary intension of 'water' picks out ice, even though Alice does not think that ice is of the same kind as water, and so the primary intension of 'water' fails to capture the cognitive significance of 'water' for her.

For this reason, the primary intension *can*—though need not—be developed in a way in which the connection to cognitive significance is made a central feature. This way of developing primary intensions is especially attractive for those who, like Chalmers, are interested in motivating a strong connection between conceivability and possibility. In the next section, we will develop this account of primary intensions, and in the sections that follow, we will discuss how it motivates the claim that conceivability is a good guide to possibility.

### *Section 3.1: Primary Intensions and Cognitive Significance*

One concern with developing an account of primary intensions in which they capture the cognitive significance of expressions is that there is no cognitive significance of

expressions *simpliciter*. All cognitive significance is cognitive significance *for some individual*. However, we cannot use the cognitive significance for any actual language user, as (a) cognitive significance of expressions varies across language users, and (b) it is extremely plausible that, given the unifying feature of the primary intension, the primary intension won't capture the cognitive significance of any actual language user.

One way to attempt to defend the view that the primary intension does capture the cognitive significance of the expression, or at least something very close to it, for many speakers, is that the property that is used to determine the primary intension, considered without the unifying feature, does seem to come close to capturing it. I take it that this is the way in which the primary intensions is intended to be understood as capturing the cognitive significance of expressions, and that the unifying feature is added for the purposes of avoiding ambiguity by ensuring that the primary intension and the secondary intension are co-extensive at the actual world.

I think that there is something attractive in this, but I think that, if one wishes, one could develop an account of primary intensions on which they *do* capture the cognitive significance of the expression for an individual. This is desirable if one is committed to the claim that the primary intension of the expression captures the cognitive significance of the expression. In order to avoid ambiguity, the primary intension of a proper name or natural kind term must pick out at the actual world the same individual, or all and only samples of the same kind, uniformly.

However, if the primary intension of a term uniformly picks out the same individual, or all and only samples of the same kind, at the actual world, then it does not capture the cognitive significance of the expression for at least many speakers. This is because it would require that the cognitive significance of the term includes all the ways that that which is in the extension of the term is presented under all conditions. In turn, this is because if the primary intension picks out its extension by virtue of that which is in the extension having the properties that capture the cognitive significance of the term, and if the extension of the term is such that it picks out that individual, or all and only samples of the same kind, under all possible conditions, then the cognitive significance of that term must be such that it includes all and only ways that that which is in the extension is (or would be) presented under all possible conditions.

To see what I mean, consider the following example. Suppose that we adopt the claim that the primary intension of 'water' captures the cognitive significance of 'water'. Since we also want to avoid rendering 'water' ambiguous by allowing the primary and secondary intensions to differ in extension at the actual world, we'll want to make sure that the extension of the primary intension is such that it is co-extensive with the extension of the secondary intension. Since the secondary intension of 'water' picks out all and only samples of H<sub>2</sub>O, we'll need to develop our notion of the primary intension of 'water' to do the same. However, it seems highly plausible that *none* of us know how water presents under *all* conditions. Thus, if one wants an account of the primary intension for proper names and natural kind terms on which primary intensions capture the cognitive significance of these expressions, we will need a property that determines



the extension of primary intensions that both picks out at the actual world the same individual, or all and only samples of the same kind, uniformly, *and* that it picks out that extension by virtue of the cognitive significance of that expression for some individual that is like us ordinary speakers in important respects, though significantly more knowledgeable regarding how the world turns out, and is likely idealized in some respects, including perception, cognition, and so on. We will develop such an account of primary intensions in what follows.

We will think of the primary intension of expressions as an idealization. In particular, we will think of it as capturing the upper bound of cognitive significance that a being with sensory capacities like ours could have for that expression. In order to understand what I mean by upper bound, think of cognitive significance as falling on a continuum<sup>81</sup>, with the lower bound as the least one can know about that which is in the extension of the expression in order for that expression to have any significance at all to that individual, and the upper bound of the continuum as the most that an individual with perceptual capacities like ours could know about that which is in the extension of the expression. The guiding idea is that the cognitive significance of an expression for an individual captures the qualitative and/or functional characteristics that the individual expects the individual or kind named by that expression to have under various conditions. That is to say, just as the cognitive significance of ‘water’ for Alice is governed by the ways that she expects water to be presented under various conditions, the primary intension of

---

<sup>81</sup> Perhaps it would require multiple continuums, since, for example, one could learn that water has a gaseous state before learning that water has a solid state, *or* one could learn that water has a solid state before learning that water has a gaseous state—each would require a different continuum. One could learn various facts about the ways that water is presented under different conditions, and each of those orderings would require a distinct continuum.

‘water’ captures all the ways that an individual who knew everything there is to know about water would thereby expect water to be presented under all conditions.

We still need to say more to address the matter of *whose* cognitive significance is captured by the primary intension. It will need to be the cognitive significance for a being with perceptual and cognitive capacities like ours in *some* important respects—to be qualified in what follows—but who knows a lot more about the way the world is than we do. This being would know how, for a given proper name or natural kind term, the individual or kind, respectively, would be presented under all conditions. For this reason, we will introduce an idealized being with capacities in important respects like ours (though idealized in other respects) but who knows how each kind would present itself in any conditions in which it could be found.

Let us say that the primary intension for an expression captures the cognitive significance of that expression for a being that is idealized in such a way as to have sensory capacities like our own, and cognitive capacities much like our own, though suitably extended to allow for her to be able to know how that which is in the extension of the expression is presented under all conditions.

Consider the potential states of cognitive significance that can be had for an expression as on a continuum, with lower bound as the bare minimal amount one would have to associate with an expression in order for that expression to have any cognitive significance for that individual, and the upper bound as the idealized epistemic state that

is such that one knows how, under all possible conditions, that which is in the extension of the expression would present itself (to beings with perceptual and cognitive capacities of the kind we have), and all the things that might be done with that which is in the extension of the expression, and all the conditions under which that which is in the extension might be found, and so on.

For any expression that is cognitively significant for us non-idealized beings, that state of cognitive significance falls somewhere on this kind of continuum. For example, if you know more than I do about that which is in the extension of the expression, the location of your cognitive significance is further along the continuum towards the upper bound than mine. Suppose that we both know that water is the predominant clear odorless liquid that falls as rain; quenches thirst; forms rivers, lakes, and oceans, and that this is all that I know about water. Suppose that, in addition to knowing this about water, you also know that in its frozen state, it is a clear cold solid. This means that you know more than I do about how water is presented under various conditions. The location of your cognitive significance on the continuum is closer to the idealized end of the continuum than mine.

As one's beliefs change about that which is in the extension of the expression, the cognitive significance of that expression for that individual changes accordingly. Suppose that I learn that water is presented as a clear cold solid when frozen: the cognitive significance for 'water' for me moves closer to the upper bound of the continuum than it had been, since it is a better approximation of the upper bound.

What must this upper bound be like? Consider the epistemic state of an individual who knows how water presents under all conditions. This includes how water will present itself under various temperatures and pressures, as well as under various levels of “resolution”, i.e., to the unaided eye, or when viewed using a very powerful perceptual aid that allows one to see the most fundamental constituents, or viewed from a very great distance, and so on. Her epistemic state with respect to water (i.e., how water presents under all conditions) is the upper-bound of the cognitive significance for ‘water’. Furthermore, just as you or I can associate with ‘water’ the property of being the predominant clear odorless liquid, or whatever other property precisely captures the qualitative and/or functional characteristics that we associate with ‘water’, an idealized being of the sort under discussion could associate with ‘water’ the property that precisely captures the qualitative and/or functional characteristics that she associates with ‘water’, given her epistemic state with respect to water.

Clearly, *how* the idealized being might come to be in such an epistemic state raises some concerns. On the practical side, one might wonder about how the idealized being could make the relevant observations in order to be in that epistemic state with respect to water such that ‘water’ has that cognitive significance for her. Is she constrained by the same technology that we have, or will have in the future, or could have had in the future if we’d only applied ourselves differently in the relevant research, etc.? Or, is she such that she could make use of technology that, even in principle, we’ll never be able to develop?

Similarly, *all possible conditions* seems to include conditions that could have obtained at our world (i.e., physically possible conditions) that happened not to have obtained. Does that require her to be able to make observations of other *worlds*? How else would she go about making these observations, even in principle?

I'm raising this point about physically possible conditions that fail to obtain at our world here as a *potential* worry. Whether such a case is a genuine concern depends on how wide-ranging the physically possible conditions are beyond the ones that obtain in our world. It might be that there are no such further conditions—that is, that all of the possible conditions occur, at some point or another, at our world. This might require that the idealized being had a lifespan much longer than our own in order for her to be able to make all of the requisite observations, or that she, in some sense, has access to other worlds.

Though this is clearly a problem in practice, it is not as clear whether it is also a problem in principle. Suppose, for example, that she is able to acquire this information within minutes through a Matrix-style information upload. Another way to respond to both this worry and the worry about possible states that do not obtain at our world is to appeal to the in-principle possibility of what, in his *Constructing the World*, Chalmers calls a *Cosmoscope*. The Cosmoscope is a supercomputer that has infinite storage capacity and infinite processing power. Provided with full information about the microphysical facts of our world (and, for our purposes, we can include the macrophysical facts as well, for those who are skeptical of the claim that the microphysical facts entail the macrophysical

facts), and the information about the physical laws, along with facts about the qualitative experiences had by all individuals, the Cosmoscope can provide the user with a simulation of observations that one could have at any level of resolution, including counterfactual states of the world. Provided that the idealized being knew enough about water to know that it is H<sub>2</sub>O—by virtue of knowing that it is H<sub>2</sub>O that has the qualitative and function properties that she associates with ‘water’—then the idealized being could use the Cosmoscope to determine where the water is located in the simulated counterfactual situations by being able to view such situations both at a level of resolution that allows her to determine whether she is observing H<sub>2</sub>O, as well as a level of resolution that allow her to observe that sample of H<sub>2</sub>O under those conditions, in order to learn how H<sub>2</sub>O presents under those conditions.<sup>82</sup>

I won’t attempt to resolve this worry about whether this is an *in principle* problem here.

Because it has interesting consequences for the view, we’ll provisionally assume that

---

<sup>82</sup> This raises an interesting question of whether this presupposes actualism, and, if so, whether that is problematic. Even the possibilist can acknowledge that worlds at which there is no H<sub>2</sub>O aren’t relevant to our interest in the ways that H<sub>2</sub>O presents under all conditions. For this reason, worlds constituted out of different stuff, giving rise to worlds that don’t contain H<sub>2</sub>O, aren’t relevant. The possibilist can then ask, what about worlds that differ only *slightly* from the domain of this world, such as having all the same fundamental kinds, save for an additional kind of subatomic particle? Suppose that there is H<sub>2</sub>O at that world. This would be a world at which there is H<sub>2</sub>O but due to the differences in domain between that world and our own, the Cosmoscope may not be able to generate a simulation of these conditions. One consideration is whether or not water presents differently at that world than it does at any of the possible states of this world. The actualist, of course, will reject that there are any such worlds, but what should the possibilist say? Suppose that water doesn’t present differently. Then there is no concern. Suppose that water does present differently. This is where the issue gets really interesting. If the possibilist is a semantic externalist then there is the question about whether the water at that world is in the extension of ‘water’. Given that H<sub>2</sub>O, but not XYZ, is in the extension of ‘water’ because we are in causal contact with the former but not the latter, then it seems that perhaps the H<sub>2</sub>O at that world isn’t in the extension of ‘water’ because we aren’t in causal contact with the H<sub>2</sub>O in that domain. That world isn’t a possible state of our own, and so we aren’t in causal contact with the constituents of that world’s domain. If it is causal contact that matters, then it seems that those samples of H<sub>2</sub>O aren’t in the extension of our word, ‘water’. This suggests an interesting question for semantic externalism, and that the view may be in need of further refinement. I won’t attempt to address the issue further here, but I intend for this to be a topic of further research.

either there are no possible conditions such that they do not obtain at our world, or that if there are, they aren't problematic in principle for the reasons previously discussed, and bearing in mind that if either there are such conditions, or that if they are problematic in principle, then those aren't consequences of the view after all.

Next, we need to say more about her idealized capacities. In what way(s) is she an idealized version of a being with capacities like ours? Her perceptual capacities are clearly required to be extended beyond what our unaided perception can do. Similarly, it looks like at least some of her cognitive capacities need to be extended beyond ours as well.

In order to answer these questions, we need to consider in more detail what work the idealization is supposed to be doing. It is supposed to be capturing the best possible epistemic state with respect to that which is in the extension of an individual or natural kind term, and such that cognitive significance of that proper name or natural kind term, respectively, is at the upper bound of the continuum.

This suggests that the idealization of her perceptual and cognitive capacities should not be such that the capacities differ in kind from our own, and should not be extended in such a way that ours could not be possibly extended, at least in principle. Regarding extending her perception, suppose that we do not idealize her perceptual capacities—that is, we allow her to have the same sorts of perceptual capacities that we have—and extend

them through perceptual aids, much as microscopes let us see much smaller things, and telescopes allow us to see much further things, than we could see with our unaided eyes.

For example, the idealized being would know how water would be presented under all possible conditions in which a being with capacities like our own, suitably extended through perceptual aids, could observe it. The idealized being's epistemic state captures something like an upper bound of an epistemic state with respect to water that a being with sensory capacities like our own could have, suitably extended through perceptual aids, such that she would have accurate expectations of what water would look like under any conditions.

Developing this account requires specifying more precisely what is meant by 'a being with perceptual capacities like ours' as well as 'all the conditions under which a being with capacities like ours could observe it'. Some challenges for the first include addressing whether what is meant is what beings with capacities like ours could see through *unaided* perception, or whether this includes what we could see using perceptual aids (telescopes, microscopes, etc.). If the latter, is this restricted to *current* technology, or current technology plus technology we will go on to develop, or current technology plus technology we could go on to develop, whether we ever do or not?<sup>83</sup>

I won't attempt to suggest an answer here. However, I will point out that the further removed the state of technology that we adopt as the answer to that question is from the

---

<sup>83</sup> Note that, in this last case, we'll want to specify what we mean by "technology we *could* develop". Do we mean given the research that we've done, or given what's physically possible, or something else?



technology that we currently have, or even that we ever do develop, the weaker the motivation for the claim that actual instances of conceivability are a very good guide to possibility, for reasons that I will explain in what follows.

#### ***Section 4: The Extended Accounts and the A Priori***

Recall the notion of a priority of interest to Jackson and Chalmers, and so to us: the sentence is a priori if we can know it to be true independently of knowing the way the world turns out in the relevant respects. What is at issue, for example, is whether I can know a priori ‘water is the watery stuff’, where ‘the watery stuff’ is supposed to be an abbreviated version of the qualitative and/or functional characterization that we associate with ‘water’, without knowing how the underlying constitution of water turns out. The important question is whether associating the kind of property that we’ve been discussing with a proper name or natural kind term provides a priori justification for these kinds of sentences. For example, does the fact that I associate with ‘water’ the property of being the predominant clear odorless liquid provide a priori justification for my belief that ‘water is a clear odorless liquid’ is true?

As we’ve discussed, the thought seems to be that the primary intension of a proper name or natural kind term is intended to capture the cognitive significance of that expression: as Chalmers says in “The Foundations of Two-Dimensional Semantics”, “the role that the expression plays in reasoning and reference”.<sup>84</sup> As Chalmers discusses, Frege argues that, in addition to the extension of expressions, we also need to adopt a kind of meaning

---

<sup>84</sup> “The Foundations of Two-Dimensional Semantics”, p. 56.

that Frege called *sense* that captures the cognitive significance of expressions, in order to explain, for example, the difference in meaning between ‘Hesperus is Hesperus’ and ‘Hesperus is Phosphorus’. The first is trivial, and the second is an empirical discovery, yet both ‘Hesperus’ and ‘Phosphorus’ refer to Venus, so if meaning were merely extension, then we are unable to explain why only the first is trivial.

Sense is intended to capture the *mode of presentation* of that which is in the extension of the expression. We won’t investigate the notion of modes of presentation here in depth, and will understand it to mean how that which is in the extension of the expression is presented to us (i.e., how it appears qualitatively, and what functional characteristics we associate with it). The mode of presentation associated with the expression captures the cognitive significance of the expression. In knowing the mode of presentation associated with the name, one knows how that which is in the extension of the expression is presented (at least under the relevant conditions), and so one knows how the world must be in order to for assertions in which that expression is a constituent to be true.

Two different names, for example, could be co-extensive, and yet differ in cognitive significance because we associate a different mode of presentation with each name. For example, if we associate the name ‘Hesperus’ with that which is presented as the brightest celestial body visible just after dusk, and if we associate ‘Phosphorus’ with that which is presented as the brightest celestial body visible just before dawn, then though the same thing—Venus—is presented in both of these ways, making these two names co-

extensive, it is nonetheless cognitively significant to learn that ‘Hesperus is Phosphorus’ is true.

Since, on the Fregean view, sense determines extension, then grasp of the cognitive significance of expressions gives us access to the epistemic truth-conditions<sup>85</sup> of assertions in which the expressions are constituents. This is because the cognitive significance of a proper name or natural kind term is the contribution of that expression to the epistemic truth-conditions of assertions in which that expression is a constituent. Recall<sup>86</sup> that it is the epistemic truth-conditions of assertions that allow us to know whether the assertion is true independently of how the world turns out in the relevant respects, on the epistemic sense of turns out. For example, it is the epistemic-truth conditions for ‘water is the watery stuff’ that allows us to know that it is true independently of knowing how water turns out in important respects, such as that water is H<sub>2</sub>O, or even whether there is one (or more) qualitatively indistinguishable kinds in the oceans, falling as rain, etc., and so we can know it a priori.

One significant concern is that, depending on the cognitive significance for an expression for an individual, sentences of the form ‘(the) N is F’, where N is a proper name or natural kind term, and F is the property that the individual associates with N (i.e., the property that captures the cognitive significance of F for her), ‘(the) N is F’ is a priori for her because the property that she associates with N is F, and so the epistemic truth-conditions for her for ‘(the) N is F’ are such that she can see that they are satisfied, and

---

<sup>85</sup> I am borrowing ‘epistemic truth-conditions’ from his “Two-Dimensional Semantics”, p. 23. He doesn’t use the expression in his “The Foundations of Two-Dimensional Semantics”.

<sup>86</sup> See my discussion of this in chapter three, section 1.2.

yet it is false because that which is in the extension of N does not, in fact, have property F.

For example, consider Kripke's example of a case of a false belief about who proved the incompleteness of mathematics.<sup>87</sup> Suppose that I associate with 'Gödel' the property of being the person who proved the incompleteness of arithmetic. In this case, N is 'Gödel', and F is the property of being the person who proved the incompleteness of arithmetic. However, suppose that, unbeknownst to me, another man, *Schmidt*, proved the incompleteness of arithmetic, but Gödel became widely believed to be the one who did so because he cleverly was able to pass off Schmidt's work as his own. In this case, even though the epistemic truth-conditions are satisfied, the truth-conditions for that sentence, determined by the secondary intensions of the constituent expressions, are false. That is, the secondary intension of 'Gödel' picks out the man we call Gödel, and, according to the scenario under consideration, he did not prove the incompleteness of arithmetic. According to Jackson's and Chalmers's views, the C-intension or secondary intension, respectively, determines the truth-conditions for the sentence. Since knowing is factive, then I cannot know 'Gödel is the man who proved the incompleteness of arithmetic' a priori.

From this kind of consideration, we can see that epistemic truth-conditions generated by the cognitive significance of expressions for beings in epistemic states like our own—we know some things about the way the world turns out, but not everything, and our beliefs

---

<sup>87</sup> *Naming and Necessity*, p. 85, footnote 36. I've slightly modified the example to better fit our purposes: whereas Kripke discusses the description, 'the man who discovered the incompleteness of arithmetic', I am discussing the corresponding property.

about the way the world turns out include some false ones—are not guaranteed to allow us to know the truth of assertions of the form ‘(the) N is F’ a priori. The worry, as we’ve seen, is that in our impoverished epistemic state, the epistemic truth-conditions determined by the primary intensions of the constituent expressions may differ from the truth-conditions determined by the secondary intensions of the constituent expressions. Since it is the truth-conditions, as determined by the secondary intensions of the constituent expressions, that must be satisfied in order for the sentence to be true, then in cases in which those conditions fail to be satisfied, the sentence is false. Thus, this allows for the possibility of cases in which the epistemic truth-conditions are met simply by virtue of the stable association between the constituent expressions and the corresponding properties, and yet the sentence is false. These sorts of cases undercut our a priori justification. In other words, grasp of the sense or cognitive significance of ‘water’ for me (i.e., a finite being with an impoverished epistemic state) does not provide me with indefeasible a priori justification for ‘water is the watery stuff’.

Still, it is interesting to consider whether a view can be developed on which grasp of the cognitive significance of expressions does provide indefeasible a priori justification for believing that sentences of the form ‘(the) N is F’, such as ‘water is the watery stuff’, are true, and so allowing that individual to know ‘water is the watery stuff’ a priori. In order for the epistemic truth-conditions to allow for a priori knowledge, the epistemic truth-conditions for an assertion must be such that they are satisfied if and only if the truth-conditions for the assertion, as determined by the secondary intensions of the constituent expressions, are also satisfied. The epistemic state of the idealized being is such that she

knows how water is presented under all conditions, and so her epistemic truth-conditions are satisfied just in case the truth-conditions of sentences of the form '(the) N is F' are satisfied.

In order to investigate the plausibility of such a view, we will provisionally develop the view on which the extension of the primary intension of a term is determined by the property that the idealized being associates with that term. This is an account on which the property associated with the primary intension of an expression captures the cognitive significance of the expression for the idealized being. We will then consider whether, given the cognitive significance of some expression N for the idealized being, captured by the property F which she associates with N, sentences of the form '(the) N is F' are a priori for her.

Let N be 'water'. Since the property that the idealized being associates with 'water' includes *all* ways that water would be presented under all possible conditions, property F is the property of having the qualitative and functional properties that water would have under all possible conditions, relativized to the appropriate conditions. If it is by virtue of the association one has between 'water' and F that one can know this sentence to be true a priori, then this significantly limits who can know this sentence a priori. Perhaps *only* our idealized being could know it a priori. Surely, for the reasons just discussed, neither you nor I could know this a priori in practice: our epistemic state is not such that *we* can know that a priori by reflecting on which property we associate with that expression

because the property that *we* associate with ‘water’ differs from the property the idealized being associates with it.

It’s interesting that, since the idealized being associates property F with ‘water’, not only is such a sentence knowable a priori, but this knowledge is also indefeasible by experience. If F really captures all the ways that water is, or would be, presented under all possible conditions, then there is no further experience that could serve as a defeater to the a priori justification for the sentence.

One way to see this is to consider again the relationship between the primary and secondary intensions of such expressions. Recall that the truth-conditions of the sentence are determined by the secondary intension of that sentence. However, if proper names and natural kind terms are semantically stable, such that the primary and secondary intensions of that sentence map all the same worlds to true, then if the primary intension maps that sentence to true, then that world is also among those that the secondary intension maps to true. The primary intension of a sentence maps a world to true if and only if that world satisfies the epistemic truth-conditions for that sentence. The secondary intension of a sentence maps a world to true just in case the truth-conditions determined by the secondary intensions of the constituent expressions are satisfied. If the primary intension maps a world to true, then the secondary intension maps that world to true, too, and vice versa. If it is among the worlds that the secondary intension maps to true, then the sentence is true at that world.

We must note a very important feature of this kind of primary intension. The property associated with a proper name by the idealized being is such that the property picks out that individual uniformly *without requiring the unifying feature*. We can see that if the property that the idealized being associates with ‘water’ is such that it is had by all and only samples of water, this renders the unifying feature for the primary intension superfluous. That is, if the extension of the primary intension of ‘water’ picks out all and only those samples that have the property—*without the unifying feature*—that the idealized being associates with ‘water’, the primary intension would pick out all and only samples of H<sub>2</sub>O.

More generally, the property that captures the cognitive significance of a proper name for the idealized being determines that the primary intension of that proper name picks out the same individual uniformly. Since she knows how that individual is presented under all conditions, the property picks out that individual under all conditions, thereby picking out that individual uniformly. For example, since the idealized being associates with ‘Aristotle’ the property that captures the way that Aristotle is presented under all conditions, then in associating that property with ‘Aristotle’, the primary intension of ‘Aristotle’ picks out Aristotle uniformly.

Similarly, the property that she associates with a natural kind term is such that the property picks out all and only samples of that kind uniformly *without requiring the unifying feature*. For example, the idealized being associates with ‘water’ the property that captures the way that water is presented under all conditions. Since she associates



this property with ‘water’, the primary intension is such that it picks out all and only those samples that have that property, thereby picking out all and only samples of water. However, note that since she knows *all* the ways that H<sub>2</sub>O presents, then unless there is another kind at another world that is qualitatively indistinguishable in all respects from H<sub>2</sub>O,<sup>88</sup> her primary intension picks out not *only* all and only samples of H<sub>2</sub>O at the actual world, but rather at all worlds.

Such a primary intension, which, without the unifying feature, would pick out that individual, or all and only samples of that kind, regardless of presentation, would thereby be co-intensive with the secondary intension for that expression. But, since that the secondary intension determines whether what is said is true on the horizontal, i.e., as evaluated counterfactually at each world in the matrix, and that the primary intension determines whether what is said at each world is true at that world, then this would mean that the diagonal would be the same as the horizontal with respect to the truth-value of the assertion at each world. Indeed, for modeling conversations among idealized beings, the diagonal would be rendered superfluous.

Also, since the idealized being knows how, say, water turns out in all respects, there won’t be any worlds still epistemically possible for her that represent alternative ways

---

<sup>88</sup> This raises the question of whether there could be two different natural kinds that are qualitatively indistinguishable in all respects from one another. I won’t attempt to resolve this here, but Chalmers seems to suggest that he favors the thesis that there is no difference without an observable difference in *Constructing the World*. If this thesis is right then natural kind terms refer uniformly across worlds. For our present purposes, we will adopt this thesis for the sake of argument, keeping in mind that if the thesis is false, then the primary intension of natural kind terms may not be co-intensive with the secondary intension of those terms. The same can be said for the primary intensions of proper names, given that the extension of the primary intension at each world is determined by the property that the idealized being associates with that proper name.

that water could turn out, for all she knows. There *aren't* any alternative ways that water could turn out, for all she knows, since she already knows how water turns out in all respects. There aren't any such alternatives epistemically open to her.

However, this account of primary intensions would render primary intensions, and the diagonals that they determine, superfluous. That the primary intension is co-intensive with the secondary intension makes it the case that both are doing the same work. We would not need a two-dimensional matrix to rule out worlds. Yet, as we've seen, primary intensions are very useful within Jackson's and Chalmers's accounts for the purposes of explaining how assertions change the state of conversations. For example, they explain why, even though what we say in asserting, 'there is water in the glass' is false in a superficially indistinguishable situation at an XYZ world at which there is XYZ in the glass, nonetheless, for interlocutors who do not yet know whether water is H<sub>2</sub>O or XYZ, they do not rule out the XYZ world in light of the assertion. For this reason, we need for the diagonal to differ from the horizontal: if it does not then we cannot account for how such assertions change the state of conversations in which the non-idealized interlocutors do not know how water turns out in all respects by appealing to the diagonal.

For this reason, we can see that we don't want to say that *this* is the primary intension to be used to model the way that assertions change the state of conversations for those conversations with non-idealized interlocutors—namely, us—after all. Perhaps the best way to think about this primary intension is that it is the primary intension for an

idealized language that is useful for some purposes.<sup>89</sup> In order for the primary intension to do the work that we want it to do with respect to modeling communicative exchanges among non-idealized beings, we need a primary intension that is such that the extension is determined by a property that non-idealized speakers associate with expressions, and that has the unifying feature. This will allow the extension of the primary intension to vary across worlds, thus making it useful for our purposes. Nonetheless, even though the primary intension that we will want to adopt for our purposes is one that uses a property that is commonly associated with expressions by non-idealized beings such as ourselves, we can still use this notion of a primary intension to motivate the relationship between primary intensions and a priority. In order to keep track of the two primary intensions within our discussion, we will call them the idealized primary intension and the non-idealized primary intension.

Suppose that it really is the case that the first-order property (which thereby lacks the unifying feature) that the idealized being associates with a proper name or natural kind term uniformly picks out that individual uniformly, or all and only samples of the same kind uniformly, across all worlds at which it picks out anything at all. This would be so if the first-order property is F. Then, plausibly, sentences of the form ‘(the) N is F’ are a priori for her, since she knows that that which is in the extension of that proper name or natural kind term has the property of being F, and that the extension of ‘the F’ is also that which has the property of being F. Furthermore, since the property of being F captures

---

<sup>89</sup> This idealized language might well be of use to those interested in a “perfect language”, or a language that captures an isomorphism between our concepts and the things in the world which they are about. Such a language would indeed be useful for modal epistemology, as the conceivability of a given thought or sentence would thereby be a very good guide to possibility, if the case can be made for such an isomorphism between our concepts and/or expressions, and the things in the world that they are about.

all and only the ways that that which is in the extension of 'N' is (or would be) presented under all possible conditions, her a priori justification for believing that '(the) N is F' is true is indefeasible by experience.

Under these conditions, it seems quite plausible to say that the idealized being can know '(the) N is F' a priori. However, this doesn't settle the matter of whether '(the) N is F' is a priori *for us*. I think that the approximating intensions can do important work for settling this matter for us, too. To see this, consider that the idealized primary intension for 'water' and the non-idealized primary intension for 'water' are co-extensive at the actual world: the extension of the idealized primary intension for 'water' is all and only samples of H<sub>2</sub>O, and the extension of the primary intension of 'water' is also all and only samples of H<sub>2</sub>O. Thus, if an individual's approximating intension for 'water' is an approximation of the primary intension of the non-idealized primary intension for 'water' then it is also an approximation of the idealized primary intension for 'water'. More importantly, we can see that the cognitive significance of 'water' for a non-idealized being, such as oneself, falls on the continuum of cognitive significance for 'water', and approximates the upper-bound of the cognitive significance of 'water'.

In order to avoid confusion since we will be discussing both in what follows, let us adopt the convention of referring to the property that captures the cognitive significance of a term for the idealized being as property F, and of referring to the property that captures the cognitive significance of that term for a particular non-idealized being as property G. We can think of the first-order property that a non-idealized individual associates with an

expression as providing her with defeasible a priori justification for believing ‘(the) N is G’, where G is the property that she associates with N, as we discussed in chapter four. For example, let N be ‘water’ and let G be the property of being the predominant clear odorless liquid. Her justification is defeasible because G does not capture all of the ways that water is (or would be) presented under all possible conditions—it only captures those ways that she expects water to be presented under the conditions she expects it to be presented at all. But she might, for example, have the wrong expectations regarding how water would present under some conditions, and she might fail to have expectations about how water would be presented under other conditions, making her a priori knowledge defeasible by experience.

Why think that associating property G with a proper name or natural kind term N provides *a priori* justification at all, even if it is defeasible a priori justification? This is another point in which the continuum of states of cognitive significance does important work for us. The cognitive significance that N has for the non-idealized individual falls somewhere on the continuum of possible states of cognitive significance that N can have for individuals, with the upper-bound being the cognitive significance that N has for the idealized being. For this reason, the non-idealized individual’s justification for believing ‘(the) N is G’ is true is the same kind of justification that the idealized being has for believing that ‘(the) N is F’. The justification is of the same kind because both have the justification by virtue of associating a property—the property F, for the idealized being, and the property G, for the non-idealized individual—with N, and properties F and G are themselves of the same kind, namely properties that capture the way that the relevant

individual expects that which is in the extension of N to be presented under various conditions.

The difference between properties F and G is that F includes *all* ways that that which is in the extension of N is (or would be) presented under all possible conditions, whereas property G only captures expectations of some of the ways that that which is in the extension of N is presented under some conditions (and some of those predictions may be wrong). But, even in spite of that difference, the non-idealized being has defeasible a priori justification because of the relationship between the cognitive significance of N for her, and the upper-bound of cognitive significance of N. While it is true that our non-idealized individual isn't in a position to associate property F with N, she is in a position to associate property G with N. Furthermore, though property G doesn't determine the extension of N, another property very much like it in the sense of capturing the ways that that which is in the extension of N presents under various conditions—namely, property F, though she doesn't know which property that is—picks out an extension that is co-intensive with the extension of N. That there is some property F that picks out an extension that is co-intensive with the extension of N, and because property G is an approximation of property F, using property G to approximate the extension of N gives her defeasible a priori justification for believing sentences such as '(the) N is G' to be true.

Though it is too extensive of a project to investigate presently, the continuum of cognitive significance raises an interesting puzzle about the relationship between

cognitive significance and the a priori, on Chalmers's and Jackson's views. Recall that it is the cognitive significance of expressions that gives us access to the epistemic truth-conditions for assertions in which those expressions are constituents. Furthermore, recall that it is the epistemic truth-conditions that allow us to know that a sentence is true a priori. This is what allows us to know 'water is the watery stuff' a priori: the qualitative and/or functional characteristics for which 'the watery stuff' is an abbreviation captures the cognitive significance of 'water', allowing us to know that, regardless of how the world turns out, epistemically speaking, we know enough to know that 'water is the watery stuff' is true. That is, we can know 'water is the watery stuff' independently of knowing how water turns out with respect to its underlying constitution, and this is what allows us to consider various hypotheses about how the extension of 'water' might turn out on various hypotheses about the way our world might turn out. But, if knowing 'water is the watery stuff' a priori requires that we don't know how the world—more specifically, how *water*—turns out in some respects, i.e., the underlying chemical constitution of water, then are there are some epistemic states that are such that we know too much about the way the world turns out to count as knowing 'water is the watery stuff' a priori?

For example, consider the epistemic state of the idealized conceiver. She knows how water presents under all possible conditions, and the cognitive significance of 'water' for her reflects this knowledge. Her epistemic state is such that there are no alternative ways that water could turn out, for all she knows. There is a sense in which this doesn't present a problem for Jackson and Chalmers: since a priori justification is justification

independently of experience, then since her a priori justification for ‘(the) N is F’ is her stable association between N and F, then this qualifies as a priori. However, they also discuss the a priori in terms of independence from empirical information.<sup>90</sup> One reasonable way to understand what is meant by ‘independence from empirical information’ is that it can be known *independently of knowing how the world turns out in certain (empirical) respects*. This way of thinking about it fits well with their notion of application conditionals: since we know that ‘water’ is to be applied to that stuff—whatever it is—that has the property of being the watery stuff, then if the watery stuff turns out to be H<sub>2</sub>O, then ‘water’ picks out H<sub>2</sub>O. However, as we’ve discussed, on their views, we can know ‘water is the watery stuff’ a priori simply by virtue of our stable association between ‘water’ and the property of being the watery stuff, and we can know this independently of knowing how the watery stuff turns out.

Consider again the property that the idealized being associates with ‘water’. This property captures all the ways that water presents itself under all conditions. Note that, in associating *this* property with ‘water’, there is a very important respect in which she knows how the world turns out in the relevant respects (i.e., with respect to the way that water turns out). Note that I’m not making any assumptions of how she came to associate that property with ‘water’: I’m not presupposing that she came to do so through empirical experience. This is a concern even if she came to associate the two through, say, a Matrix-style upload, such that she didn’t have any empirical experience. The point is that, in associating this property with ‘water’, she knows how water turns out in the

---

<sup>90</sup> See their “Conceptual Analysis and Reductive Explanation”, section 3.



relevant respects, and so she cannot associate this property with ‘water’ independently of knowing how water turns out.

This second understanding of the a priori—regarding what can be known independently of knowing how the world turns out in the relevant respect(s)—suggests a slippery slope worry. The epistemic state of an individual who doesn’t know the underlying chemical constitution of water is such that the cognitive significance of ‘water’ for her allows her to know ‘water is the watery stuff a priori’. The epistemic state of the idealized being is such that the cognitive significance of ‘water’ for her doesn’t allow her to know ‘water is the watery stuff’ a priori. This suggests that there is a cut-off point on the continuum below which ‘water is the watery stuff’ is a priori, and above which ‘water is the watery stuff’ is not a priori. How do we determine that point? Relatedly, given that I know more about water than, say, Alice, but less about water than the idealized being suggests that the cognitive significance of ‘water’ for me falls somewhere on the continuum of cognitive significance between that of the former and that of the latter. Does this mean that ‘water is the watery stuff’ is a priori for me or not? This is relevant in light of Jackson’s and Chalmers’s commitment to the claim that *we* can know ‘water is the watery stuff’ a priori. I won’t attempt to resolve this here, but I think it raises some interesting questions about the relationship between cognitive significance and the a priori, and is deserving of future research.

If the aforementioned concerns don’t significantly threaten this relationship between the a priori and cognitive significance, this also allows for us to have (defeasible) a priori

justification for what Jackson and Chalmers call *a priori application conditionals*.<sup>91</sup> These are of the form ‘if x has such-and-such property (that is associated with a given proper name or natural kind term N), then x is (an) N’. For example, one such application conditional might be ‘if x is the predominant clear odorless liquid, then x is water.’ If one uses property G—the property that one associates with N—in place of ‘such-and-such property’, then one has defeasible a priori justification for believing this conditional to be true. Furthermore, when presented with something that has property G, one has defeasible a priori justification for believing that it is (an) N.

In the next section, we will discuss whether, on this notion of a priority—bracketing, for the most part, the concerns about this account of a priority that we’ve raised so far—conceivability is a good guide to possibility.

### ***Section 5: Is Conceivability a Good Guide to Possibility?***

The proposed accounts have consequences for whether conceivability is a good guide to possibility. Since the secondary intension of a sentence determines the truth-conditions for that sentence, and the property that the idealized being associates with an expression and the secondary intension for that expression are co-intensive, then in knowing what property the idealized being associates with an expression, one thereby has access to the truth-conditions for that sentence. If a world in which the truth-conditions of that sentence are satisfied is conceivable, then such a world is possible.

---

<sup>91</sup> See their “Conceptual Analysis and Reductive Explanation”.

This puts the ideal being in an excellent position to use conceivability as a guide to possibility. Since, for each proper name or natural kind term, she associates with it the property that captures how that which is in the extension of that expression would present itself under all possible conditions, she knows how the world must be in order for an assertion in which that expression is a constituent to be true, and thereby has access to the truth-conditions of that assertion.

As non-idealized beings, the rest of us are at a disadvantage here: we don't know which property the idealized being associates with the primary intension, and so we are not in nearly as good of a position to use conceivability as a guide to possibility. However, because the property, G, that a non-idealized individual associates with the proper name or natural kind term is determined by the cognitive significance of that expression for that individual, and the property, F, that the idealized being associates with the expression is determined by the cognitive significance of that expression for the idealized being, and because the cognitive significance of the former approximates the cognitive significance of the latter, conceivability is at least a somewhat reliable guide to possibility. Furthermore, the better the approximation that property G is to property F, perhaps determined by the overlap between the extensions of the two properties at each world, the better guide conceivability is to possibility in that case. Unfortunately, since we aren't in a position, epistemically speaking, to know how good of an approximation that G is to F, we aren't in a position to know how good of a guide conceivability is to possibility in a given case.

## ***Section 6: Concluding Diagnostic Remarks and Looking Forward***

As I have argued, I think that any of proposed extensions that I developed are very useful for modeling how assertions change the state of conversations. Not only can they get the right result in the cases that Stalnaker's, Jackson's and Chalmers's views can accommodate, I think they make the right predictions in the cases that prove problematic for their views.

However, the proposed extensions that I develop do not use the contingent extension intuition in the same way that Stalnaker, Jackson, and Chalmers do. Recall that we discussed two ways of understanding the role that the contingent extension intuition could play in modeling the force of assertions. The first way, used by Stalnaker, Jackson, and Chalmers, has what I called the problematic unifying feature—the same individual or kind is picked out uniformly. The second way, used on the proposed extended versions of their accounts, does not have this problematic unifying feature, and instead picks out all and only those things that are presented in the way the interlocutors expect that which is in the extension to present. Like the first way of understanding the contingent extension intuition, the second way of understanding it also captures the interlocutors' understanding that proper names and natural kind terms refer uniformly. This is because it captures the interlocutors' (perhaps only tacit) belief that if the expression picks out all and only that which is presented in the ways they expect that which is in the extension to present itself, the expression *is* thereby picking out the same individual or kind uniformly. For example, someone who thinks that water only ever presents as a clear

odorless liquid will take it that in picking out all and only clear odorless liquid, ‘water’ thereby *does* uniformly pick out all and only water.

Note that this is a substantial difference in how we are using these intuitions for modeling the force of assertions. In my proposed way of thinking about the intuition for the purposes of the three-dimensional pragmatic reinterpretation, we model the way that this interpretation (or reinterpretation) would go at each world in the context set as follows. At each world in the context set, the extension of the expression would be (re)assigned as that which has the property (or satisfies the description) that the interlocutors associate with the expression (or that it is useful for us to model them as associating with that expression because it is what they know about that which is in the extension of the expression): the expression does not *uniformly* pick out that individual or kind that has that property (or satisfies that description), because it fails to pick out the same individual in circumstances in which that individual fails to have the property, and samples of the kind that fail to have the property.

The way that Stalnaker, Jackson, and Chalmers use the intuition might very well capture how such individuals would (or at least *should*, given their understanding of the semantic or metasemantic facts) take the (re)interpretation to proceed if they could know all the details of how each of the worlds in the context set turn out. For example, if they knew all the details of how the actual world is, including that the stuff in the oceans, lakes, and rivers, etc., is of the same natural kind as that which is in the basin, then they would conclude that what Bruce said, as (re)interpreted at each ice world, is true at each ice

world, and so would not rule out the ice worlds. The point here is that, as used on their views, this intuition is doing *normative* work here: it is telling us that, given that these individuals have this uniform extension intuition (and that they at least tacitly believe that the semantic or metasemantic facts for ‘water’ are such that ‘water’ picks out a clear odorless liquid, and that ‘water’ refers uniformly to the same kind), this is how they *should* take (re)interpretation to proceed at each world in the context set.

I think this is an inappropriate norm to which to hold them accountable for these purposes. It is, as we’ve seen, not obviously true that we *are* able to give a determinate answer regarding whether our expressions have extension at a given world, described in full detail, on the hypothesis that that world is actual—recall, for example, the borderline cases—but, in any case, that is certainly not often, if ever, the situation in which the interlocutors find themselves while engaged in a conversation in which they are trying to narrow down the space of ways this world might be. They *aren’t* given a full description of, or otherwise provided with full information about, the way that this world turns out. They’re working with partial information—they know some things about the way this world is, and not others, and some of the things they believe about the world are mistaken—and they use this partial information in updating the way they take the world to be in light of assertions.

This is why I think that the second way of thinking about the contingent extension intuition is much more appropriate for the purposes of modeling the force of assertions, and why I think that using it is much more successful at making accurate predictions of

the way that an assertion changes the state of a conversation in the cases that proved problematic for Stalnaker, Jackson, and Chalmers. We consider not only what the interlocutors (perhaps only tacitly) take to be the semantic or metasemantic facts for the constituent expressions, but also what they take the worlds in the context set (i.e., the ways this world could turn out to be, for all they know) to be like. They don't have full information about these worlds, and this is an important consideration.

For example, in the ice in the basin example, among the worlds that are in the prior context set are worlds at which there is ice in the basin. This is because it is a way that the world could turn out, for all his students know, prior to Bruce's assertion—in fact, as we set up the example, it is among the hypotheses that they are actively considering. Since they don't have full information about the way this world turns out—including the information about the underlying sameness of chemical constitution of liquid water and ice—they aren't in a position to determine that if there is ice in the basin then there is (frozen) water in the basin. That is, that there is water in the basin under the cold ambient conditions doesn't preclude the possibility that there is ice in the basin. They are, instead, operating with the false belief that ice is of a different natural kind than water. Thus, it is inappropriate to model the force of the assertion in the first way because it holds them to norms to which they could not possibly follow, given their relatively impoverished epistemic state.

In using the second way of thinking about the intuition, we use the property or description that interlocutors associate with the proper name or natural kind term, but we

do not take the expression to apply uniformly. Instead, we take it to apply in such a way *that they judge it to have been applied uniformly*, namely as picking out all and only those things that are presented the way that they expect that which is in the extension of the expression to be presented. As we saw, using the intuition this way, as we do on the accounts that I propose, makes better predictions of the force of assertions than using the intuition the first way.

Unfortunately, in adopting the second way of thinking about the intuition, the accounts that I develop do not have particularly interesting consequences for a priority, or for conceivability being a guide to possibility like those promised by Jackson's and Chalmers's accounts. At most, it gives us an account on which we can have defeasible a priori justification for believing sentences of the form '(the) N is F' to be true, and that conceivability might be a good guide to possibility, though we cannot know how good for a guide since we cannot know how closely the property that we associate with the relevant expression(s) approximates the property that an idealized being would associate with it. One important thing to note is that, on both of the accounts that I develop, that a non-idealized being associating some property G with an expression does not thereby give that non-idealized being access to the truth-conditions of sentences in which it is a constituent. Consider Alice again: she associates the property of being the predominant clear odorless liquid with 'water', and yet she isn't in a position to know the truth-conditions for Bruce's assertion. She *does* have approximating truth-conditions for it, which explains why she took Bruce's assertion to rule out the actual world and other ice worlds, but this led her to *mistakenly* take these worlds to be ruled out.



This suggests that, at least in the sorts of epistemic states that we non-idealized individuals can be in, practically speaking, conceivability cannot be used as a *very* good guide to possibility. Since we will not, in a large number of cases, know *precisely*<sup>92</sup> which proposition was expressed by an assertion, and so we won't know the truth-conditions of it, then we won't know exactly what the world would have to be like in order for those truth-conditions to be satisfied. That is to say, even if we can conceive of a situation in which the truth-conditions would be satisfied, we aren't in a position to *know* that the truth-conditions would be satisfied in that situation, because we aren't in a position to know what the truth-conditions are in the first place.

Precisely *how* good of a guide conceivability is to possibility will depend on how closely one's approximating truth-conditions for a sentence approximate the truth-conditions for the sentence. The more closely one's approximating intension for the constituent expressions is to the primary intension of those expressions, the better one's approximating truth-conditions for the sentence approximate the truth-conditions for that sentence. However, given our epistemic state, we won't be in a position to *know* how good of an approximation our approximating truth-conditions are of the truth-conditions of the sentence (because we don't know *how much we don't know* about that which is in the extension of the constituent expressions—perhaps there are a great many ways that they could present that are unbeknownst to us, or perhaps the property that we associate

---

<sup>92</sup> This is a point of epistemic modesty: even given our best contemporary science, there are still many ways the world could turn out, for all we know. This means, among other things, there are still many ways that water could turn out, for all we know. Even though we know that water is H<sub>2</sub>O, there are still things we don't know about the ultimate constituents of H<sub>2</sub>O, for example.

with the expression captures nearly all the ways they could present, but, either way, we can't know this from the epistemic state that we are in), and so we aren't in a position to know *how* good of a guide a particular instance of at least apparent conceivability of a sentence is to the possibility of that sentence.

## References:

- Block and Stalnaker, "Conceptual Analysis, Dualism, and the Explanatory Gap"  
*Philosophical Review* 108 (1):1-46 (1999).
- Coffa, A. *The Semantic Tradition from Kant to Carnap: To the Vienna Station*.  
Cambridge University Press. (1991).
- Chalmers, D. and Jackson, F. "Conceptual Analysis and Reductive Explanation".  
*Philosophical Review* 110 (3):315-61 (2001).
- Chalmers, D. *Constructing the World*. Oxford University Press. (2012).
- Chalmers, D. "Does Conceivability Entail Possibility?" In Gendler, T. and Hawthorne, J.  
(eds.) *Conceivability and Possibility*. Oxford University Press. (2002)
- Chalmers, D. "On Sense and Intension". *Philosophical Perspectives* 16 (s16):135-82.  
(2002).
- Chalmers, D. "Ontological Antirealism" In D. Chalmers, D. Manley, and R. Wasserman  
(eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford  
University Press, 2009.
- Chalmers, D. "The Components of Content" in David J. Chalmers (ed.), *Philosophy of  
Mind: Classical and Contemporary Readings*. Oxford University Press. (2002).
- Chalmers, D. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford  
University Press. (1996)
- Chalmers, D. "The Foundations of Two-Dimensional Semantics". In Manuel Garcia-  
Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and  
Applications*. Oxford University Press. (2006).
- Chalmers, D. "The Nature of Narrow Content". *Philosophical Issues* 13 (1):46-66.  
(2003).
- Chalmers, D. "The Two-Dimensional Argument Against Materialism". In Brian P.  
McLaughlin & Sven Walter (eds.), *Oxford Handbook to the Philosophy of Mind*.  
Oxford University Press. (2009).
- Chalmers, D. "The Tyranny of the Subjunctive". (unpublished. Accessed on:  
[Http://consc.net/papers/tyranny.html](http://consc.net/papers/tyranny.html))
- Chalmers, D. "Two-Dimensional Semantics", in Lepore, E. and Smith, B., eds., *The  
Oxford Handbook of the Philosophy of Language*. (2006).

- Davies, M. "Reference, Contingency, and the Two-Dimensional Framework". In Manuel Garcia-Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press. (2006).
- Davies, M. and Humberstone, I.L. "Two Notions of Necessity". *Philosophical Studies* 38 (1):1-31. (1980).
- DeRose, K. "Contextualism: An Explanation and Defense", in J. Greco and E. Sosa, ed., *The Blackwell Guide to Epistemology*, Blackwell Publishers, 1999.
- DeRose, K. "Contextualism and Knowledge Attributions", *Philosophy and Phenomenological Research*. Vol. LII., No. 4, Dec 1992.
- Evans, Gareth. "Comment on 'Two Notions of Necessity'". *Philosophical Studies* 118 (1-2): 11-16. (2004).
- Evans, Gareth. "Reference and Contingency". *The Monist* 62 (2):178-213. (1979)
- Garcia-Carpintero, M. and Macia, J., eds., *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press. (2006).
- Gendler, T. and Hawthorne, J. (eds.) *Conceivability and Possibility*. Oxford University Press. (2002)
- Grice, H. "Logic and Conversation", in Cole, P., and J.L. Morgan, eds., *Speech Acts*. New York Academic Press, 1975 p. 41-58.
- Jackson, F. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press. (1998)
- Jackson, F. "Why We Need A-Intensions" *Philosophical Studies* 118 (1-2):257-277. (2004).
- Kaplan, D. "Demonstratives". In Joseph Almog, John Perry & Howard Wettstein (eds.) *Themes from Kaplan*. Oxford University Press. (1977)
- King, J. "What in the World are the Ways a World Might Be?" *Philosophical Studies* (133) (3):443-453. (2007).
- Kripke, Saul. *Naming and Necessity*. Harvard University Press. (1980).
- Peacocke, C. *A Theory of Concepts*. MIT Press: Cambridge. 1992.
- Putnam, "It Ain't Necessarily So" *Journal of Philosophy* 59 (22):658-671. (1962).

- Putman, "The Meaning of 'Meaning'". *Minnesota Studies in the Philosophy of Science* 7: 131-193. (1975).
- Soames, S. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press. (2002).
- Soames, S. "Kripke, the Necessary A Posteriori, and the Two-Dimensional Heresy" In Manuel Garcia-Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press. (2006).
- Soames, S. *Reference and Description: The Case Against Two-Dimensionalism*. Princeton University Press. (2005).
- Stalnaker, R. "Assertion" in *Context and Content*. Oxford University Press. (1999)
- Stalnaker, R. "Assertion Revisited" In Manuel Garcia-Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press. (2006).
- Stalnaker, R. "Conceptual Truth and Metaphysical Necessity" in *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays*. Oxford University Press. (2003).
- Stalnaker, R. *Context and Content*. Oxford University Press. (1999)
- Stalnaker, R. *Inquiry*. Cambridge University Press. (1984)
- Stalnaker, R. "On Considering a Possible World as Actual" in *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays*. Oxford University Press. (2003).
- Stalnaker, R. "On What Possible Worlds Could Not Be" in *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays*. Oxford University Press. (2003).
- Stalnaker, R. "Pragmatics" in *Context and Content*. Oxford University Press. (1999)
- Stalnaker, R. "Pragmatic Presuppositions" in *Context and Content*. Oxford University Press. (1999)
- Stalnaker, R. "Reference and Necessity" in *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays*. Oxford University Press. (2003).
- Stalnaker, R. "Responses." *Philosophical Studies*. 133. 2007.
- Stalnaker, R. *Ways a World Might Be: Metaphysical and Anti-metaphysical Essays*. Oxford University Press. (2003).

Yablo, S. "Coulda, Woulda, Shoulda". In Gendler, T. and Hawthorne, J. (eds.)  
*Conceivability and Possibility*. Oxford University Press. (2002)