2018

Meaning and Modality

Jesse Fitts

Recommended Citation

MEANING AND MODALITY

A Dissertation Presented

by

JESSE J. FITTS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2018

Philosophy

MEANING AND MODALITY

A Dissertation Presented

by

JESSE J. FITTS

Approved as to style and content by:

_____
Phillip Bricker, Chair

_____
Joseph Levine, Member

_____
Christopher J. G. Meacham, Member

_____
Seth Cable, Member

_____
Joseph Levine, Department Chair
Philosophy

DEDICATION

To Rachel and my family

ACKNOWLEDGMENTS

There are many acknowledgements to go around, some of which I'm bound to forget.

First, I thank the chair of my committee, Phil Bricker. Phil read every chapter of this dissertation multiple times. In philosophy you often find yourself in a quagmire of distinctions, multiple iterations of theses, and complicated dialectics. In our discussions, Phil could ask a single question, or make a single comment, that so quickly got to the heart of the issue at hand—a heart that I had been seeking for some time. Phil's mentoring has profoundly impacted me.

Second, I thank the rest of the committee. Chris Meacham mentored me during my first few years of graduate school. He greatly improved chapter 3, which began as a paper for his Bayesianism seminar. Seth Cable taught me formal semantics, and I began thinking about ideas that would turn into chapter 4 in one of his formal semantics seminars. Joe Levine gave me valuable feedback on chapter 1 and provided me with some foundational ideas in philosophy of mind in his seminar on Jerry Fodor.

Third, there are the other philosophers that have mentored me along the way. First and foremost, I thank Ian Dove. If it weren't for Ian, I wouldn't have applied to graduate school in the first place. Ian inspired my first interest in philosophy and continued to mentor me thereafter. James Woodbridge was also responsible for introducing me to contemporary philosophical problems.

Fourth, I thank my mom and dad. My parents are intellectual in their hearts and souls. In some other possible world, they would have perhaps both been PhD students. In this world, they made this one possible.

Last, I thank my wife, Rachel. She's read every paper I've written, discussed every single idea I've ever had, improving each and every one, and has been with me every step of this long, long journey.

ABSTRACT

MEANING AND MODALITY

FEBRUARY 2018

JESSE J. FITTS

B.A., UNIVERSITY OF NEVADA LAS VEGAS

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Phillip Bricker

I intended to write four papers whose topics faintly concerned separate issues in meaning and modality. As it turned out, chapters 1-3 all roughly concern the same topic: propositions. While I argue for two different theses in chapters 1 and 2, I try to understand the changing propositions literature in both. In addition to arguing for the respective theses in chapters 1 and 2, accounting for this change is a parallel goal for the chapters taken together. Chapter 3 examines particular propositional roles—the objects of the attitudes and the objects of credence. Finally, chapter 4 changes the subject to the second conjunct in the title—modality, specifically of the epistemic kind.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ICONIC PROPOSITIONS

Abstract

I motivate the need for, and then sketch, an *iconic* theory of propositions according to which propositions are *like* or *similar* to their objects of representation. Propositions on this theory are properties that the mind instantiates when it simulates the world. I connect the theory to recent developments in the propositions literature as well as to a strain of cognitive science that explains some kinds of mental representation in terms of iconicity. To fill out my theory, I compare it to prominent theories in the contemporary literature.

1.1   Introduction

I explore the following argument in this paper:

1. Mental representation is fundamental and propositional representation derivative— what I will call the *mind-first* view.

2. If the mind-first view is true, then the nature and existence of propositions depends, in some way to be determined, on how the mind, in fact, represents.

3. The mind represents, *to some extent*, not *linguistically* but *iconically*.

4. Therefore, the nature and existence of *some* propositions depends on iconic representation.

I say "explore" rather than "defend" for the following reasons. Premise 1 I provisionally accept for the sake of this paper because the mind-first view is significant enough to explore its consequences, but I won't defend the mind-first view itself. Premise 2 is a trivial fact about the mind-first view. I want to be most explicit

about the status of premise 3. The literature on iconic representation, from multiple sources, is vast, and writers have explored iconic representation in many forms— mental models, mental maps, mental imagery, to name a few. It only matters to me that such views, in some shape, are plausible, for I only want to explore the consequences of combining the mind-first view with iconic representation *supposing* premise 3 is true. Given this supposition, I bracket any objections to iconic representation itself. For the bulk of the paper, then, I explore the consequences of the conclusion. Combining the mind-first view and iconic mental representation raises an important question for the propositions theorist: How, metaphysically, ought we understand propositions on such a combination? I answer that iconic propositions are a kind of property—one that both the mind and the world instantiate when agents successfully simulate the world. In other words, when the mind represents iconically, its content represents because, at some level of abstraction, it's *similar* to its object of representation.

The structure of this paper will follow the above argument. §1.2 covers premises 1 and 2. This consists in contrasting the mind-first view with its opposite, the *proposition first* view, which takes propositional representation as fundamental and mental representation as derivative. In explicating the mind-first view, I review its two main proponents: Jeff King and Scott Soames. In this section I also consider the issue of pluralism about propositions. If the existence and nature of propositions depend on the mental processes that result in representation, and if the mind represents multi-modally, then the possibility of pluralism arises. §1.3 considers cognitive scientific views of iconic representation. §1.4 presents my theory sketch of iconic propositions, compares it with King's and Soames' theories, and evaluates the theory on various *desiderata* in the current propositions literature.

1.2   Mind and Proposition First

In this section I contrast the traditional approach to propositions, proposition first, with the new approach, mind-first. I want to flag that I'm not primarily interested in the history of propositions—whether the timeline of old/new matches up perfectly to proposition first/mind first. What matters is that we have two broadly different approaches and that, at least currently, the focus is on mind-first style approaches.

1.2.1   General character of each thesis

Schiffer (2003) captures how philosophers have traditionally understood propositions. According to this view, propositions are abstract objects that have truth conditions essentially and absolutely—i.e., respectively, it's necessarily true that a proposition has its particular truth conditions, and a proposition has those truth conditions without relativization. Contrast that essentiality and absoluteness with sentences: It's contingently true that sentences have their particular truth conditions, and they have their truth conditions only relative to a given language with its various conventions. Though Schiffer doesn't use this language, traditionally, propositions represent, or have their truth conditions, intrinsically—i.e., they represent the world purely in virtue of the way the proposition is. Any duplicate of a proposition will represent like the original copy. Last, propositions are mind- and language-independent in two senses. First, their existence doesn't depend on humans or agents of any kind. Second, multiple languages can express the same proposition, but any one proposition doesn't belong to a given language.

The important aspect of this traditional picture is the mind-independent bit. The recent propositions literature has taken an empirical turn in which the nature and existence of propositions is in some way mind dependent. To start, we can understand this turn to mind independence in terms of whether propositional

representation is *derivative* or *fundamental*. Some things represent because we confer that power upon those things. The sentences on this page represent in virtue of various conventions set up regarding written language. Thus we humans partially conferred representational properties on written language. But what about our human representational capacities? Are we the ultimate wellspring of representation, or do our mental states represent derivatively—because they have as their contents abstract objects—viz. propositions—which themselves are the ultimate wellsprings of representation? Note that this question is importantly distinct from the question of whether propositional representation is intrinsic or extrinsic. The source of a property and whether an object has that property intrinsically are different issues.

The mind-first view rejects the fundamentality of propositional representation. The two mind-first representatives that I discuss below, King and Soames, see this rejection as perhaps *the* key innovation of their views. It's unclear, historically, whether philosophers really held that propositions were the ultimate source of representation or whether the issue simply wasn't investigated. Whether philosophers such as King and Soames are bucking tradition in this respect or exploring a before-now under-explored issue is unimportant. And while the question of the source of representation is important, it's too large to explore in this paper. The issue for us is, *supposing* minds, not propositions, ultimately represent, how does that affect, if at all, our theories of propositions? Since I will be referring back to the mind- and proposition-first views repeatedly, let's set off and define each:

Proposition-First (P1) The thesis that propositions exist, and have their nature, independently of minds, and propositions are the ultimate source of representation while mental representation is derivative.

Mind-First (M1) The thesis that propositions depend for the existence and nature on the mind, and minds are the ultimate source of representation while propositional representation is derivative.

Both theories will agree on aspects of Schiffer's characterization of propositions. According to both, propositions will be abstract objects. Both views will have some way of understanding representation, truth conditions, or some general notion of propositions characterizing the world,[1] and that world characterization will be in virtue of an intrinsic feature of propositions. The main disagreement between P1 and M1 will be in terms of the priority of the propositional and the mental.

1.2.2   P1 theories and their benefits

The usual way of dividing up theories of propositions is in terms of grain— whether propositional identity cuts finer than truth conditions—and structure— whether propositions contain semantically significant constituents. P1 theories come in different combinations, and most of these theories are familiar, so I will be brief. The main unstructured, coarse-grained theory is the possible worlds theory, according to which the proposition that p is the set of worlds where p is true (or a function that pairs p-worlds with truth and non-p worlds with falsity). According to the tuple theory, the main fine-grained, structured theory, propositions are structured entities comprising some kind of entities depending on one's semantic commitments. The basic idea is that the proposition that Rachel reads comprises an entity having to do with Rachel and something to do with the property of reading, and the proposition is structured in some way related to the syntax of the sentence "Rachel reads," something like the following:

$$\langle R, r \rangle \,.$$

---

[1]I'm being deliberately vague about this aspect because of the variety of theories in the literature. Some theorists think propositions must *have* or *determine* truth conditions while other theorists think propositions *are* truth conditions. Yet other theorists think that propositions don't represent at all—which may sound like heresy—but still in some way can determine truth conditions—this is Jeff Speaks' view. Any P1 or M1 theory will have some way of characterizing various states of affairs in a way that is independent of those states of affairs obtaining.

If one is a Russellian, then Rachel, the person, and the property of being a reader comprise the proposition. For a Fregean, the relevant senses do. The other combination in the literature is fine-grained and unstructured.[2] These are theories such as Schiffer's in which propositions are *sui generis* entities that lack structure and that have primitive, fine-grained representational capacities.

All such theories are *supposed* to adhere to Schiffer's general definition. Sets of worlds (or characteristic functions thereof), object- or sense-containing tuples, and *sui generis* propositions are all abstract. *If* each theory succeeds, propositions are or determine truth conditions that are held intrinsically, absolutely, and essentially.[3] And what's more, sets, functions, and tuples don't depend on humans for their existence or their representational capacities, and *sui generis* theories usually stipulate that their propositions likewise are mind independent.

We should say why one may be attracted to P1 for reasons other than age and historical inertia. Perhaps the greatest virtue of P1 theories is their ability to make sense of some of the metaphysical roles of propositions. First, propositions bear alethic properties, such as contingent and necessary truth. There were truths (and falsities) before any beings existed, and there will be truths once we're gone. Whether there are truths and falsehoods doesn't seem to depend on whether there are beings with representational powers. If there are truths before beings, we need something to bear those alethic properties, and P1 theories can provide such propositions.

Relatedly, there are not just infinitely many truths, but *continuum* many. If this is the case, then we need that many propositions. As Armstrong (2004, p.9), citing Stephen Read, says, "[f]or all being, there is a proposition (perhaps one

---

[2]I'm unfamiliar with any coarse-grained–structured theories. I don't think such a combination is possible.

[3]I say "if" because, as we'll see shortly, there are good reasons to doubt that such propositions represent intrinsically.

never formulated by any mind at any time) that truly renders the existence and nature of this being."[4] But if the existence and nature of propositions depends on beings with representational powers, we straightaway face a problem: a dearth of propositions. The representational capacities of any being are in some way limited, and if propositions depend for their existence on such beings, then propositional existence will in some way be limited. As we'll soon see with King's theory, one may employ various recursive devices to generate infinitely many propositions—in the way that a grammar can generate an infinity of well-formed sentences. Be that as it may, recursive devices will generate at most denumerable infinities, with the result being that there are truths with nothing to be true.[5]

The last benefit of some P1 theories is their ability to understand logical relations.[6] Propositions are supposed to be the entities that bear logical relations to one another—relations of entailment, logical strength, etc. We can understand these relations elegantly with world or tuple theories. Take the worlds theory, which understands logical relations in terms of set relations and operations. A proposition p is true on the worlds theory just in case the actual world is a member of p; a proposition p entails q just in case p is a subset of q; and so on.

Pluralism—whether one class of entities plays every propositional role or not—complicates how important P1's benefits are (and whether it's a major cost to M1 should they fail on such issues).[7] If pluralism is true, then it may not be a cost

---

[4]This quote isn't a direct rejection of the mind-first program. The mind-first program needs to account for propositions that no being has actually entertained. They do this, as we'll see, by appealing to recursive devices.

[5]For a specific objection along these lines to Soames' theory, see Keller (2017).

[6]It's unclear how *sui generis* theories understand such relations unless they appeal to set theoretic devices as models, which I discuss in chapter 2.

[7]This area is tricky because there is a pluralism of pluralisms. The kind I'm discussing here is pluralism about roles—that we may need different kinds of content for different kinds of propositional roles, which naturally split into metaphysical, semantic, and epistemic roles. This is a different issue

to the M1 theorist if she can't account for the objects of logical relations and the objects that bear alethic properties before there are representing beings. We should, however, primarily address the issue of having the right amount of contents to serve as the contents of any given (iconic) content-bearing mental state. And we should understand the objects that enter into logical relations in the sense of the objects that enter into the reasoning of representing agents.

1.2.3 M1 theories and their benefits

My explication of M1 will be more in depth given M1's heterodoxy. I start with the two main examples of M1 theories from King and Soames. I then say something about how we should understand M1 theories in general before listing the problems that M1 theories are meant to solve. To close this section, I argue how an acceptance of M1 can motivate the view of propositions of this paper.

Mind dependency is the key to an M1 theory, but different theories do this in different ways. The general M1 strategy is to identify propositions with abstract objects whose existence is the result of mental processes that are responsible for mental representation. The main M1 theories come first from Jeff King and then from Scott Soames. King's propositions are based on syntactic and semantic mental processes: The members of our lexicon bear semantic relations to their semantic values, and lexical objects also bear syntactic relations to each other. Humans put together this information to calculate content—a process that is familiar from formal semantics in which, for a given sentence, one composes semantic values at each non-terminal node in a syntactic tree to reach a proposition at the top for a declarative sentence. The result of these calcuations is an abstract object—a fact—that King identifies with propositions. Soames' theory is similar but employs mental events

---

than the pluralism that arises in the M1 approach in which, if beings represent multimodally, we'll have a pluralism of proposition types for each mode.

rather than facts based on formal semantics—though it's unclear how different King's and Soames' theories are. I will now quickly explicate King's and Soames' theories in more detail, but if the reader is familiar with the theories, the reader can skip to §1.2.3.3.

1.2.3.1 Example 1: King's theory

Propositions, for King, are facts, and a fact, according to King (2007, p.26), is an object possessing a property or n objects standing in an n-place relation. This might sound striking: A fact is typically taken to be the part of reality that true propositions represent, and rather than being true or false, facts *obtain* or they do not obtain. The kind of fact that is the proposition that the following express—

(1.1) Rachel reads

—must obtain regardless of whether the fact that Rachel reads obtains or not, though the fact contains the components of the proposition: Rachel and the property of reading.[8] If the fact that Rachel reads obtains, then the fact that is the proposition that Rachel reads will be true, and it will be false otherwise. So what kind of fact is a proposition?

Consider the LF representation, in tree form, of (1):

(1.2)

<p style="text-align:center">Rachel   reads</p>

A syntactic relation, which the tree represents, obtains between "Rachel" and "reads"—call it R. While syntax relates the lexical items "Rachel" and "reads" to each other, both lexical items stand in semantic relations—they have Rachel, the person Rachel (the person, on one view of the meaning of proper names), and the property of being a reader, respectively, as their semantic values. We can represent being a semantic value with asterisks:

---

[8]Though the proposition contains more.

(1.3)

```
        ⌒
  Rachel   reads
    |        |
 Rachel*  reads*
```
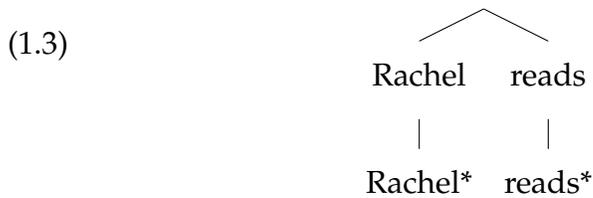
We're building propositions—i.e., meanings independent of any particular language—so we existentially generalize over the specific lexical items, represented with stars:

(1.4)

```
        ⌒
   ★       ★
   |       |
 Rachel*  reads*
```

Now what is it about this fact that gives it its truth conditions and thus gives us good reason to believe that it is the proposition that Rachel reads? According to King, nothing at this point. For, a syntactic relation could have different semantic significance in different languages. Another way of stating this is that the syntax provides instructions to the semantics regarding how and whether semantic values compose. It's in principle possible that a syntactic relation could instruct the semantics to combine Rachel and the property of being a reader and return truth just in case Rachel does *not* have the property of being a reader. Thus, we need to make sure that the syntax deploys the *instantiation function* in order to get the truth conditions: The syntax instructs the semantics to interpret syntactic concatenation as the instantiation function. Leaving out some details dealing with indexicals, we're in a position to picture and describe King's full account:

(1.5)

```
              I
              |
           ⌒
    ★        ★
    |        |
 Rachel*   reads*
```
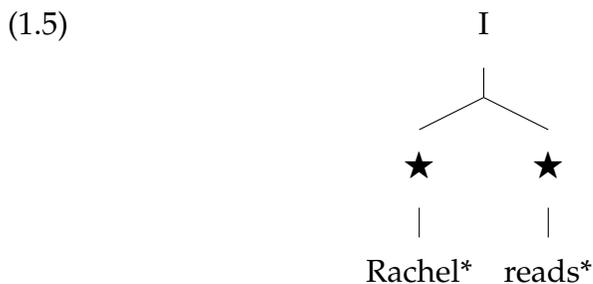
I represents that *R* encodes the instantiation function. Here is the proposition that (5) expresses, which will serve as a statement of King's account applied to a single

proposition: There are lexical items x and y of some language L occurring at the left and right terminal nodes, respectively, of the sentential relation R that in L encodes the instantiation function, where the semantic value of x is Rachel and the semantic value of y is the property of being a reader. What's important for us is that propositions come to exist on King's account because humans interpret R as I. In doing so, we brought into existence the facts that King identifies with propositions.

1.2.3.2   Example 2: Soames' theory

According to Soames, propositions are certain kinds of *event types*. According to this account, to entertain a proposition is to engage in a cognitive activity—the mental act of predication, which for Soames is a primitive mental act, basic among others. When an agent sees o as red, there is an event token of her predicating redness to the object o. Now of course this event token can't be the proposition that o is red. Instead, Soames takes the cognitive event *type* of predicating redness to o to be the proposition that o is red expresses. And more complex propositions and differing propositional attitudes are built upon the mental event act of predication. Regarding other attitudes, some examples: To judge that o is red is to predicate redness to o and then to endorse that predication; to believe that o is red is to predicate redness to o and to be disposed to judge that o is red (and so to be disposed to endorse the predication of redness to o). Regarding more complex propositions, an example: To entertain the proposition that o isn't red is first to predicate redness to o and then to predicate not being true to the result of that first predication.

1.2.3.3   General characterization of M1

Now let's see if we can draw out the distinctive features that King and Soames share that qualify them as mind-first theorists. Toward that end, let's introduce a distinction, between the *basis* of content and content itself, that will help to capture

the mind-first program in general.[9] The basis of content is whatever mechanism is responsible for some certain kind of content. In humans and other animals, the brain and its activities are the mechanism that underlie content. To put it differently, if some entity has a function, then there must be a mechanism that carries out that function. The function of the kidneys is to filter blood, so there is a mechanism that carries out this function: many small filter cells compose the kidney, each of which has a mechanism that captures waste and sends non-waste back to the blood stream. One of the brain's functions is to characterize the world, providing information, and thus there is a mechanism that carries out that function. To discharge this function the brain's activity (somehow) creates an inner representation, a vehicle for content, that is a representation because it expresses some proposition. The brain, e.g., may represent via inner language-like entities. Or, as will be relevant for us later, the brain may represent not linguistically but via iconic simulations of the world. To have an attitude in a proposition will consist in the following, where r is a content vehicle, S is an agent, Att is a propositional attitude, and p is a proposition:

(1.6) $\exists r(S \; Att \; r \; \& \; r \; expresses \; p)$

For S to have an attitude whose object is p is to bear a relation, functionally defined as a given attitude, to an inner content vehicle whose content is p. For the language of thought theory, the content vehicle will be a sentence of mentalease. For a simulation theory, the content vehicle will be some inner simulation.

Here is an impressionistic way to understand M1: The M1 theorist pays greater attention to r and especially the process that gives rise to r when crafting a theory of p in the second conjunct. Conversely, a proposition-first theorist can craft a theory of p while ignoring some facts about the mechanism by which the brain carries out its representational function. What this means in practice is that the M1 theorist

---

[9]See McGinn (1989, p.182) for the distinction.

first considers *how* the mind represents—through some series of actions that result in a mental representation. Then, the M1 theorist will extract abstract objects from those series of actions. If the M1 theorist doesn't have good reasons to show that the mind in fact represents the way she claims, then we have no reason to accept the existence of the abstract objects, i.e. her propositions. The details of these mental processes constrain one's account of propositions. For example, with King, since his propositions depend on a mechanism that individuates content according to syntactic difference, difference of syntax, for King, entails difference of proposition. Yet this is *prima facie* unintuitive since a single content that can stand over minor differences in the details that gave rise to that content was one reason we wanted propositions in the first place.

The propositions-first theorist doesn't face such a dependency. Stalnaker (1987, p.22) makes the point here:

> The pragmatic picture and the possible worlds definition of proposition does not then deny that beliefs are internally represented. But it remains neutral on the form that those representations must take. It should be emphasized that the possible worlds conception of proposition does not present an account of the form in which propositions are represented which is a rival to the linguistic account. Accepting this account of propositions does not, for example, commit one to a psychological hypothesis that our minds represent a space of possible worlds point by point, with individual representations of individual worlds. The aim of the definition is to give an account of the structure of what is represented while leaving open questions about the means by which this is accomplished.

Thus, there may be some relation between the traditional properties of grain and structure that both the content vehicle and its associated proposition share. But the proposition-first theorist can safely ignore the underlying process that gives rise to mental states that have such properties. In fact, the possible worlds theorist can remain neutral regarding the mechanism by which the mind represents, as Stalnaker (1987, p.22) notes here:

In attributing beliefs and desires, we are attributing certain kinds of internal causal properties which have the structure that tends to reflect the world in ways that make it appropriate to call them representations. These representations could conceivably take the form of sentences of a language of thought written in the belief center of the brain, but they also could take the form of pictures, maps, charts, or graphs, or (most plausibly) a diversity of redundant forms, none of which are very much like any of the forms which our public representation take.

Now that we have an understanding of M1, let's list its virtues. The driving idea behind these virtues is that M1 theories give propositional representation a kind of independent reality that doesn't depend on the philosopher for interpretation. Because these propositions and their representative powers have an independent reality, M1 theories have the prospect of solving the following issues that P1 theories have difficulty with.

1.2.3.4 Benefits of M1

The Benacerraf problem. Since chapter 2 covers this problem exclusively, I'll be brief. A Benacerraf problem can arise when a philosopher reduces or identifies one class of entities to another. In the classic mathematical version, the problem targets the reduction of numbers to sets. The problem proceeds in the following steps. First, the objector that raises a Benacerraf problem locates multiple, non-equivalent reduction/identification bases. By non-equivalent here I mean simply that the reduction bases are different entities. Second, the objector argues the candidate reduction bases are equally explanatorily virtuous, so that choosing one base over another would be arbitrary. The final step concludes that the reduction/identification fails.

The problem straightforwardly applies to set-theoretic theories of propositions. For example, a set of worlds and its characteristic function are, metaphysically speaking, different things, but both work equally well satisfying various propositional roles by the world theorist's own lights. So, for a given proposition p, it would be arbitrary to identify it with either the set of p worlds or the function that

pairs *p* worlds with truth and non-*p* worlds with falsity. Therefore, *p* is neither the set nor the function.

In chapter 2, I will argue that P1 theories, such as sets of worlds or tuples, in my view, are best seen not as propositions themselves but *indices* of propositions. A *theorist* employs an index—typically a set of abstract objects—to individuate aspects of reality, and the index is theoretically fruitful when it shares relevant structure with the object of its index, and when there is a function that pairs members of the index with states of reality. For example, we use real numbers to index temperatures because states of reality that we want to individuate in terms of heat share a relevant similar structure with real numbers, and there is a function that pairs the states of reality with the numbers. In this case, there is no risk of confusing heat itself with the numbers used to measure that heat. But it's my hunch that we risk conflating an index with its object when the object of the index is abstract. Such is the case with sets of worlds or tuples, which, in my view, are best thought of as such indices: They are logical devices that index, among other things, abstract objects that are themselves the content of mental states. But if we suppose that traditional propositions are instead actual propositions and not indices thereof, we run into trouble.

Let's move to the benefits of M1. The first is the prospect of solving the Benacerraf problem. For the mind-first theorist, the mechanism that gives rise to content fixes a unique set of abstract objects that serves as that content. For example, King claims we have good reason to believe that the brain in fact composes semantic values in the way that syntacticians and semanticists tell us, giving rise to the abstract objects that he identifies with propositions. The mind could have composed semantic values in different ways, giving rise to different abstract objects, but the mind will really work some way or other. But, if the empirical results about the mechanism are correct, then those facts pick out a unique set of abstract objects. For

Soames, minds engage in a series of actions that result in a representation of the world. While the mind may have engaged in some other series of actions, it in fact engaged in some specific series of actions, which determines a unique event token that is a token of a unique event type. And it will be these mental actions that non-arbitrarily demarcate the propositional facts/events from the non-propositional facts/events. The ability to solve these arbitrariness problems, which is no small feat, is tied to the next issue.

Another arbitrariness worry: The demarcation problem. There is a newer, less discussed arbitrariness worry in the propositions literature:[10] Non-*sui generis* propositions are members of a class of entities that contain non-propositional members. We've seen the M1 theorists identify propositions with either facts or events, yet there are many facts and events that are not propositions. The world and tuple theorists have identified propositions with sets of worlds or object-containing tuples, respectively. But there are lots of sets, functions, and tuples that aren't propositions. We may ask any propositions theorists, what is it about the entities that you identify as propositions that makes them different from the non-propositional entities? Call this, specifically not having an answer to this question, the *demarcation problem*.

The demarcation problem seems to apply to some P1 theories. Why does the set of white-snow worlds represent that snow is white while the set of even numbers doesn't represent anything? Why does the tuple that contains Rachel and the property of being a reader represent that Rachel reads but a tuple containing a dog and a cat doesn't represent anything? No answer seems forthcoming. The second benefit for the M1 theorist, then, is the prospect of avoiding this problem. King's facts are different from other facts because they contain a relation that encodes

---

[10]I first found this worry in King (2007, ch1)

instantiation, and Soames' events are different because they involve mental actions that are responsible for creating content.

The source and nature of propositional representation. There are a few important, related issues regarding propositional representation. A successful theory of propositions will make sense of how propositions characterize the world in a way that doesn't involve outside interpretation. We need some way to make sense of some basic propositional *desiderata*, such as being able to say the same thing as someone else, believing the same thing as someone else, etc.—the notion of shared content. Two sentences say the same thing if they have the same content; two people believe the same thing if they each have beliefs that share the same content; and so on. When two people believe the same thing, it is not in virtue of us, from the outside, interpreting their mental states to have the same content. Rather, there's something about their mental states, or two sentences that say the same thing in different languages, such that it makes sense to say they share something in common. But if propositions require our interpretation to represent, then they're no different from sentences: We would have to attribute content to them. But then what is the nature of that content and what is the source of its representational powers?

The problem with P1 theories is that their representational powers seem to come from the outside. Take, e.g., the function version of the worlds theory. Why choose truth values over a function from worlds to the set $\{1, 0\}$? Or, to make a point that Soames has made, why not a function from worlds to a dog and a cat? Put this way, the claim that traditional propositions represent intrinsically seems absurd on its face. The only reason various set-theoretic structures, both structured and unstructured, serve equally well for various purposes, is because, to represent, they require outside interpretation. If representation requires interpretation, propositions no longer intrinsically represent and thus are no different from sentences. But if that's true, then we no longer have a way of making sense of shared content.

The third benefit of M1 is the promise to solve this problem. The mind-first theorist combines intrinsic and derivative representation. The source of a proposition's representative powers derives from the activity of the mind, but such activity endows propositions with structure and properties that make a plausible case that propositions intrinsically represent. Let's return to King and Soames, yet again. For King, humans at some point interpreted syntax to have a particular semantic significance. This interpretation brought into existence abstract objects whose syntax has that semantic significance as an intrinsic property, which endows King's propositions with their truth conditions and representational powers. Thus, although the representational capacity of King's propositions derive from humans' representational powers, the property of representing is intrinsic to King's propositions. What's more, there is something about King's propositions *in and of themselves and absent our interpretation* that explains how they represent the world. For Soames, humans engage in a variety of mental actions among which predication is fundamental: When I think that Rachel reads, I predicate the property of being a reader to the person Rachel. Such an event is a token of a cognitive event type that Soames identifies with propositions. The nature of the representational property of Soames's propositions is the same as King: intrinsic yet derivative. The propositions come to represent as a result of our mental lives, but these propositions have the intrinsic property of representing because there is an intrinsic connection between a token's and type's representing. Like King, there is something about Soames's propositions, in and of themselves and absent outside interpretation, that accounts for their representational capacity.

Propositional unity. I would be remiss to leave out the *problem of propositional unity*, which looms large in the literature. This is a problem for structured theories

in particular—theories of propositions with parts.[11] On the one hand, consider a list of two objects and a relation—Desdemona, the loving relation, and Cassio—and on the other, the proposition that Desdemona loves Cassio. What's the difference? What it is about the latter that explains its unity, so that it is not a mere collection of individuals and a relation? The only structured P1 theory, the tuple theory, doesn't seem to have a deep answer to this problem. The difference between Desdemona, the loving relation, and Cassio, on the one hand, and $\langle Loves, \langle Desdemona, Cassio \rangle \rangle$ seems to simply be two sets of brackets.

While we generally accept the force of the Benacerraf problem, and while we want to account for the nature and source of propositional representation, the force of this issue is less clear to me.[12] The tuple theorist's answer may just be what we just said: the brackets provide structure, and a tuple is a unified, single thing while a list is, well, an un-unified list of entities. But the problem may not be simply what unifies the proposition but rather two deeper problems: First, why *that* unity rather than another—why $\langle Loves, \langle Desdemona, Cassio \rangle \rangle$ rather than $\langle \langle Desdemona, Cassio \rangle , Loves \rangle$? And second, what is it about those brackets that confers propositional properties on that tuple? But these two questions point back to two previous problems that M1 theories promise to solve—Benacerraf and the source of representation. While it's unclear exactly the force of the problem of unity, M1 theories have an account of unity: The special glue that holds propositional constituents together is mental activity, and that activity pins down a reduction base the members of which have important propositional properties intrinsically.

---

[11]We can refine the target of the problem more since, after all, a possible world, in some sense, is a part of a set of possible worlds, and possible worlds theories are usually thought of as the paradigmatic unstructured propositions. Rather, a structured proposition has parts, and these parts are *semantically significant* and, usually, something like syntax structures these semantically significant parts.

[12]Soames (2010) makes this point.

1.2.4 Motivating iconic propositions

Why be interested in an iconic theory of propositions? If we accept the general M1 approach, as we provisionally are for this paper, then we commit to propositions that depend on some etiology that is responsible in the end for some kind of human representation. This is what the two major contemporary theories of propositions, King's and Soames', do. King claims that humans create content via semantic and syntactic operations, and Soames claims that humans create content via mental actions, and they both disagree about the ontological category of the resulting proposition—fact or event.

We've noted that if the M1 theorist is wrong about the etiology that is responsible for their content, then we can call into question her theory of propositions. If mainstream formal semantics is seriously mistaken in the way that we calculate semantic content, then King's theory may be in trouble. But even if the M1 theorist is right about the ground-level processes that are responsible for the content on which she bases her theory, it's important to see the scope of such a theory. An M1 theory will be a theory of *that* kind of content—the kind that a given mental process is responsible for. For example, for King, it's most plausible that his theory is a theory of *linguistic* content. The empirical processes he appeals to, after all, attempt to explain how humans come to calculate linguistic meaning. But there are multiple kinds of content-bearing mental states, and it's not obvious that a single processes will be responsible for each kind of content. For example, if one thinks that perceptual states are content-bearing, it's not clear that the same mental processes that are responsible for calculating linguistic content are also responsible for producing perceptual content. In presenting his theory, Soames often uses perceptual examples: For example, Soames (2016) says, "When I perceive or think of o as red, I predicate redness of o, which is to represent o as red." Such mental

events seem plausible for calculating linguistic content, but I see no reason to accept, *prima facie*, that such mental actions occur at the perceptual level.

King recognizes this point. King (2007, ch3) wrestled with the worry of the origins of his propositions: to explain how our ancestors first brought Kingean content into the world. To explain, e.g., how lexical items came to have semantic values, it seems we would have to appeal to our ancestors' beliefs and desires, so propositions would need to exist in the first place. In his book, King sketches a quick story about how our pre-linguistic ancestors had "proto" intentional states before the dawn of propositions. Later, King (2014, p.60) says a bit more about these proto-intentional states, that they may have been perceptual states that had their own kind of content, which is different from linguistic content. King (2014, fn.19) thus recognizes that the etiology that produces his content should limit the scope of his theory: His book, as he says, should have been titled *The Nature and Structure of Linguistic Content* instead of just content. He then goes on to say the following:

> I believe that many things have content other than sentences of natural languages. Maps, diagrams, perhaps pictures and, most importantly for present purposes, perceptual experiences have contents. In the case of each sort of thing that has content, there will be an account of those contents in the spirit of the present account of the contents of natural language sentences. Due to limitations of space, time and knowledge, the details of the theory of the contents of perceptual experiences that is in the spirit of the present account of contents of natural language sentences cannot be sketched here.

This is exactly what I'm beginning to do in this paper: I'm trying to give an account, in the spirit of King's, of a certain kind of mental content that we have some reason to believe may not be linguistic in nature.

To close this section, I note that M1 theories have a built in limitation: Such theories provide content for a certain kind of mental state. It may turn out that, as an empirical fact, that all content arises from the same process, but I think it's likely that we'll find that humans represent multi-modally. The takeaway from

this discussion is that M1 theories, depending on how the empirical facts end up, may result in a kind of pluralism: we may need an M1 theory for linguistic content which may not, as King recognizes, carry over to other kinds of content. Pluralism about propositions is woefully under-discussed. If pluralism has been raised at all, it has been in terms of propositional *roles* that may conflict. I suggest that if M1 is true, a pluralism of a different kind arises—a pluralism of etiologies that give rise to different kinds of contents and so different kinds of propositions.

For the rest of the paper, I want to stress the limited scope of an M1 theory given this pluralism, and this scope limits the force of certain objections. For example, since King is giving a theory of a certain kind of content, it's not enough to object that his content can't play some role that the objector is interested in. In what follows, I will attempt to provide content for certain kinds of mental states, and objections that such content can't play some different roles may be irrelevant.

1.3   Varieties of Iconic Representation

The topic of iconic representation is vast, and the following is but a mere sample. I give a small survey and then try to establish some important properties of iconic representation that will qualify it as a basis of content suitable for a theory of propositions. The idea that the mind represents iconically is not by any means new. Aristotle, in *De Anima*, defends the view that the mind becomes like its object of representation in a way that's appropriate for the mind. As Shields (2016) puts it:

> Aristotle is reasonably understood as holding that *S* thinks some object of thought *O* whenever *S*'s mind is made like that object by representing salient structural features of *O* by being directly isomorphic with them, without, that is, by simply realizing the form of *O* in the way *O* does.

Aquinas endorses and builds upon Aristotle's view, and in the early modern era, Hume, under a certain interpretation, has an iconic view.[13] F.P. Ramsey is the modern source of something like the view in his adage that "belief is a map. . . by which we steer." Let's quickly motivate, from different directions, the idea that the mind represents in this way.

1.3.1 Perceptual content

Let's start small, with a particular kind of iconic content in the perception literature, specifically the conceptual vs. non-conceptual content debate. Those in favor of non-conceptual content claim that the contents of some of our perceptual experiences don't depend on our conceptual recourses (I'll still see this or that shade of green regardless of whether I possess that fine-grained concept). According to one sub-view on the non-conceptual side, some perceptual content is iconic and, in virtue of that, non-conceptual. In the philosophy literature, Fodor (2007) brought this view to light. Fodor distinguishes between two different kinds of representations: discursive representations, which involve concepts, and iconic representations, which don't. For Fodor, both kinds of representations are compositional, meaning that the syntactic structure and semantic content of the parts of a representation determine the syntactic structure and semantic content of the whole representation. While discursive representations satisfy compositionality in the usual way, Fodor offers what he calls the "picture principle" to account for the compositionality of icons:

Picture Principle If P is a picture of X, then parts of P are pictures of parts of X.

What distinguishes discursive from iconic representations, according to Fodor, is that iconic representations don't canonically decompose while discursive represen-

---

[13]See Jacobson (2013) for historical views of iconic representation.

tations do. What this means is that discursive representations, being language like, contain *constituents* in the sense that the parts of the representation decompose in a unique way. For example, in the sentence John loves Mary "loves Mary" is a part but "John loves" isn't. With icons, any grouping of its parts still represents some part of its object of representation. What's important for us is that, as Fodor notes, icons are *semantically* evaluable—meaning that they are icons *of* some object. Yet, for Fodor, iconic representations play a limited role. They occur only at the beginning of perceptual cognition, and their role, unsurprisingly, is sub-personal. Iconic representations are stored briefly as an intermediary between being a representation on the surface of the retina to being a representation in short- or long-term memory. Let's now consider wider views of iconic content, specifically mental maps and models.

1.3.2   Mental Maps

We often employ non-linguistic representation to navigate the world. To get a feel for this, for example, when I make my way around the library, I consult an inner, map-like representation of the library. This map isn't linguistic at all—it doesn't say, in sentences, that these books are on this floor, next to these books, and so on. Maps and pictures come to mind when we think of non-linguistic representation, but we categorize any kind of representation in which isomorphism plays a significant role as a mental map. The isomorphism involved in pictures may involve physical similarity while the isomorphism involved in a diagram, say a Venn diagram, may be highly abstract. This contrasts with linguistic representation in which isomorphism plays no role in explaining representational capacity. The case of perception presents a kind of iconic representation that is, if Fodor and others are right, a sub-personal, very technical sense of representation. The content of such perceptual states doesn't seem to enter into rational relations with other iconic

states, though presumably they can somehow enter into relations with conceptual content. The contents of mental maps, on the other hand, enter into rational relations. Just think of updating the mental map of the library from the example in light of experience: I knew where the philosophy books were in the library, but I thought the logic books were toward the west side of the floor but are in fact toward the middle. The next time I am in the library I will consult the updated mental map to find the logic books, and the process of updating the mental map is a *rational* process.

Heck (2007) proposes mental maps as one among other examples of non-conceptual content, which humans employ in cognition, along with conceptual content. On the other hand, there are at least two prominent sets of defenders of the view that the mind represents in a map-like way in general. First, Braddon-Mitchell and Jackson (2007, ch.10) offer mental maps as an alternative to the language of thought, and so seem to claim that thought in general occurs in a map-like medium, and Armstrong (1973) presents a similarly holistic view of map-like mental representation.

1.3.3   Mental models

The mental maps literature in philosophy shares a likeness to the large, and more empirically informed, mental models literature in cognitive psychology, beginning with Craik (1967). The following quote encapsulates the mental models view:

> If the organism carries a "small-scale model" of external reality and of its possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and the future, and in every way to react in a much fuller. safer, and more competent manner to the emergencies which face it.

Like the mental map theory, according to the mental models perspective, the mind represents in an iconic way, creating a simulation of reality. Our creation and

manipulation of mental models occurs both sub-personally and personally, as the theory is usually coupled with a dual process theory of reasoning wherein the mind reasons both unconsciously and automatically, on the one hand, and consciously and deliberately on the other. What's more, humans reason by manipulating these models—i.e. running simulations on them—without having to manipulate external reality directly.

This reasoning occurs in different mediums. For example, humans seem to make visual inferences on mental models. According to one experiment, participants were asked to decide whether two images of objects were the same objects. Cubes that appear glued together at right angles composed the objects. Subjects seem to rotate a three-dimensional model of one onto the other to decide if it's the same object, and the larger the size of the angle of rotation, the longer a decision tended to take. It's important that the manipulation took place in three dimensions—not just the spinning of a two-dimensional representation of a three dimensional object. The model in this case is a three-dimensional model from which a two dimensional representation is derived. Thus, if we were to understand the mental map of the library earlier in this way, we wouldn't have an egocentric, two-dimensional internal picture of the library. We would construct such a picture *from* a mental model, which would in this case be a three-dimensional array, similar to the way a computer employs a three-dimensional array, analogous here to a mental model, to project a two dimensional picture on a screen.

1.3.4 Computationalism

Finally, iconic representations have been couched in terms of the computational theory of cognition (CTC) according to which we reason by manipulating symbols according to rational rules. The CTC is a variegated theory *type*. The theories it houses all agree that cognition consists in symbol manipulation. On the CTC,

propositions are, unsurprisingly, the objects of propositional attitudes: An agent has an attitude A to a proposition p in virtue of standing in a relation to an inner symbolic representation whose content is p and whose functional role determines which attitude A is—belief, hope, etc. On the standard interpretation of the CTC, symbols bear their content, i.e. have the semantics they do, in a way that is similar to the way language bears its content.

Iconic representations in the CTC provide an alternative to the language-like representations of the standard interpretations. Such representations have been employed in AI implementations of both human and non-human (viz. insect) cognition. In philosophy, Cummins (1989) gives an account of *s-representations*—"s" for "simulation." Defenders of iconic representations in the CTC claim that the representational capacity of the symbols of the CTC are required to understand our inferences that the CTC underwrites. To use an example from Ramsey (2007, pp.81-85), a cognitive system may use something akin to a family-tree graph to represent familial relations and work out a relative relation of some node on the graph. If a cognitive system uses a simulation that the symbols participate in in order to perform some task, then we can't understand the success of a given task absent representational features of the simulation. Ramsey asks us to imagine a Chinese Room-like thought experiment in which someone works out familial relations yet is semantically ignorant of the input-output representations. In addition, the person is ignorant about the labels as well as the meanings of the tree connections. In other words, the person employs syntax alone to pair the correct input representations (questions) and output representations (answers). But unlike the Chinese Room, we can't understand this person's *success* without her operating on a simulation that has an isomorphic pattern of relations among its constituents as reality does. As Ramsey puts it, this person's success can't be explained by appealing to the fact that she focuses on shapes alone and follows shape-involving instructions. The

graph must represent family relations in order to understand this person's success; likewise, we must treat a mental simulation as representing.

1.3.5 Iconic content as a viable base

To motivate iconic propositions, an iconic basis of content needs to satisfy certain properties. Unfortunately, it's easy to muddle some of these issues, so we must be careful here. We discussed the issue of structure in section §1.2.2 in the context of P1 theories. In the propositions literature, *structure* refers to whether propositions contain semantically significant parts. The underlying issue in the debate around structure in the propositions literature concerns the grain issue. If we have, e.g., two syntactically distinct sentences that intuitively express different propositions, but those sentences have the same truth conditions, we can appeal to the constituents that are associated with sub-sentential items to distinguish the propositions. What's especially important is that we can appeal to a difference of constituents to distinguish various necessary truths, which all have the same truth conditions and thus all express the same proposition on unstructured, coarse-grained theories.

Structure differs slightly in the context of mental representations—as a way to make sense of two key *desiderata* for a theory of mental representation: productivity and systematicity. A system of thought is productive if there is, in principle, no upper bound to the construction of novel thoughts and is systematic if there are systematic connections between the kinds of thoughts thinkers can think. For example, if one can think "John loves Jane" then one can think "Jane loves John" or "John loves. . ." for any object that you fill in the eclipses with. Philosophers in this area argue that we must appeal to the structure of thoughts to explain how they are combinatorial in a way that satisfies productivity and systematicity. It's not clear to what extent the notions of structure differ in the propositions and

mental representations literature, but let's just say this: The propositions literature tends to focus on whether propositions comprise constituents that correspond to sub-sentential entities, and the mental representation literature tends to focus on whether a system of thought is combinatorial in the right way, which requires the system to have parts, but those parts don't necessarily need to be sub-sentential.

Productivity and systematicity are easy to come by if thinking occurs in a language-like system—a language of thought—with a structure similar to first-order logic, with the mechanisms of connectives, quantification, predication, etc. While LOT theorists often speak as if their thesis requires thought to occur in a medium like first-order logic, the official definitions given by Fodor only require that thought proceed in a medium that has a syntactic structure and compositional semantics.[14] What I want to note here is that various philosophers have argued that certain iconic content is structured, systematic, and productive, which is my main concern and that, as such, that content qualifies as a kind of language of thought in a weak sense.

Let's see what notion of structure may be relevant to iconic content. Braddon-Mitchell and Jackson (2007, p.178) claim that a content is structured if the similarities and differences of content are systematically related to the similarities and differences of the objects of the content. If we have representations $R_1, R_2, ...R_n$ that correspond to states of the world $W_1, W_2, ...W_n$, differences in the Ws will correspond to differences in Rs. The level at which we find this correspondence may be highly abstract. To use an example from Braddon-Michell and Jackson, Arabic numerals are structured representations the objects of which are numbers since there is a systematic correspondence between the similarities and differences of the

---

[14]I found a distinction between strong and weak sentientiality that I'm working from here in Rescorla (2009, p.397)—strong sentientiality requiring first-order logic like structure, and weak requiring only syntax and combinatorial structure. Resorla notes that the official definition of mentalease from Fodor (1987, pp.134–8) only requires the weak sense. Also see Camp (2007, p.152).

Arabic numerals and their number objects. The idea when applied to iconic contents is that there will be systematic differences and similarities in iconic representations that correspond to similarities and differences among their objects.

Braddon-Mitchell and Jackson (2007, p.181–182), Camp (2007, pp.152–154), and Rescorla (2009, §7) all argue, in different ways, that iconic representation can proceed in a medium that is combinatorial, systematic, and productive. The specifics of the arguments for those properties will depend on the details of the iconic representational system, and here I merely want to mention how this might work. Braddon-Mitchell and Jackson and Camp both focus on maps in the ordinary sense. Both argue that such maps contain reoccurring, combinable elements. As Camp (2007, p.154) puts it, mental maps are made up of "recurrent formal elements that make a common semantic contribution each time they occur." There is no upper limit to recombining these elements—satisfying productivity. What's more, if one can entertain one combination of locations and relations among them, one can entertain those elements in a nearby configuration, satisfying systematicity. The above two philosophers have ordinary maps as their paradigm, but there are many other formats that iconic content can take. For example, Rescorla (2009) notes, according to one account of a kind of iconic representation, cognitive systems employ vectors that participate in computational models that interpret those vectors in terms of Euclidean distance metrics. Since these vectors comprise coordinates, an organism that can entertain a mental map vector composed of a certain set of coordinates can also entertain another mental map comprising the same coordinates swapping one out for another, satisfying systematicity.

1.3.6 Which content is a good base?

Mental representations—understood broadly as any semantically evaluable mental state—play different roles, and not every role is suited to underwrite a

theory of propositions. We've seen iconic representation implicated at a technical, sub-personal level, and, on the other side, we've seen that mental models, when coupled with a dual-aspect view of cognition, operate on a personal level, being available to consciousness. It is common to distinguish three levels of explanation regarding the mind: the hardware (neural level of the brain), software (abstract structures that the hardware system realizes to engage in cognition), and the personal level (consciousness experience, beliefs, desires, etc.).[15] The basis of content for a mind-first theory of propositions will be the software level—the level at which the mind represents that is common enough among humans to ground a theory of propositions. This is clearly where King grounds his theory, and Soames is less clear on this issue, though his theory would make the most sense if his mental actions took place at the software level.[16] For the purposes of this paper, any software level iconic content, such as mental maps or models, will serve as a basis for iconic propositions.

1.4 Iconic Propositions

We've motivated the need for iconic propositions and have laid out some ground-level details. For the sketch of a theory of iconic propositions, there will be some key pieces and some decision points. The first key piece of the sketch is to understand iconicity in terms of a string of ideas: We account for iconicity in terms of similarity, similarity in terms of isomorphism, and isomorphism in terms of properties. This is the most important part of the account. Next, we'll need some way of ensuring that iconic contents are both intrinsically representational and determinate. We face a decision point in how to solve these problems, and I'll take the theory in one

---

[15]See McGinn (1989, p.185).

[16]For example, Soames often uses perceptual examples—when we see that a ball is red we predicate redness to a ball. Presumably we don't consciously perform such actions.

direction—accounting for them in terms of teleology, though I'm not ruling out other ways. Last, we face another decision point—whether holistic or atomistic representation takes priority. I take the theory sketch in the holism direction, again not ruling out the alternative. And before moving on, I want to stress again the message of §1.2.4: I'm not giving a univocalist sketch of propositions. Rather, operating under an M1 and pluralistic framework, I'm giving a sketch of one kind of proposition among others.

1.4.1 Iconicity's core: Similarity and isomorphism

Iconicity ties together the disparate kinds of content that I've surveyed, but what is iconicity? I've glossed iconicity in terms of similarity, but I have something more precise than similarity in mind—in particular, isomorphism. Generally, an isomorphism is a function in which relational structure is preserved. Swoyer (1991, p.452), which focuses on what he calls "surrogative reasoning," captures the notion I'm after thus:

> [T]he pattern of relations among the constituents of the represented phenomenon is mirrored by the pattern of relations among the constituents of the representation itself. And because the arrangement of things in the representation are like shadows cast by the things they portray, we can encode information about the original situation as information about the representation. Much of this information is preserved in inferences about the constituents of the representation, so it can be transformed back into information about the original situation.

Isomorphism is the first major part of the theory I state below, and officially, I implicate isomorphism generally in the theory. What I mean by this is that there are different *orders* of isomorphism, and, at this point, I don't commit to any order or other being *the* correct order of isomorphism. I say this because first, cognitive scientists are still actively investigating this question, and second, different types of orders may capture different modes of iconic representation. I should note that various authors have chosen second-order isomorphism as their preferred order.

The kind of isomorphism that Swoyer describes above is second order. The idea is that, while first-order isomorphism will typically capture the sense in which a representation literally resembles its object, second-order isomorphism will capture similarity in abstract relational structure.[17]

The pattern of relations contained within an iconic representation such as a mental map will be highly complex, but in order to state the theory, let's zero in on the complex property of having *that* pattern of relations. Entertaining a mental map, then, will consist in the mind/brain instantiating this complex property, and successful representation will consist in the *co*-instantiation of properties— the mind's of a pattern of relations and the world's instantiation of the complex property of the mirrored pattern of relations. And it is this mirroring of relational patterns that partially underwrites the usefulness of these representations—because it allows us to deploy, to use Swoyer's phrase, surrogative reasoning, in which one reasons with a structural representation to draw inferences about the object of that representation.

The first piece of sketching the theory, then, is that when the mind represents the world, it instantiates a complex property—the property of having a certain pattern of relations—and when the mind represents veridically, the world instantiates a complex property—the property of the mirrored pattern of relations. If we set aside for a moment the niceties of the context of propositions and the order of isomorphisms and such, the general idea is that a map comprises a certain pattern of relations—it will be some way or other. And when the map is useful, the world will be the same way, under isomorphism. And the idea stays the same when the iconic representation becomes more abstract—say, with a Venn diagram. Isomorphism is only one part of the sketch because similarity, resemblance, isomorphism—all

---

[17]For cognitive scientific literature on the second-order view, see Shepard and Chipman (1970) and Choe (2002).

the basic concepts under which we explain iconicity—are all too easy to come by. Anything is isomorphic to anything else *under some interpretation*. A map doesn't represent merely because it's isomorphic to some aspect of reality. It represents because it has the right kind of isomorphism to make it useful to us. It is our *use* to which we put the map *plus* its pattern of relations' being isomorphic to some aspect of reality that explains how the map comes to represent reality. And as we've seen, part of the problem with proposition-first theories is that they require interpretation. Surely it cannot be "we" who interpret our own inner iconic representation on pain of an infinite regress of inner homunculi.

1.4.2   Interpretation and securing unique content

Avoiding such a regress isn't a problem alone for a theory that employs iconic representations, and there are various strategies to avoid it. Here is one way, employing teleology, that we may avoid the problem.

The strategy to avoid the problem for iconic representations is to show that such representations can be run in a cognitive system that lacks any kind of interpretative intelligence, yet the use to which the icons are put, in addition to isomorphism, underwrites the icons' representative powers. And this is the second piece of the foregoing sketch—that iconic propositions represent, to borrow a phrase from McGinn (1989), partially in virtue of *natural teleology*. This idea is developed by writers both in[18] and outside[19] the mental models and maps literature. E.g., Ramsey (2007, §6.1) illustrates the mindless strategy with a two-step analogy. The first step involves an intelligent agent's interpretation conferring representative powers upon an icon; the second step involves removing the agent while keeping the guiding function of the icon, the result being that its representational capacity still

---

[18]See Cummins (1996).

[19]For example, see the work of Ruth Millikan, such as Millikan (2002).

remains. The analogy compares the iconic view with another prominent view of representation—the "receptor view"—in which neural states represent because they reliably detect certain external conditions. To illustrate:

Step one. Two windowless cars, car A and B, respectively, traverse an "S"-shaped path. They do so as follows:

- *Icon*. A driver guides the car with a map of the "S" combined with dead reckoning.

- *Receptor*. When the car approaches a wall, sensors on the front bumper push inward, closing a circuit, which illuminates a light, which the driver uses to guide the car through path.

Step two. Now we remove the drivers, and the cars traverse the path as follows:

- *Icon*. The map is replaced with an "S"-shaped groove, isomorphic to the track, in which a rudder runs. The rudder moves along the groove and controls the steering wheel via arms connected to the rudder, guiding the car through the path.

- *Receptor*. When the car approaches a wall, sensors on the front bumper push inward, closing a circuit, which controls a rod that pushes the steering wheel in the opposite direction, which guides the driverless car through path.

Once we remove the driver, the most natural way to interpret the process in the *receptor* car is that a mere causal process guides the car. There is no component of the process—the sensors, the circuit, the rod—that seems representational. On the other hand, the best way to explain the success of car A in step two is that the groove is a map, a representation, of the track. At some much, much more complicated level, the iconic representations that guide our behavior are analogous to the groove, and the conclusion that Ramsey draws, that I want to employ, is that

35

our cognitive systems can mindlessly use iconic representations to target aspects of reality, and in so doing, the icons acquire their representational powers.

Natural teleology can solve another problem for the iconic content—that such content is indeterminate. This is perhaps one of the most obvious worries for any theory of representation based on isomorphism or similarity. The worry is that a single iconic representation can target any object—as Fodor puts it (tweaked for our purposes) one model may represent a chess game or the Six Day War if, at some level of abstraction, the two targets are functionally identical. The source of this objection comes from thinking about everyday models as well as scientific models. We're familiar with, at least before the days of Google Maps, scrawling a map of one's neighborhood that could serve to represent any number of things given the right purpose. But everyday and scientific models, maps, etc., require interpretation from us—they only extrinsically represent. However, *if* we pair such iconic representations with a use to which they are put, then we do get a unique content. We wouldn't be impressed if, in the midst of discussing directions via my hand-drawn map of my neighborhood, someone objected that my map represents Napoleon's strategy in the Battle of Austerlitz. Natural teleology performs a similar purpose except that the "interpretation"—or the use to which it's put to—isn't extrinsic, and this use can pin down what an iconic representation represents.

This determination of content also resolves another immediate worry for the theory: The representation relation should be typically asymmetric, for the objects of representations don't represent their representers. Language-like representation obviously satisfy this *desideratta*. I've grounded the representational capacity of my propositions partially in terms of a kind of similarity, in particular a kind of isomorphism. Yet similarity is a symmetric relation. If this is the case, then I face an objection: If iconic proposition p represents r, then r will also represent p, which seems strange at the very least. The solution is this: Iconic contents and their targets

may be symmetrically isomorphic, but the targets of such content aren't put to any use and so don't represent anything.

1.4.3   Holism and atomism

The last decision point is a priority issue. Iconic contents may not encode the entirety of an agent's belief state at some time. Still, such contents can encode lots of information. Yet at the very least, to give a sketch of a theory of propositions, we need individualized contents. Let's first consider, at an intuitive level, how we may extract such contents. If the content of an agent's mental state with respect to some issue at some particular time is a complex, holistic map, then an individualized content will be some part of that map. Just as a map may represent some expanse of land, some portion of that map represents a part of that land. More abstractly, a Venn diagram may represent various properties and relations between sets as a whole and particular relations and properties when we consider an aspect of the diagram—that, e.g., the intersection of two sets is non-empty. If we can understand a mental map as a complex property of having a certain pattern of relations—one that a cognitive system uses to target, via isomorphism, some aspect of reality—then an individualized content will be some portion of that mental map. In other words, it will still be a property of having a certain pattern of relations, but the pattern of relations will be some portion of the agent's holistic content.

How do we isolate the right parts that intuitively correspond to individualized contents? This process, of extracting individualized contents from holistic content, isn't new, especially for possible worlds theorists such as David Lewis.[20] And there is, in fact, a close connection between possible worlds and iconic representation.[21]

---

[20]Lewis (1986, pp.32-34) gives four options to extract individualized content.

[21]Mental models theorists, Johnson-Laird in particular, have used possible worlds to precisify mental models.

There are various options to precisify the holistic iconic content with possible worlds from which we can derive individualized content. Here is a typical strategy: Braddon-Mitchell and Jackson (2007, pp.190-191) give a functionalist pairing of sets of possible worlds with belief states the content of which are mental maps. Their view is this: each belief state, when paired with an agent's desires, causes actions. These actions would satisfy an agent's desires in some worlds but not others. Those *some worlds* are the content of that agent's belief state. Or we consider what possibilities an agent's mental maps rules in, rules out, and leave indeterminate. The set complement of the set ruled out will be an agent's doxastically accessible worlds. Some of the supersets of these worlds will correspond to what we intuitively think about as individualized contents. The Braddon-Michel and Jackson method is (but) one way to pick out the right supersets.

Let's revisit the library example. The relevant iconic content may be a highly complex, albeit incomplete, map of the library. Now one aspect of this content is that I represent the world in such a way that the logic books at the university library are near the middle of the twentieth floor. The way that we would extract this content from my holistic map according to the above strategy is that given my desire to check out a logic book (for example), if the aspect of my map regarding the logic books is accurate, then it would successfully steer my behavior to satisfy my desire. In terms of possible worlds, one superset of my doxastically accessible worlds will have in common the logic books being near the center of the university library on the twentieth floor.[22] Again, we're considering one possible way of moving from holistic contents to individualized contents, and for the sake of stating the theory in the next section, let's say that there is a *content pairing* between a holistic content

---

[22]This set of worlds isn't the individualized content itself. Rather, it's a *theorist's* tool for precisifying, out of my holistic content, a single aspect of it that corresponds to what we might specify with a belief sentence.

and an individualized content when in fact there is such a pairing on a process like the one I just described.

1.4.4 Statement of theory

To recap: According to the mind-first trend, the fundamental source of representation is the mind upon which all others forms of representation, including propositions, are derivative, though this doesn't impugn the intrinsicality of propositional representation. This view promises to solve some recalcitrant problems, such as locating a plausible fundamental source of representation, solving the Benacerraf problem, and solving the problem of propositional unity. The mind-first trend, then, requires us to pay mind to the details of mental representation. Toward this end, we've considered a diverse view according to which the mind represents partially iconically. Iconic representations represent not only in virtue of isomorphism, but also, according to the option we took, in virtue of a cognitive system's natural teleology—a kind of use that doesn't require inner homunculi—that delivers a determinate content for the system. I've understood an isomorphism between mind and the world as the sharing of a pattern of relations, and I've considered the property of having some particular pattern of relations as one's iconic content as a way of specifying that content. We've then considered aspects of that highly complex property that are an individualized content, and we've sketched different ways to precisify that individualized content. With this in hand, we can now give a statement of the sketch that we'll refine and expand upon:

Iconic Propositions (IP) An iconic representation c (a mental map, model, diagram, etc.) is a complex pattern of relations that form an isomorph that a cognitive system uses to navigate the world. The content of c is the property i of having the pattern of relations involved in the isomorph. An iconic proposition p is a property r that isolates an aspect of i for which there is a content pairing with

i, and the mind represents the world veridically in terms of *p* when both the mind and world instantiate *r*.

To explicate the theory, let's see how various propositional issues play out according to it.

1.4.5 Basic propositional issues

I first consider how the sketch deals with basic issues, which will set us up to discuss the more complex, contemporary issues from §1.2.3.4.

General content. Can iconic propositions represent anything general? If they couldn't, I wouldn't take that as a damning objection. We would still have a software-level process that leads to non-general content for which we should still account. Unfortunately the general–non-general distinction is somewhat vexed since ways of understanding the distinction, say in terms of the singular–general proposition distinction, are themselves a matter of debate. Fortunately, I think there's a clear sense in which at least some iconic content can represent general facts, and the focus on concrete-style maps may lead one to think otherwise. For example, diagrammatic content clearly represent general facts—that all dogs are mammals, which isn't about any particular dog or mammal.

Misrepresentation. There is no representation without misrepresentation, and there is some disagreement about how misrepresentation works among those that write about iconic representation.[23] Misrepresentation comes about from a disconnect between a representation and its target. The disconnect on the traditional picture is between what an externally existing object describes and how the world is. For IP, the disconnect comes between the content that isomorphism and natural teleology fix and how the world is. If, according to the foregoing sketch, we

---

[23]See Cummins (1996) for a detailed view.

understand veridical representation in terms of the co-instantiation of properties—the mind and the world both instantiating the property of having a certain pattern of relations—then it's natural to understand misrepresentation as the asymmetric instantiation of properties. But the world instantiates all kinds of properties that the mind doesn't instantiate, and in many cases we don't want to say that we're misrepresenting the world. Misrepresentation, for IP, is a disconnect between the fixed content of an icon plus natural teleology and how world is—a disconnect between an "intended" target and the actual target. This idea should be familiar on the analogy with everyday iconic content. We may intend to map a certain part of the neighborhood for a party but misrepresent the streets, or intend to capture family relations with a diagram and misrepresent those relations, etc.

Non-existant relata. A necessary condition on the satisfaction of many relations is that the *relata* of that relation exist. If the loving relation obtains between Abelard and Heloise, then both need to exist. Not so for representation. Agents can represent the nonexistent: E.g., I can represent Pegasus, though such a being is nonexistent. It's important to note that accounting for non-existent entities, properties, etc., implicated in a *relatum* of the representation is a general problem. Here is how IP makes sense of this: The mind instantiates a property that the relevant chunk of reality does not instantiate. The reason that a given simulation represents some non-reality is because that's the use the cognitive system puts the model to.

Shareability of content. The way that IP accounts for the shareability of content is perhaps one of the most interesting aspects of the theory in my view. There is a sense which any number of people can say, believe, doubt, etc., the same thing. And two sentences of the same or different languages can mean the same thing. Indeed, we commonly give examples of shared content to introduce propositions to the uninitiated. Above, in the midst of stating the theory, we considered how the propositions of IP come to have determinate content rather than representing

anything given the right purpose. Considering the key *desiderata* of shared content will give me a chance to elaborate on the determinacy issue.

It appears to me that some philosophers take the notion of shared content quite literally. When two people believe the same thing, the content of the states is *literally one and the same abstract object*—the proposition. Similarly for natural language sentences. This is the case for tuple theories as well as possible worlds theories: You and I both believe that snow is white in virtue of our mental states having literally one and the same proposition as content, whether it be $\langle W, s \rangle$ or the set of snow-is-white worlds, etc. Some modern writers hold to this literal picture, for example Schiffer (2003). Traditionally, then, to share content is to glom onto one and the same object.

It's not clear to me, since Soames and King don't address these issues directly, how we should understand their vision of shared content. It seems to me that Soames' picture must be different, for when you and I believe the same thing it is in virtue of *doing* the same thing on his picture: When you and I both believe that Rachel reads, we both predicate the property of being a reader to Rachel. Though we engage in separate event tokens, both of our tokens are of the same predicative event type. This mutual participation in the same event type is supposed to make sense of shared content.[24]

IP diverges from the traditional picture of shared content. My notion of shared content consists in a notion of sharing that we're familiar with—that of properties. Various minds will process information in different ways, but, when they represent the world as being one way or another, they do so in virtue of an agents' iconic contents representing it as such. What these minds share is that, at some level of

---

[24]It's unclear how to understand King on this issue. I would want to know more about how King understands the nature of the syntax in his theory, an issue I raise in the next chapter. If his syntax is idealized psychology, then I think we should understand his notion of shared content in a way similar to Soames'.

abstraction, they're all *similar*—both with each other and with the object of representation. And they're all similar because they all instantiate the same property—a sample (or attempted sample) of the world. So when agents share content, it's not in virtue of accessing the same object but in virtue of being relevantly similar to one another.

One may worry that iconic representation is a poor basis for mind-first propositions because it just seems that mental maps, models, and others kinds of iconic representation are just too variegated from person to person to ground any notion of shared content. I think there is something to this worry, but I think it is a general worry for the mind-first program, which I want to quickly address and note that I consider this issue in more detail in chapter 2.

For Soames, and possibly for King, when we look closely at their respective bases for their propositions, minor, seemingly trivial, details in individual representation may seem to either impugn the shareability of content or push them in a direction in which traditional problems, namely the Benacerraf problem, resurface. Let me briefly say what I mean and why I don't think these problems are genuine.

For Soames, we may consider how events are individuated. It seems plausible that if we take a series of actions, and rearrange the order of those actions, then the result is two event tokens that don't fall under the same type.[25] But if this is the case, then trivial differences in event order will result in different propositions when what we want is shared content. For example, suppose you and I are watching a subtitled foreign news broadcast, and the following line flashes on the screen:

(1.7) Fighting in Aleppo intensified, and Turkey entered the Syrian conflict today.

---

[25]Soames doesn't discuss whether he's committed to this individuation of events, but the individuation conditions seem extremely plausible to me. If I go to the store and then go to the bar, and you go to the same bar and then the store, we seem to engage in separate events.

I read the sentence normally, from left to right, and someone blocks your view so that you read the conjuncts out of order. We both come to believe the same thing, and although actions involved in either of our event tokens were the same, they were performed in different orders, resulting in different event types and thus different propositions. Soames' account, because of this temporal aspect, is especially vulnerable to the worry underway—that minor differences in the basis of mind-first propositions will lead to distinct content when we want shared content. For example, the series of actions in processing ⌜If p, q⌝ may be different from ⌜q if p⌝ though we would plausibly want shared content here.

For King, we don't have this worry since there is no temporal aspect to the order in which semantic values, given some syntactic structure, are composed. Yet for King, much depends on exactly what the syntactician and semanticist are up to at a metaphysical level. While there is a strain of semanticists that see themselves as describing a Platonic grammar, it's my hunch that many semanticists and syntacticians see themselves, if only implicitly, as modeling psychological processes at some level of abstraction. And for King, the Platonic option is out anyway if he is to be a mind-first theorist. Given all this, a kind of dilemma arises. If we stick to the mind-first program to the letter, then trivial differences in bases threaten the notion of shared content. If we depart from the mind-first program and base propositions on models of psychological processes, then a Benacerraf problem of a different kind arises. There may no be fact of the matter—at a metaphysical, not epistemological level—whether one syntactic/semantic structure or a closely related one is *the* syntactic/semantic structure.

These are complicated issues, but my point in bringing them up is that pinning down a notion of shared content presents a general kind of difficulty for the mind-first theorist, and IP doesn't face a distinctly threatening version of it. At some level of detail, individual iconic representations will surely differ from brain to brain.

But if a crowd of people have mental maps that all represent the logic books as on the twentieth floor of the library, then there will be some property based on that representation that all of their minds instantiate, and they will share content in virtue of that. If anything, IP fares especially well with the dilemma currently under discussion because the notion of sharing is the familiar notion from the sharing of properties. Many red objects can differ is innumerable ways, and may even differ in their exact shade of redness, yet they may all share the property of being red. Similarly for iconic representation: many minds may vary in exactly how they represent some given fact, but they can still all share a property in common.

1.4.6 How IP fares with P1's benefits

In §1.2.2 we noted that the P1 theorist's main benefit is that she provides *enough* propositions: The P1 theorist can provide a proposition for every possible content-bearing mental state. The P1 theorist can also make sense of the objects that agents entertain when they engage in logical inference. These are the main benefits that I want to consider. We also noted that the P1 theorist can make sense of truths and falsehoods before the existence of representing beings.

Let's start with the issue of the number of propositions. Recall the problem: Both King and Soames, as M1 theorists, employ bases of content from actual human psychology. In order to generate propositions that no one has thought of yet (or will think of), they both appeal to recursive devices familiar from syntax, and such devices won't generate enough propositions. While actual human agents employ the cognitive models of IP, the theory can provide properties to ground the representation of any iconic mental state. This is because the properties of my theory aren't human-centric, and isomorphism is an account of iconicity generally. This being so, these properties could underwrite the representation of any being that simulates the world.

The main issue for M1 theories is a dearth of propositions, but this ignores the other aspect of a parity between being and representation—that of *too much* representation so that we end up with "propositions" that shouldn't be propositions. To my knowledge, no one has discussed this issue, but, *prima facie*, King has the problem. Any syntactically well-formed sentence with lexical items that bear semantic relations to their semantic values, for King, will express a proposition. If this is the case, then so called category mistake, non-sense sentences will express propositions: e.g., "Colorless green ideas sleep furiously." The defectiveness of such sentences, for some, goes beyond mere falsehood: Such sentences are *meaningless*. And if we analyze meaninglessness as the failure to express a proposition, then King would have propositions where he shouldn't. There is a lot to say about this issue, but I want to note that IP doesn't seem to have the too-many-propositions problem: properties come prepackaged with instantiation conditions that prevent non-sense from being some way the world could be.

As to logical relations, the idea of truth or satisfaction conditions starts from the idea of simulating a fact that obtains. Now this might not fit the mold of traditional truth conditions, but I also think this is OK: We need some understanding of what it means for the content of someone's mental state to be true, and IP gives us that in terms of property co-instantiation. And we can begin to understand the logical relations that exist among mental contents in terms of properties; one can't simulate something's being maroon without simulating its being red, one can't simulate a square's having three sides, etc., which means that certain kinds of properties instantiate together, some can't co-instantiate, etc. And given that much of the iconic representation literature, mental models in particular, was introduced to account for reasoning, much of that literature can be worked into an account of the logical relations that hold among mental contents.

As to the issue of truths before the existence of representing beings: For every way that a cognitive being might simulate the world, I can provide a property to ground such simulation, though before there were agents, there were no cognitive systems to employ such simulations. On this worry, I have nothing new to add—I would point to King (2007, ch3), in which King distinguishes between *there being no truths* and *the world not being a determinate way*. Without representing agents, the former will be true, but not the latter: the world will still be some way—i.e., it will instantiate various properties—but there will be no representers to represent this world. Presumably, things would be similar for Soames, but with event types. For King, Soames, and I, there will be no propositions before agents and no propositions at worlds lacking agents. For me, this means that the world was some way without there being a representing agent to simulate the way the world was at that time.

In the end the metaphysical role category, however, in my view, is best understood at the level of propositional models—for me, as noted above, in terms of sets of possible worlds. Here, I mean "model" in the sense of a scientific model—an idealized and theoretically fruitful representation.[26] But this isn't a demerit of IP alone—that I can't account for the metaphysical roles at the level of my propositions and not at the level of their model. I think it holds for both King and Soames as well. For any Kingian or Soamesian propositions, it's not clear whether some given logical relation holds between them. King and Soames thus will employ models, and that's just the situation that I am in: IP goes hand in hand with a worlds model that elegantly analyzes logical relations in terms of set-theoretic relations.

1.4.7 How IP fares with M1's benefits

Recall from §1.2.3.4 that M1 theories promise to solve the Benacerraf problem as well as a the related arbitrariness worry that I called the demarcation problem.

---

[26]I discuss this issue in depth in the next chapter.

They also promise to explain the ultimate source of representation and to explain the unity of the proposition, although we noted that that benefit may not be as important as the other issues.

First, the Benacerraf problem. The Benacerraf problem arises, to recall, when we have multiple propositional candidates to choose among the choice of which is arbitrary. This problem is most salient when we have multiple formal objects with the right propositional properties, such as grain and structure. This is the classic propositional version of the problem that plagues set theoretic theories of propositions. IP identifies propositions with properties, not formal devices. The decision to employ such formal devices is what leads to the surfeit of choices. If IP had a Benacerraf problem, it would arise from indeterminacy of content—if we had no way of choosing which of any iconic content may represent some way the world is because any icon can represent anything with the right interpretation. If this were the case, then to choose one icon over another would be arbitrary. But we provided one way of pinning down unique content, which would prevent a Benacerraf problem of this kind.

We answer the other demarcation problem with the same way that we pin down a unique content, which for us was natural teleology. Sets of objects and tuples don't, in general, represent—so why do the traditional propositions? There is seemingly no answer. King and Soames, on the other hand, do seem to have an answer: Facts and event types don't, in general, represent, so why does the subset that are propositions represent? Because, for King, humans interpreted syntax to have a particular semantic significance, and for Soames, because those particular mental events are composed of tokens that involve mental actions. For me, properties aren't in general proposition like, but the properties of IP *are* the way the mind is when it simulates the world.

IP is able to face these two arbitrariness worries because the ultimate source of representation, as a mind-first theory, is the mind. It is the mind's activity that creates simulations of the world. We know something about how this happens, as we've seen in §1.3, but it is still, probably, mostly a mystery—a subject for further inquiry. From §1.4.2, I sketched arguments from Ramsey (2007) that iconic contents represent intrinsically. And as King oughtn't be committed to the fine details of the final true syntax and semantics, I shouldn't be committed to the fine details of iconic representation. The important part is that the *broad* outline of the program, the very idea that the mind represents partially iconically gives me the materials to extract abstract objects whose representational capacity depends on the creation of simulations, which, together with natural teleology, intrinsically represent. My answer to the problem of explaining the source of representation, then, is similar to King's: to extract abstract objects from the general outline of a theory of how the mind represents without committing to the exact details.

What about the problem of the unity of the proposition? Though, as I said, it may not be as serious an issue, I should say something about it given its large role in the propositions literature. Iconic representations contain something like parts. I need to explain the difference between a collection of things and a proposition that involves the members of that collection. The difference is that, in an iconic content, the mind uses those parts in a simulation of a world while alone the parts don't participate in such a model. The fact that our cognitive systems employ simulations with parts unifies those parts and explains why they represent, which is a fact, given the use, about the proposition all on its own.

CHAPTER 2

THE PROPOSITIONAL BENACERRAF PROBLEM

Abstract

Writers in the propositions literature consider the Benacerraf objection serious, often decisive. The objection figures heavily in dismissing standard theories of propositions of the past, notably set-theoretic theories. I argue that the situation is more complicated. After explicating the propositional Benacerraf problem, I focus on a classic set-theoretic theory, the possible worlds theory, and argue that methodological considerations influence the objection's success.

2.1   Introduction

How can any given propositional candidate *be* the proposition that p when there are equally good candidates? Call this the *propositional Benacerraf problem*.[1] This objection originates in the philosophy of mathematics, targeting set-theoretic reductions of numbers, but generalizes to other entities, including some views of propositions. Yet in the propositional case, we perhaps don't fully understand the problem, which the fast-changing nature of the propositions literature complicates. Investigating and better understanding the problem will be this paper's main goal. The problem, we'll see, most obviously targets set-theoretic theories of propositions. These theories tend to come in two varieties: tuple theories and worlds theories. After explicating the propositional Benacerraf problem itself, I consider its application to the worlds theory as a case study. I claim that there are (at least) two plausible defenses of the worlds theory from the Benacerraf problem. These defenses serve

---

[1]Benacerraf (1965) is the original source of the eponymous problem.

to explicate the worlds theory in a way that coheres with realistic versions of the worlds theories—not contrivance to show that it's logically possible for the worlds theory to avoid the Benacerraf problem.

On the first defense, I argue that in some instances the problem fail because the objector confuses a philosophical representation for the object of that representation. On the second, I argue that, given a certain view of philosophical analysis, the arbitrariness at the heart of the Benacerraf problem is benign. This second defense may bring out a more general, big-picture goal: to make sense of two impression-istically distinct ways of thinking about propositions. Historically, some writers have treated propositions as *tools* that they put to work in some larger theory. As such, a theory's requirements define and exhaust the nature of these tools. Writers in this historical thread *stipulate* and *define* rather than *discover*. On the other hand, some have approached propositions directly, believing them worthy of study in their own right, with a nature to uncover rather than to define or stipulate. Here I have in mind, e.g., the recent interest in propositions as such, notably from Jeff King and Scott Soames.[2] Such confusion, in my view, results in objections like this:

> When I believe (doubt, justify, assert) some proposition, do I believe (doubt, justify, assert) a function? On the face of it, this is not plausible. Advocates of this reduction seem to have lost "the naive eye." (Bealer, 1998, p.5)

Do philosophers such as David Lewis simply miss the forest for the trees? No. I argue such objections only make sense absent the target's methodology, a method-ology that I argue provides a view of propositions that the Benacerraf objections misses.

---

[2]For book-length statements, see King (2007) and Soames (2010).

2.2 The Benacerraf Problem

We'll start with the original Benacerraf problem in the mathematical context, then move to the standard version of the problem in the propositional context, looking closely at the problem with respect to developments in the propositions literature, which gives rise to a new kind of Benacerraf problem whose force is less clear than the standard version. I close with typical responses to the Benacerraf problem.

2.2.1 Benacerraf 1965

Benacerraf raised a problem for the reduction of numbers to sets. For such a purported reduction, various progressions of sets work equally well. Two conspicuous contenders are von Neumann's ordinals and Zermelo's, each progression which goes, respectively, as follows:

(2.1) $\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \ldots$

(2.2) $\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \ldots$

These two progressions may serve equally well as *representations* of the natural numbers, but the question is, *Are they the natural numbers*? Benacerraf says no. Benacerraf presents the objection in parable form in which different logician parents bring up two children, Johnny (for John von Neumann) and Ernie (for Ernst Zermelo), teaching them their respective progressions instead of usual counting and mathematical instruction. Upon the set-theoretic instruction's completion, the children quarrel: Ernie insists that 3 is a member of 17 and Johnny disagrees. This disagreement brings out the first step of the Benacerraf objection: the candidate reduction bases are *non-equivalent*. The second step is the arbitrariness step. The naturals can't be both progressions since they're non-equivalent by step one, so the naturals must be one or the other (or some yet other progression). If the naturals

are, say, the von Neumann progression rather than Zermelo's, then there must be an argument for that choice. Yet, no argument is forthcoming—what could the reasoning possibly be since both work equally well foundations-wise? The final step is to conclude that the purported reduction/identification fails—in this case, that numbers aren't sets.

We can generalize the problem thus:

Benacerraf problem If there are multiple, non-equivalent classes of entities that contend for an identification (or reduction) base for some class of entities E, and there's no non-arbitrary reason to choose one reduction/identification base over another, then E in fact isn't (or can't be reduced to) any of those bases.

2.2.2 The Propositional Benacerraf Problem: The Classic Version

The objection applies straightforwardly to worlds theories. Step one: Worlds theorists employ both sets of worlds and functions from worlds to truth values. We may have the set of worlds at which snow is white—

$$\{w_1, w_2, w_3, ...\}$$

—and the function that pairs those worlds with 1 (or true, or whatever) and the others with 0—

$$\{\langle w_1, 1 \rangle, \langle w_2, 1 \rangle, \langle w_3, 1 \rangle, ..., \langle w_n, 0 \rangle, ...\}.$$

And it's clear that, ontologically,

$$\{w_1, w_2, w_3, ...\} \neq \{\langle w_1, 1 \rangle, \langle w_2, 1 \rangle, \langle w_3, 1 \rangle, ..., \langle w_n, 0 \rangle, ...\}.$$

Thus, the two are non-equivalent. Step two: Both sets and their characteristic functions work equally well by the world theorist's lights. Conclusion: Propositions

are neither sets of worlds nor their characteristic functions. Call this the *classic propositional Benacerraf problem*, or just the *classic problem* since we'll focus mostly on propositions, and I'll flag the mathematical case below.

2.2.3 The Propositional Benacerraf Objection: Further Thoughts and Another Version

The key feature of a Benacerraf problem of any type is untoward arbitrariness. About that arbitrariness, I raise two points that I discuss at once. First, the *source* or *nature* of the arbitrariness is important, for it gives rise to importantly different kinds of Benacerraf problems the strength of which may depend on the source of that arbitrariness. Second, different *kinds* of potential constraints may eliminate the arbitrariness.

The arbitrariness in the original mathematical Benacerraf problem involved, as Benacerraf himself presented it, the epistemic idea of *cogent reasons*: If one doesn't have a *cogent reason* to favor that $a = b$ rather than that $a = c$ (where $b \neq c$), or vice versa, then it's arbitrary whether $a = b$ or $a = c$. And if it's arbitrary whether $a = b$ or $a = c$, then $a \neq b \neq c$. But it can't be merely that if we lack cogent reasons, at present, whether $a = b$ or $a = c$, then a Benacerraf problem arises. If either the butler or the gardener solely committed the murder, and we don't have a cogent reason to identify one or the other as the murder, that may give us reason not to accuse or indict either—that would be arbitrary. But it's not arbitrary, holding fixed the disjunction, which *is* the murderer despite our ignorance. It mustn't be that we merely lack a cogent reason, but that, for some reason, no cogent reason is forthcoming.

Why is no cogent reason forthcoming? In the original mathematical Benacerraf problem and in the classic propositional Benacerraf problem, it seems that not only is no cogent reason forthcoming, but no reason could *ever possibly* be forthcoming.

In no possible world could we obtain cogent reasons that would favor choosing one set progression over another in the mathematical problem. In the classic propositional case, if we grant that there are roles for coarse grained propositions, then, as in the mathematical case, there is no possible world in which we could have cogent reasons to favor the set version versus the function version. In both the mathematical and classic propositional Benacerraf problems, we're choosing among candidates according to role suitability—whether, respectively, some entities behave like natural numbers (in that we can define arithmetical operations on them, e.g.) or whether some entities behave like coarse-grained content (in that factually equivalent propositions are identical, logical relations are definable, etc.). And given that the candidates in both reduction/identification projects are formally equivalent, there can be no *possible* cogent reasons to favor one candidate over another.

We solve a Benacerraf problem when we resolve this arbitrariness, and we resolve the arbitrariness on the basis of some kind of criteria or other. Yet there are different *kinds* of criteria, and recent developments in the propositions literature give rise to a new kind of criterion that can affect the modal profile of the just-mentioned cogent reasons. We're familiar with the two typical kinds of criteria on which we judge candidates in an identification/reduction. The first I'll just call *role-filling* criteria. These criteria are simply whether the candidate class of entities really satisfies the recognized *desiderata* for a theory. In the original Benacerraf problem, if we couldn't define the successor relation on either the von Neumann or Zermelo progressions, then the arbitrariness would obviously resolve. For propositions, any entities that we identify with propositions have to play whatever propositional roles we intend those propositions to play. It may be that one class of entities can play all of the propositional roles, or perhaps only some roles if "[t]he conception we associate with the word 'proposition,'" as Lewis (1986, p.54) says, is "a jumble of conflicting *desiderata*."

Whatever the case may be, if we set out to give a theory of, e.g., the objects that enter into logical relations, then we need to be able to define logical relations on those objects. If one class of entities satisfies this criteria and another doesn't, then that favors the role-filling class of entities. I may be stating the obvious—that *desiderata*-satisfaction favors one class of entities over another—but this arbitrariness resolver contrasts with another: instrumental criteria, such as simplicity, economy, etc. While such criteria can perhaps be called upon to break ties—simplicity in particular has in the propositions literature—such criteria are applied *after* role-filling criteria. If we're presenting a theory of the objects of logical relations, and we can define such relations on one class of entities and not another, then it doesn't matter how simple, e.g., the latter entities are.

The propositions literature has recently taken an empirical turn, what I will call the *mind-first* movement, and with this turn comes a different kind of tie-breaking criteria. I won't dwell on whether this kind of criteria is fundamentally different from role-filling criteria. According to the mind-first movement, the mind is the ultimate source of representation from which propositions inherit their representational capacity, and the *basis of content*—the workings of the mind that give rise to content—constrain the mind-first theorist's proposed propositions. What this means is that in addition to the propositions filling their various roles, propositions must somehow be related to the actual, empirical workings of the mind. Thus, if we have an apparent tie between two candidates both of which satisfy their proposed roles equally well—and both are equally simple, yet one *actually* is derived from the basis of content while the other is somehow in error in its understanding of the empirical facts that give rise to content—then the tie is broken in favor of the empirically informed entities. For example, for King, suppose we focus on two propositional candidates for the proposition that p, one of which includes binary branching syntax only while the other includes some ternary

56

branching in its syntax. The candidates may serve equally well as the objects of the attitudes, or as the content of assertive sentences, or whatever. Further suppose that linguists inform us that the binary branching hypothesis is true. Then, since one candidate gets the empirical facts about the basis of content wrong, we break the tie.

Let's see how this material plays out with some concrete examples from the literature. The Benacerraf problems that writers have recently raised have mostly targeted King's theory.[3] For example, Caplan and Tillman recently raise two Benacerraf problems for King's theory. Here I focus on an aspect of King's theory relevant to the Benacerraf problem. Consider the simple sentence, "Rachel reads." As the words "Rachel" and "reads" stand in the sentential relation R, the *person* Rachel and the *property* of being a reader stand in a *propositional* relation. According to King's *original account*, the propositional relation was this:

(PR_1) Rachel stands in the PR_1 relation to the property of being a reader iff in language L, ___ is the semantic value of a lexical item x, and ___ is the semantic value of of lexical item y, occurring at the left and right terminal nodes, respectively, of the sentential relation R.

King's original account of the proposition that Rachel reads was the fact that $PR_1(Rachel, reading)$. King (2007) amends the account in a few ways. First, he notes that syntax, in this case the relation R, has semantic significance—that R "instructs" the semantics to deploy the instantiation function at syntactic concatenation points. PR is updated thus:

(PR_2) Rachel stands in the PR_2 relation to the property of being a reader iff in language L, ___ is the semantic value of a lexical item x, and ___ is the semantic

---

[3]But see Dixon and Gilmore (2016) for a recent Benacerraf problem targeting Speaks' theory.

value of of lexical item y, occurring at the left and right terminal nodes, respectively, of the sentential relation R, which in L encodes instantiation.

Second, King worried that it is contingent that syntax instructs semantics to deploy the instantiation function rather than, say, the anti-instantiation function. Because of this, King claims that PR itself encodes instantiation, which, according to King, secures the intrinsicality of propositions' truth conditions—call this King's *new account*:

(PR_3) Rachel stands in the PR_3 relation to the property of being a reader iff in language L, ___ is the semantic value of a lexical item x, and ___ is the semantic value of of lexical item y, occurring at the left and right terminal nodes, respectively, of the sentential relation R, which in L encodes instantiation, and PR encodes instantiation.

King's new account comes with a story to explain how PR came to encode instantiation.[4] Since King's theory is mind-first, the representational capacity of his propositions is derivative. So there is something humans *did* to *something*—call it the "proto"-proposition—to endow propositions with their representational capacity. The proto-propositions are simply the facts involving the relevant entities standing in the PR_2 relation. At some point in our distant past, "speakers attempted to make use of structured contents, in part by beginning to employ that-clauses to talk about them, they would have implicitly taken the structured content of a sentence to have the same truth conditions as the sentence" (King, 2009, p.267). In doing so, the propositional relation came to encode instantiation.

Caplan and Tillman (2013) claim that we can't break the tie between King's account involving PR_2 and his new account involving PR_3. Both have truth conditions, although the new account has them intrinsically while the old account has

---

[4]See the "First Objection" section from King (2007, ch3) and King (2009, pp.266-67).

them extrinsically. Yet, they claim, the old account is simpler while the new account is more complex. Here I think our different criteria and their order of importance shows why this is not a genuine tie, and the impression that we have a Benacerraf problem arises from mixing the importance of criteria. Intrinsic truth conditions is a *role-filling* criteria while simplicity is an *instrumental* criteria and the former take priority over the latter.[5]

They raise a second Benacerraf problem for different candidates for the *proto-proposition*. They claim that King's original account involving PR_1 and PR_2 are equally well-suited. As they say:

> The difference between . . . [King's choice for the proto-proposition, involving PR_2] and . . . [King's original account] is that only the former includes as a conjunct the fact that [R] encodes instantiation. But, as long as [R] does encode instantiation, the [original account] will have the right truth-conditions and so shouldn't be a worse candidate for being [a given proposition].

As Fletcher (2014, pp.8-10) notes, Caplan and Tillman's second proto-propositional Benacerraf problem gives rise to a full-fledged Benacerraf problem. If King's original account involving PR_1 is the proto-proposition, then the following will be the final account of full-fledged propositions:

($PR_{2.5}$) Rachel stand in the $PR_{2.5}$ relation iff in language L, ___ is the semantic value of a lexical item x, and ___ is the semantic value of of lexical item y, occurring at the left and right terminal nodes, respectively, of the sentential relation R, and PR encodes instantiation.

The only difference is whether R encodes instantiation. But so long as PR does, then it doesn't matter, says Fletcher, whether R does. So we have a tie between the fact that PR_3(Rachel, reads) and the fact that $PR_{2.5}$(Rachel, reads) for the proposition that Rachel reads.

---

[5]Fletcher (2014, p.8) makes this point without my language of criteria and their order of importance.

While the relative importance of criteria resolves Caplan and Tillman's first Benacerraf problem, their second, and Fletcher's improvement, may be solved with the new kind of criteria—the empirical criteria. Let's grant that Caplan, Tillman, and Fletcher are right when they say that both candidates are *eligible*—in my language, that both serve equally well in terms of role-filling criteria. Something may be said in favor of the proto-proposition involving PR_1 (and so the full proposition involving $PR_{2.5}$) in terms of simplicity, but that doesn't seem very appealing. The Benacerraf problems under discussion revolve around whether the proto-proposition included R's encoding instantiation or not. Here things get complicated—which is why I said empirical criteria "may" resolve the problem—since there are different views on the nature of linguistics.[6] Since King recruits Chomskyan syntax as his favored syntax that provides propositional structure, I think it's fair to say that the syntax involved in his propositions is an idealization of psychological processes. But if this is the case, then, regardless of the role suitability of the candidates, the evolution of our syntactic and semantic modules may break the tie here. It's hard to know how seriously King takes his evolutionary story that culminates in his final account, but regardless of that, there will be some empirical fact of the matter regarding R. Regarding syntax providing instructions to the semantics, King (2007, p.34) says the following:

> [O]n any approach to compositional semantics for natural languages, even one that eschews propositions altogether, one will have to invoke this idea. For on any compositional semantics, the semantic values of expressions at terminal nodes need to compose to yield new semantic values for non-terminal nodes, which themselves must compose, until we get a sentence level semantic value. The syntax gives the instructions as to how the semantic values are composed. . . . Semantic approaches differ only on what they claim is the instruction that a given bit of syntax provides. They are all stuck with the idea of syntax providing instructions.

---

[6]I cover some of these complications in chapter 1.

Here, King (2009, pp.266-67) describes his evolutionary story that brought the propositions in his final account into existence:

> Now here let me tell a highly idealized and mythological story about speakers coming to interpret the propositional relation of the fact that is the proposition that Dara swims in the way they do. Consider a time when languages had come into existence, bringing such facts into existence, but the propositional relations had not yet been endowed with the significance that gives propositions their truth conditions. As speakers attempted to make use of structured contents, in part by beginning to employ that-clauses to talk about them, they would have implicitly taken the structured content of a sentence to have the same truth conditions as the sentence. Now consider all the facts that consist of Dara and the property of swimming bound together by some relation and that would have been candidates to be the proposition that Dara swims. My thought is that whichever of these had been 'chosen' to be the proposition that Dara swims, the relation binding together Dara and the property of swimming in the proposition inevitably would be interpreted as ascribing the property of swimming to Dara in virtue of the fact that speakers implicitly and without thinking took the proposition to be true iff Dara possesses the property of swimming.

To the Caplan and Tillman objection, there will be some fact of the matter—*if* the whole evolutionary story is true—regarding the role of the proto-proposition. It will be an empirical matter in linguistics what the nature of *R* was in our distant past, and that will, on one view of syntax, be a question of our past psychology vis-a-vis syntax. And if this is the case, then to Fletcher's objection, there will be some fact of the matter regarding the nature of *R* in King's final account. Thus, one may criticize King's evolutionary story and his metaphysics of syntax. And it may turn out that King will have to tweak his account in light of further empirical investigation. These are all potential problems, but they won't be *Benacerraf* problems.

We've just gestured toward empirical criteria breaking ties. Yet the empirical turn in the propositions may give rise to a kind of indeterminacy, previously foreign in the propositions literature, that can crop up amidst any empirical investigation, and it's unclear whether this indeterminacy results in a Benacerraf problem or not. We're familiar with indeterminacy arising from vagueness, future contingents,

translation, reference, etc. The list goes on and isn't restricted to such philosophical mainstays: For example, given the historical city-planning facts, it's indeterminate whether Princeton is the borough of Princeton or the borough in addition to the surrounding township.[7] Why is this relevant to us? Because that's what the Benacerraf problem involves in other words—irresolvable indeterminacy. But the kind of indeterminacy that often arises in metaphysical investigations is different from the kind that arises in empirical investigations, especially investigations that employ modeling methodology.[8]

I have in mind scenarios involving a non-theoretical level of description underdetermining facts about a theoretical level of description whose purpose is to account for that non-theoretical level. Here are a few examples:

- Consider formal epistemological facts about someone's subjective confidence in some proposition—say whether it will rain tomorrow. There are, on the one hand, the facts about the person's brain and her behavior (say betting behavior), and on the other, her subjective confidence in rain, say it's .5. There may be no fact of the matter whether her credence in rain is .5 or .50000...1 (add in as many zeros as you like). This is because the one level of description—the facts about her brain, her behavior—whatever goes into determining whether and to what degree someone believes something—underdetermines what her credence is.

- Consider the Lewisian view about the relationship between languages and language in Lewis (1975/1983). On the one hand, we have language as a convention-governed social phenomenon—conventions that Lewis accounts for game theoretically; and on the other, we have language as a formal seman-

---

[7]This example is from Lewis (1988, p.129).

[8]"Modeling" here means the methodology discussed in philosophy of science; for example, see Weisberg (2013) for a contemporary book-length discussion.

tic function from strings to meanings, where meanings determine extensions in possible worlds. Given that there will be some indeterminacy in the regularities that govern linguistic conventions, which in turn determines which language some population speaks, there will be indeterminacy in the relation between a given population's language as a social phenomenon and the formal semantic account of that language.

- Suppose at the end of syntactic inquiry, there is no evidence to distinguish between whether every branching node is binary or whether some branches are n-ary. Further suppose that syntacticians model actual human psychological processes, as many take themselves to be doing. Of course one reason that such a situation may arise is that our science will just never be good enough. But it may be that human physical and behavioral facts underdetermine fine-grained theoretical distinctions at the level of the syntactic model

The *level* at which this indeterminacy resides is, respectively, a given agent's subjective confidence, the correct semantics for a natural language, or human syntactic knowledge. I'm *not* interested in the models of these—respectively, candidate credence functions, set-theoretic formal semantics, and tree- (and so set-) theoretic syntax. As we'll see, Benacerraf objections don't apply at the level of a theoretical model.

The above scenario types look similar to the classic mathematical and propositional versions: We have multiple theoretical candidates to choose from among which no candidate is better than others in terms of cogent reasons. For the classic propositional Benacerraf problem, only role-filling and instrumental criteria could break the tie. Like I said, the lack of cogent reasons is necessary and the indeterminacy is necessarily irresolvable. Yet there is a feeling of contingency once we introduce empirical investigation into theories of propositions. It may very well turn out that certain facts will remain unresolved not merely as a matter of our

clumsy science but even at the end of actual ideal science. In these cases, it seems to me that it's *possible* that some evidence could distinguish the candidates. It's just that (or at least we've stipulated) that there is no *actual* evidence to distinguish the cases. If we can distinguish between .9 and .1 credences with some confidence, then it seems possible that for some very fine-grained distinction, there could be some behavioral or brain fact that distinguishes them. Similar considerations go for the other two examples: there will always be indeterminacy in the relation between Lewis' two notions of language, but for two given formal semantics, we can produce a world where *that* indeterminacy resolves. And if we have constituency tests that give us cogent reasons to identify obvious constituents, we can imagine speech and brain facts that would resolve the indeterminacy between binary and *n*-ary branching. That we're left with these indeterminacies and not some others at the end of science just seems to be a contingent fact about our world.

I want to be clear with what I mean about "contingent" indeterminacy. When I say that familiar cases of indeterminacy, for example vagueness, are necessarily irresolvable, what I mean is this: Once we fix upon *our* concept of, e.g., baldness, and we fix upon some borderline bald person, it doesn't matter how other facts vary from world to world—that person will be a borderline case in every world. Once we fix our concept of what a proposition is and fix that a given proposition *p*'s identity is indeterminate between a set of worlds and a characteristic function thereof, we can vary other facts and the indeterminacy will remain irresolvable. I'm contrasting this case with a different kind of indeterminacy, and this different kind of indeterminacy is relevant in our current propositions literature. This difference, in the propositional case, arises from theorists introducing empirical considerations into their theories. For example, King employs syntax in his facts that are his propositions. Syntax is, of course, an ongoing filed of investigation, and King

employs syntax in his propositions *however, with some constraint,*[9] *syntax turns out in the end*. Now, in our world, it may turn out that, with respect to some syntactic facts, those facts don't turn out one way or another. There may be some irresolvable *actual* indeterminacy between the facts about our brains and speech behavior and syntax. But how things turn out *at some other world* may be different—such indeterminacies may resolve. And this is what I mean by the indeterminacy being contingent.

This matters for us because which candidate proposition is *the* proposition that p may depend on the resolution of some actual indeterminacy that may, in our world, be irresolvable. The binary/n-ary branching example is just one example. Thus we may end up with multiple propositional candidates equivalent in role and instrumental suitability, and because of empirical indeterminacy, they may be equally empirically well-suited. It may seem, then, in such cases that we have a Benacerraf problem on our hands, but I think this is the wrong conclusion to draw. In this set up, where empirical investigation stalls out, I think the indeterminacy is benign and should be chalked up to the complexities of the empirical world. If a Benacerraf problem were to threaten *here*, at the propositional level, then we would have an explosion of Benacerraf problems wherever there is indeterminacy in the interface between philosophy and science.

2.2.4 Reactions and responses in the Literature

The Benacerraf problem, as I noted in the outset, plays an important role in the contemporary propositions literature. King's theory was partially constructed to be Benacerraf immune (as he was reacting against theories he thought succumbed to the problem) and has been dismissed for nevertheless succumbing. Soames leveled a Benacerraf objection against what I thought was a promising theory—

---

[9]For example, syntax not providing instructions to the semantics to perform the relevant compositions may doom King's program.

his own "deflationary account"—that leads him to his current "cognitive realist theory" according to which propositions are event types that involve mental acts. Others have drawn dire conclusions in the face of the arbitrariness worries like the Benacerraf objection. Moore (1999) and Bealer (1998) argue from the Benacerraf objection to the conclusion that propositions are *sui generis* entities; Armour-Garb and Woodbridge (2012) employ the problem to argue for a fictionalist account of propositions; and Jubien (2001) presents a dilemma a horn of which is the Benacerraf objection to argue that propositions don't exist. The reaction has been that the Benacerraf problem is not merely a demerit for a theory but a theory ender, and the possibility of a general Benacerraf problem has lead to the rejection of propositions all together.

Despite obvious disanalogies between the mathematical and propositional cases, the responses to the propositional problem mirror the mathematical problem—platonism, nihilism/fictionalism, and structuralism. The propositions case, however, differs from the mathematical case in that platonism dominates. Most current propositional proposals identify propositions with some kind of abstract entity, but as noted, Armour-Garb and Woodbridge take a fictionalist line that shares similarities with Hartry Field's work in the mathematical case. Benacerraf inaugurated structuralism in the mathematics case in response to his own problem. Though structuralism is conspicuous in the mathematics case and seemingly absent in the propositional case, there is a coherent structuralist-like view of propositions to be had. Inferentialism, the view that the meaning consists in the licensing of various inferences, may lend itself to a structuralist interpretation. The content of the claim that p, according to the inferentialist, consist in the inferences from and to p. Thus we can think of a web of inferences as a certain kind of structure, and we have a structuralist analogue in the propositional case.

Having brought out the complexities of the propositional Benacerraf problem itself, let us turn to the complexities in its application.

2.3 Worlds theories

This section splits into two. I consider one of the main, and what seems to be (one of the) easiest, targets for the propositional Benacerraf problem, the worlds theory, and I consider two different ways to respond. I focus on two different contexts among others in which worlds theories may be employed. I argue that, once we look more closely at methodological issues, the Benacerraf objection against the worlds theorist in either context isn't an inevitable winner. For each section, I will bring in various philosophers as representatives of that kind of worlds theorist. My categorization of each philosopher squares with what I think is the best interpretation of that philosopher, or at least a compatible interpretation. But this section isn't principally hermeneutic. Each of the following contexts constitutes a plausible direction to take a worlds theory regardless of which philosopher actually fits into which context.

2.3.1 The modeler

One kind of worlds theorist employs sets of worlds (or characteristic functions thereof) to *represent* or *model* propositions. Of course, one properly lodges the Benacerraf problem at the level of the *target* of a model. If the Benacerraf problem fails at the level of the modeler's target, then this kind of worlds theorist may escape the Benacerraf problem. But the cost of such a view may be that the modeler isn't really giving us any insight into the *metaphysics* of proposition—she "merely" provides representations, as the attitude goes. I want to suggest that we can plausibly combine the worlds theory and modeling methodology, and that the target of such models are Benacerraf resistant, at least on some ways of understanding

the approach. The plausibility of the approach, then, will rest upon whether the targets really are Benacerraf immune and whether the attitude of "merely" providing representations can be resisted. To show that the latter part is plausible, it needs to be shown that theorizing about propositions obliquely via a model is a fruitful approach to propositions. To show as much, I think we need to appreciate two things. First, that modeling in metaphysics isn't isolated to just propositions but constitutes a mainstream methodology. Second, that when we look at what propositions *do*, even on contemporary, mind-first theories—ones which have taken to investigate the metaphysics of propositions "directly"—what propositions *do* on such theories is only clearly revealed at the level of the model, just as with the world-theoretic modeler.

If sets of worlds represent or model propositions, then what, at the level of fundamental ontology, are propositions? There are two general options: The target of the propositional models may be *sui generis* entities or not. It's not always clear what a given philosopher means by *sui generis*. The idea is supposed to be that some entity is of a type that we can't reduce to some class of entities we already recognize. In the propositional context, we may mean that propositions can't be reduced or identified with some other entities such as sets of worlds, event types, syntactic-semantic facts, etc. So, in one sense, propositions, if *sui generis* form their own ontological "category" in that we count them among the basic entities in the universe that we don't identify with or reduce to something more basic. The individuation of what counts as an ontological category is somewhat unclear, though, so in another sense, we may still assimilate propositions to this or that ontological "category" in the sense that we may think they are objects or properties, abstract or concrete, etc. For example, according to Schiffer's theory,[10] propositions

---

[10]See Schiffer (2003).

are *sui generis*, yet they are things rather than, say, properties, and they are abstract rather than concrete.

Many may find the *sui generis* view implausible, but there is a feel of added implausibility when we combine the view with modeling methodology. Sets of worlds model various properties and relations that a given proposition may have and stand in. When we represent a proposition as having a given property or standing in a given relation to some other proposition, it's hard to see why such modeling works. We have the propositions at the level of ontology and their representations at the level of the model. In virtue of what makes one given model rather than another successful? One standard way to think of the success of a model is in terms of isomorphism: The model is explanatory because it is, in some useful way, isomorphic to the target. But with *sui generis* entities, what could account for the success of one model over another? What about the proposition that snow is white, at the level of fundamental ontology, accounts for its pairing with the set of worlds where snow is white rather than the snow-is-green set? The answer may just be a restatement of the view that entities can primitively represent, so that the correct pairing captures the primitive representational content of the proposition. There is, however, a powerful argument against any such view in Lewis' argument against magical ersatzism in *On the Plurality of Worlds*.[11] Philosophers haven't perhaps appreciated the scope of Lewis' argument, which applies to any primitively representational entities, not just ersatz worlds.[12]

Here is an extremely compressed version of the argument: *Sui generis* propositions will stand in the *selecting* relation to this concrete world just in case a proposition represents it. This relation is either internal—depending only on the intrinsic natures of the elements—or external—depending on the intrinsic natures of the

---

[11] See Lewis (1986, §3.4).

[12] I got this idea from Bricker (2006).

relata taken together. If it's internal, then there must be some differences among the *sui generis* propositions to explain why the snow-is-white proposition stands in the relation while the snow-is-green proposition doesn't. Being a vast collection of abstract, *sui generis* entities, the only difference between the two propositions would seem to be that one represents this world while the other doesn't. This leads to a vicious circularity: A proposition *p* is selected because it represents the world as thus and so. But why does it represent the world as thus and so? Because it is selected if and only if the world is thus and so. If the relation is external, then we have an unacceptable necessary connection between various propositions (those that represent this concrete world) and this concrete world. But the goings on at this concrete world and the relations that it enters into with abstract *sui generis* propositions seem related, if at all, only contingently and not necessarily. This is Lewis' argument adapted for *sui generis* propositions.

I want to be sure that I don't characterize Lewis' argument as decisive—it's complicated, as usual. van Inwagen (1986) argues that he isn't sure what, exactly, is wrong with the argument, but if the argument does work, then it will threaten the relation of set membership, which may count as a kind of reductio of Lewis' argument. Jubien (1991) provides a comprehensive overview of Lewis' argument, weighing in against. In this dialectic, I should note two things. I'm raising the possibility that sets of worlds are models of *sui generis* propositions. First, to move onto non-*sui generis* views, I am noting that such a view of propositions has problems on its face—that Lewis' argument, *if* one finds it plausible, applies straightforwardly to such a view. And I'm noting that the combination of such a view with modeling seems extra strained, especially when we consider the main, developed view of *sui generis* propositions—Schiffer's—according to which, in addition to being *sui generis*, propositions are simple (i.e. structureless). Jubien (1991, p.266), an opponent of Lewis' argument in the abstract, would seem to agree:

> [I]f we agree that propositions are simples, then it certainly *does* seem far-fetched to think that their intrinsic properties could be such as to represent ways the world might be. Perhaps there could be enough of them to do the representing, with each intrinsically different from every other, but it does not seem that the *nature* of their intrinsic features would make any one of them flat-out *be*, say, the proposition that a donkey talks. Assuming they are simples is in this respect a little like reducing them to *numbers* (say, unreduced ordinals). We might be able to concoct a *model* of "proposition theory" from a certain class of numbers, but there would of course be no intrinsic reason to select some one particular number as the proposition that a donkey talks; any other number would do just as well.

Some class of simple, structureless entities may be useful as the *model*—if there are enough of them—but not as the *modeled*. Yet one may still insist on identifying *sui generis* entities with propositions at the level of fundamental ontology and model them with world-theoretic devices, and whatever problems such a view may incur, it won't be from a Benacerraf problem. The problem fails at step one since there isn't a surfeit of candidates by stipulation.

The more plausible option is that sets of worlds model propositions that, at the level of fundamental ontology, are non-*sui generis* entities. What kind of entity? *Properties* are a natural answer. First this squares, hermeneutically, with the ways that some theorists, such as Stalnaker,[13] describe possible worlds and set-theoretic devices—that the former are "ways the world might be" and that the latter *explain* or *explicate* such ways of the world's being. And second, there are explicit views

---

[13]Stalnaker often says that possible worlds *explicate* or *explain* propositions, or that propositions *determine* a set of worlds/function from worlds to truth values: "The *explication* of *proposition* given in formal semantics is based on a very homely intuition: when a statement is made, two things go into determining whether it is true or false."(Stalnaker, 1970/1999, p.32, my emphasis on "explication") "If we *explain* propositions as functions from possible worlds into truth-values, they will have the properties that have traditionally been ascribed to them."(Stalnaker, 1970/1999, p.33, my emphasis) "[A]ny proposition *determines* a set of possible worlds." Stalnaker, before this quote, writes that "a propositions is a function from possible worlds into truth-values." But soon after he writes in typical analyst language: "A proposition—the content of an assertion or belief—is a representation of the world as being a certain way. But for any given representation of the world as being a certain way, there will be a set of all the possible states of the world which accord with the representation—which are that way."(Stalnaker, 1999/1978, p.79, my emphasis)

of propositions in the literature according to which propositions are properties. The views in the literature differ, but the overarching idea identifies propositions with ways the world might be.[14] According to one view, propositions are 0-adic properties (or relations).[15] We canonically represent an *n*-addic relation as a sentence with *n* places missing, e.g. the two-place relation of *loves* as *_loves_*. When we put some particular object, say Bob, in one of the slots, we get the property *Bob loves_*. Then if we add another object, say Mary, we get the zero-addic property *Bob loves Mary*, which is the proposition that Bob loves Mary. These properties that are propositions, though, aren't (typically) maximal in that they don't settle every aspect of a possible world; they only settle one if instantiated—that it is a *p* world. One nice feature of this account is that it provides a clean demarcation between properties that are propositions and those that aren't in terms of adicy.

Lewis and Chisholm both advance property views not in the 0-adic tradition. Both endorse a view in which propositional attitudes are relations between an agent and a property while sentences don't express properties. They both count as pluralists, with properties playing the role of the content of mental states, and propositional attitude ascriptions are analyzed in terms of the self-ascription of properties. For both Lewis and Chisholm, propositional properties aren't 0-place but 1-place. Perhaps the most salient recent property view, Speaks', is similar to Lewis' and Chisholm's views in that propositional properties are 1-place, but his view differs in two ways. First, Speaks is a univocalist about propositions, with properties playing every role. And second, although his properties are 1-place, they are not self-ascriptions. Rather, either *everything* or *nothing* instantiate Speaks'

---

[14]This requires, among other things, abundant properties.

[15]van Inwagen (2004, fn.17) endorses the zero-adic view, and Swoyer and Orilia (2014, §7.5) explicate such a view.

propositional properties. Speaks identifies the proposition that p as the property (or 1-adic relation) of *being an x such that p*.[16]

Speaks[17] says that he doesn't understand what a 0-place property is, and there is something right about this. We have an intuitive grasp of properties and two-or-more-place relations from which we start our philosophical analysis, but not with 0-place properties. What unites the 0-place view with views like Speaks', Lewis', or Chisholm's? To say that propositions, in some role, are properties is to say that something *is some way*. For Lewis and Chisholm, in the limited role of propositional attitudes, that thing is oneself. When I believe that snow is white the way that I think I am is that I inhabit a snow-is-white world. For Speaks, that means that something (viz. every *thing*) is such that snow is white. The zero-place view, I think, is perhaps best understood as similar to Speaks' view. If 1-place properties say how some thing is, 0-place properties say how everything or nothing is.[18] Thus the overarching idea of the property view is that propositions are some way that something is, be that a person or everything—call these *world properties* since the model will be pitched in terms of worlds. Our modeler, then, will understand p's being the case in terms of property instantiation. And there is good reason to think that at this level, at the level of properties, the Benacerraf objection fails. However the details of a property view play out, there is good reason to think that the Benacerraf problem doesn't apply to properties. If I say that the world is a certain way, there won't be multiple candidates to be that way—just as if I say some particular object is some way, say that a ball is red, there won't be multiple

---

[16]My knowledge of Chisholm's view is indebted to Dixon and Gilmore (2016). Also, as they note (p.276), Speaks' theory may need to be stated in a way less connected to language, but I'm merely getting at the general idea here.

[17]See King et al. (2014, p.90, fn.36).

[18]As Williamson (2016, p.762) says: "What is a 0-place property or relation? It is a state of affairs, in a sense to be refined by analogy with whatever theory of properties and relations is in play."

candidates to choose from in accounting for the ball's redness. In other words we either won't have a Benacerraf problem or if we do, it will explode to include properties, in which case it won't be this theorist's problem in particular. For the rest of this sub-section, let's focus on a property view like Speaks' or the 0-place view.

Like the *sui generis* view, the Benacerraf problem doesn't apply, but unlike the *sui generis* view, the property view is more plausibly combined with world-theoretic devices as models. The *sui generis* and modeling methodology clashed because it was difficult to see why any given set of worlds (or characteristic function) would pair with one proposition rather than another, at least if the *sui generis* view is something like Schiffer's. The property view has enough structure and detail to set up a fruitful isomorphism. Stalnaker (1987, p.9) captures the idea that I'm after with an example concerning the relation between certain physical properties and numbers. Why is it that we represent certain physical properties, such as height, in terms of numerical properties relative to a scale—say, being 5'9"? It's because physical properties such as height and weight have a structure in common with (positive) real numbers such that we can pick out these properties with numbers. In doing so, we understand having a certain physical property in terms of an object relating to a number given a unit. Thus we have a natural intermediary to approach physical properties via numerical properties given a unit, and we give a rigorous account of such properties via measurement theory, which we cast in terms of numerical quantities. In this number–height case, there is a natural pairing between the elements of the model and elements of the target and likewise for world theoretic devices. We pair the the world-property p with the set of worlds that have one thing in common—that they instantiate p—but otherwise vary in every other possible way. In the number-height case, there is a systematic relationship between the similarities and differences in the model and in the target. The same

goes for world-theoretic devices and world properties. Such a relationship allows us to explore relationships among the members of the model, which often lead to discoveries, and transfer those insights to the target.

The move that I'm considering—that a world theorist can avoid the Benacerraf problem by employing modeling methodology—isn't clearly available to other theorists. The other common target of the problem is the tuple theory,[19] according to which propositions are n-tuples containing constituents related to the sentence expressing the related proposition and whose order mirrors that sentence's syntax. On one way of expressing the theory, the proposition p is the tuple $\langle \langle e_1, ..., e_n \rangle, R^n \rangle$ where $\langle e_1, ..., e_n \rangle$ is an n-ary tuple of objects and R^n is an n-ary relation, and p is true just in case the entities stand in the relation. The theory is an obvious Benacerraf target since, ontologically,

$$\langle \langle e_1, ..., e_n \rangle, R^n \rangle \neq \langle R^n, \langle e_1, ..., e_n \rangle \rangle,$$

and since both tuples work equally well by the relevant theorist's lights. In particular, as Schiffer (2016, pp.2553-2554) notes, working well for this theory means making clear what the *truth conditions* of a given proposition are—as we've just laid out—and the identity conditions, which depend, some say to the theory's credit, not only on the truth conditions but also on the proposition's constituents and their order. But, as Schiffer further notes, if we take tuples as models of propositions, it's hard to see what the target is. In other words, there won't be some non-*sui generis* entity that answers to those two defining features, for they can't be any kind of set-theoretic entity since the Benacerraf problem will remain, and no other sort of target entity is obviously forthcoming. Of course, the tuple theorist can just posit

---

[19]I use "tuple theory" instead of the more common "Russellianism" since Russellianism is both a thesis about structure and propositional constituents. The tuple theory can be combined with Freagean constituents as well.

entities with those two features to serve as the target of her tuples, but such a move would seem more like a wish list for a theory rather than a theory.

I want to set up a comparison between the world-theoretic modeler and contemporary accounts. We've discussed King's theory above, but let's get on board perhaps the other main theorist in the propositions literature: Soames. Soames identifies propositions with *event types*. On Soames' account, propositions' representational capacity derives from the representational capacity of agents. According to this account, to entertain a proposition is to engage in a cognitive activity—the mental act of predication, which for Soames is a primitive mental act, basic among others. When an agent sees *o* as red, there is an event token of her predicating redness to the object *o*. Now of course this event token can't be the proposition that *o* is red. Instead, Soames takes the cognitive event *type* of predicating redness to *o* to be the proposition that *o* is red expresses. And more complex propositions and differing propositional attitudes are built upon the mental event type of predication. Regarding other attitudes, some examples: To judge that *o* is red is to predicate redness to *o* and then to endorse that predication; to believe that *o* is red is to predicate redness to *o* and to be disposed to judge that *o* is red (and so to be disposed to endorse the predication of redness to *o*). Regarding more complex propositions, an example: To entertain the proposition that *o* isn't red is first to predicate redness to *o* and then to predicate not being true to the result of that first predication.

On to the comparison: Perhaps *the* most basic requirement that we ought to have of a theory of propositions is that it provides the correct logical structure among the propositions. By this I mean that propositions relate to one another in terms of logical strength, entailment, as contradictories, as contraries, etc. The properties view, at the level of ontology, translates naturally to the worlds view, which provides a very elegant understanding of these basic logical relations, all in terms of set theory, which illuminates relations among property instantiations.

So, e.g., p will entail q, at the level of ontology, just in case p can't be instantiated without q also being instantiated—or, at the level of the model, just in case p is a subset of q. *Gigi is a cat* entails *Gigi is a mammal*, so, to use Speaks' theory, *being something such that Gigi is a cat* can't be instantiated without *being something such that Gigi is a mammal* is also instantiated. The model has this as the Gigi-cat worlds being a subset of the Gigi-mammal worlds. These relations between sets mirror relations between the instantiation of world properties. And we can build more complicated structure with the worlds apparatus as our various theories demand—again, with the full power of set theory. For instance, we capture the structure of the domain of a credence function in an elegant way.[20] A similarity in structure between world-properties and set-theoretic models explains the ability of the latter to so fruitfully represent the structure of the former. As we advance our understanding of physical quantities via measurement theory defined on numbers, we advance our understanding of propositions via systematizing propositional relations defined over set-theoretic constructions.

Now take any two of Soames' propositions. That is, consider two event types the tokens of which involve agents predicating properties of objects and other more complex events involving mental acts. Is there a clear answer to the question of whether, for example, they relate as contraries? Or for any two, is one logically stronger than the other? Likewise consider two Kingean propositions. These will be complex facts comprising semantic properties and relations and syntactic relations that encode semantic operations which on one interpretation are similar to Soames' mental operations. Is there a clear answer to whether standard logical relations hold between those two propositions? I hope it strikes the reader that at the level

---

[20]For various technical reasons, we take the domain of the credence function to be a σ-algebra over the space of possibilities rather than just the space of possibilities itself, and the worlds framework captures the structure of such an algebra elegantly.

of ontology, the answer is either "no," "I don't know," or "the question doesn't make sense." Of course both theorists can construct formal models to capture these basic propositional *desiderata*, but now we're making my point: That contemporary views are no better off than the classic worlds view when it comes to the most basic propositional requirements.

To recap what I've said in other words, we can think of proposition theory construction as proceeding in two steps.[21] In step one we find some entities, *sui generis* or otherwise, whose existence we accept. I think the modeler we're considering is in at least of a good a spot as King and Soames. Properties, one might think, are less controversial than propositions, so reducing propositions to properties counts as progress since accepting properties into one's ontology is commonplace.[22] In step two, we give good reason to think that those entities are propositions. Whereas historically a lot of the debate was on step one—and perhaps still is for *sui generis* theorists—I think, nowadays, step two is where all the action is, and this is the step where the Benacerraf problem arises. If we are on step two and decide that some entity *e* isn't the proposition that *p*, then we'll think that because *e* doesn't fill the right propositional roles, *e* isn't instrumentally acceptable, or *e* doesn't satisfy empirical criteria if the proposition theory is mind first. If I've been successful in this subsection, then this particular worlds theorist escapes at least the Benacerraf problem. As I noted at the outset, we must also resist the attitude that such a theorist "merely" models/represents propositions instead of telling us what they *really* are. We've said, in this case, what propositions really are. The point just underway was that much of the theoretical payoff occurs at the level of the model—not only for the worlds theorist but also for King and Soames. True, King

---

[21]This is how King (2007, p.25) describes giving an account of propositional structure.

[22]Events and facts, especially facts, are perhaps more controversial than properties. E.g., see the debate between Lewis and Armstrong over states of affairs, with Lewis supporting the view that there only things, not facts. See Bricker (2015) for an overview of the debate.

and Soames have deep stories about how their propositions come to have structure, represent, and have truth conditions. One may find the modeler under discussion to be lacking in this area, but we can't easily dismiss the theory on the grounds of a Benacerraf problem.

Construing world-theoretic devices as models isn't an eccentricity to avoid the Benacerraf problem. Indeed, Paul (2012) argues that modeling is *the* main methodology of metaphysics, the result being that both science and metaphysics share methodology but differ on subject matter. According to the mainstream view of modeling in the philosophy of science literature, the *semantic view*, a theory consists in a family of structures—models—that themselves consists of entities and relations, usually mathematical, along with an interpretation of those structures. Then, the theory is true if, and only if, one of its models is relevantly isomorphic to its target.[23] Such methodology is familiar in science. To use an example from Paul (2012, p.10)[24] biologists model population growth in ecology with the following logistic equation:

$$\frac{dx}{dt} = rx(1 - \frac{x}{k}).$$

The equation itself involves abstract mathematical entities and relations. Paired with an interpretation, the entities represent population density (x) and rate of increase (r). If the model is isomorphic to population growth rates, then the target satisfies the model.

We understand metaphysical models on the semantic approach, and the difference between the scientific case and the metaphysical case is just the subject matter. Paul (2012, p.12) gives the following example:

> [A] theory of composition can be thought of as a class of models of composition relations between parts of composite objects. Consider a

---

[23]Paul (2012, §2).

[24]Which comes from Lloyd (1988).

theory of the composition relation such that some *x*s compose a *y* if and only if the activity of the *x*s constitutes a life. The models we can take to be the theory are structures of abstract objects that represent activity-constituting objects standing in necessitation relations to abstract objects that represent composites or wholes of the activity-constituting objects. The theory is a class of (suitably abstract) models, where these models are isomorphic to various instances of the activity-constituting relations between parts and wholes.

The view that I've given in this subsection fits this mold. The abstract family of structures are sets of possible worlds or functions from worlds to truth values. The target, as we've seen, are world properties. The interpretation connects the set structures to the instantiation of these properties. One structure will be a set that collects worlds that have one thing in common—the instantiation of *p*. The interpretation is especially conspicuous on the function version—for it doesn't really matter *what* we pair the *p* and non-*p* worlds with, so long as we have two distinct elements of the model. True and false are common, but we could use 1 and 0, or, to use Soames' example meant to show that such "propositions" don't really have truth conditions, a dog and a cat. These structures are then isomorphic to various world properties, and we study the world properties and their relations obliquely via the set-theory-powered model. That the interpretation component is essential brings out the absurdity of thinking that world-theoretic devices are themselves intrinsically representational—a point that has sometimes went unrecognized. For example, we find Field (1986, p.101), in a review of Stalnaker's *Inquiry*, saying that

> [I]f we view take as the object of the belief-state [that Caesar crossed the Rubicon] the set of possible worlds in which Caesar crossed the Rubicon. . . that too is a conception of objects of belief as intrinsically representational.

The reason being is that that the set "can be construed as representing the world as such that Caesar crossed the Rubicon." Of course it *can* be construed as he says, but it needn't be. I have a hunch that Stalnaker never took his set-theoretic devices to be intrinsically representational, and Lewis, the other main worlds theorist and

subject of the next subsection, definitely did not on the grounds of his argument against magical ersatzism.

2.3.2 Lewis

Perhaps the most conspicuous worlds theorist, Lewis (1970a/1983, p.193), said, somewhat famously now,

> [i]n order to say what meaning *is*, we may ask first what meaning *does*, and then find something that does that.

This paper ("General Semantics") pre-dates both Lewis (1970b/1983) ("How to Define Theoretical Terms") and Lewis (1972/1999) ("Psychophysical and Theoretical Identifications")—both of which firmed up his methodology. The aspect of Lewis' methodology that we'll be concerned with is the so-called "Canberra plan."

Lewis first articulated a narrow methodology in "How to Define Theoretical Terms," giving a method for defining theoretical terms within the context of philosophy of science that F.P. Ramsey inspired. He then generalized the method to any term-introducing theory—terms that don't necessarily have to be *new*, as is the case with Lewis' well-known generalization of the method to apply to folk psychology. The methodology proceeds in two steps. Step one: Make explicit implicitly defined functional roles for target terms. Here, we gather platitudes if the target term is a folk term, or we gather the sentences from some scientific theory if the term is theoretical. These sentences will contain terms that we want to define and other terms. The outcome of this step is the production of functional roles. Step two: Find non-trivial (i.e. not just the target terms) role fillers that respect our methodological commitments (e.g., for many, naturalism). In other words, we go about finding something that plays the various roles that have been defined in terms of their (usually causal) relations to the other terms.

With respect to propositions, the first step is more fraught than I think enough writers have recognized for reasons I've alluded to above. The worry for the

81

Canberra plan when generalized past the scientific domain is that there is a greater threat that the functional role or roles will be, as we quoted Lewis saying earlier, conflicting *desiderata*. When we apply the Canberra plan to, say, electrons, it's very likely that we'll have a univocal role for some naturalistic entity to play (a trivial one—viz. electrons). On the other end, we may have some folk term—say, the term "snack"—whose platitudes are obscure enough to foil the Canberra plan. Is a snack something you eat between meals? When you're not hungry? Does a snack have to be small? Can't one snack for dinner and snack even though one is full? The right attitude, it seems to me, is, Who knows? We thus need the various roles that propositions play to be univocal enough to find suitable naturalistic realizers.

To apply the first step of the Canberra plan to propositions, we identify the roles they play and determine whether there is some univocal role that admits of a single realizer. The best-case scenario is a single entity. If nothing answers to the roles, then on the one view, the term "proposition," like "phlogiston," fails to refer to anything. Another option would be pluralism, in which different sorts of related entities fulfill various semantic, epistemic, and metaphysical roles. This would occur if the roles are incommensurable, yet each role is itself clear and univocal so as to admit some entity to fill that role. And in such a case, I don't think much will turn on whether we want to call this plurality different kinds of the same entity or distinct kinds of entities all together. The roles that propositions play divide naturally into two categories: the metaphysical roles and the epistemic/semantic roles, and there is, despite received opinion, *prima facie* reason to believe that there isn't one entity type that plays these roles. Perhaps the main reason is this: The metaphysical roles seem to be only sensitive to the facts while the non-metaphysical roles are sensitive to more.

Though Lewis at one time associated propositions with sets of (real) worlds, he is perhaps most well known for identifying propositions, in the role of the objects of

attitudes, with properties. The distinction between the two was never a difference in the *kind* of entity each is but rather which of various roles the entity in question played.[25] For Lewis, sets of things are properties and worlds are just big things. As we saw above, the objects of the attitudes, for Lewis, are properties of the type *being in a world such that thus and so*. And he then analyzes propositional attitudes in terms of self-ascribing such properties. And then Lewis analyzes such self ascriptions in terms of sets of possible individuals. To believe thus and so is to locate oneself among various sets of possible individuals. As we noted above, the content of an assertive sentence will still be a garden variety set of worlds (although for lewis, concrete worlds of course).

The status of Lewis' set-theoretic devices has puzzled some theorists.[26] Are they tools playing a role in an analysis of various philosophical phenomena, or does Lewis actually identify them with properties or propositions? This may sound like an important question since Lewis makes for a prime target of the Benacerraf problem. Yet Lewis has remarkably little to say about the problem, though Lewis (1970a/1983, p.201) does say the following, in the midst of a semantic project:

> It may be disturbing that in our explication of meanings we have made arbitrary choices—for instance, of the order of coordinates in an index. Meanings are meanings—how can we *choose* to construct them in one way rather than another? The objection is a general objection to set-theoretic constructions, so I will not reply to it here. But if it troubles you, you may prefer to say that *real* meanings are *sui generis* entities and that the constructs I call 'meanings' do duty for real meanings because there is a natural one-to-one correspondence between them and the real meanings.

Though here he mentions indices of evaluation, the attitude plausibly extends to any point in an analysis that involves an arbitrary choice among entities. Given

---

[25]This is the case before Lewis (1979).

[26]E.g., Weatherson (2014) is indecisive between whether, e.g., properties *are* or *determine* sets for Lewis.

the extensive use of set theoretic devices, the Benacerraf problem would seemingly extend to many aspects of Lewis' philosophy, including his view of properties, relations, and propositions.

Asking whether Lewis' world-theoretic propositions *really are* propositions or are mere "tools" that represent propositions is a question that doesn't properly take into account Lewis' methodology. Playing a role and "really being" something is a distinction without a difference for Lewis. To apply the Canberra plan to propositions, we first gather the platitudes about propositions. As I noted above, this may be difficult since, *prima facie*, the different propositional roles pull in different directions. Furthermore, the curious status of propositions complicates the straightforward application of the plan. The plan usually has two types of applications. In the first instance, theorists applied the plan to scientific theories, with the roles usually being causal roles that physical entities realized. Later, Lewis applied the plan to folk theories, like folk psychology, but the roles were still causal. Later still, theorists applied the plan to metaphysics and ethics in which the roles weren't obviously causal. The first step in the non-causal scenario is generalized to gathering non-causal platitudes about some term.

Propositions are odd because they straddle the technical-folk border. Some philosophical notions have a clear intuitive and non-theoretical status. E.g., free will, knowledge, belief, desire—these all have a clear intuitive status. Of course the debates about these terms can become technical, and at least the epistemic notions can be used in technical endeavors, but there are a robust set of folk platitudes in these examples from which to start the plan. I have a hunch that much of the talking past that occurs in the propositions literature has to do with the hybrid status of propositions. The folk sources of propositional platitudes all surround the notion of shared content: two sentences from the same or different languages "saying the same thing" or various people having propositional attitudes toward the "same

content." But the folk view of meaning is somewhat of a mess and perhaps should be barred membership from the regimented platitudes that make it into step one of the plan—this, for reasons that Lewis (1980/1998) first noted:

> Unless we give it some special technical meaning, the locution 'what is said' is very far from univocal. It can mean the propositional content, in Stalnaker's sense (horizontal or diagonal). It can mean the exact words. I suspect it can mean almost anything in between.

One of the main complaints against King's theory has been his theory's fineness of grain is far too fine, for any difference of syntax results in difference in propositions. In response, King echoes Lewis:

> There just is no *pre*theoretical notion of a proposition to which our theories of propositions must do justice. (King, 2013, pp.767–78)

If Lewis and King are right, then there will only be fully theoretical roles that we gather in step one. If this correct, then all we need from propositions is that they are the meanings of assertive sentences that result from composing sub-sentential meanings modeled in formal semantics; that together propositions have the correct logical structure; that we can construct the correct domain for a probability function taken as a credence function; etc. Possible states of the world are *one* kind of entity that fill some of these roles.

Supposing that a Lewisian completes step one, the major questions of this paper will be raised on step two. In step one, we will have found a role for coarse-grained propositions. This is in fact the step at which much of the debate will take place—many will object to coarse-grained content on the basis of folk intuitions. But setting that aside, the question—translating the Benacerraf problem to the idiom under discussion—is, What do we do when we have multiple realizers of the platitudes? Both the set and characteristic function version of the worlds theory are exactly equal deservers of the role if they're deserves at all. And further, once we fix upon the roles in step one, the two versions will likely be *perfect* deservers of the role.

Lewis changed his mind about this kind of situation over his career. Initially, Lewis (1970b/1983) required that the roles be uniquely realized. It's important to note that that paper concerned scientific theory where the uniqueness requirement is most plausible. But as Lewis generalized his method, he relaxed the uniqueness requirement. For example Lewis (1984, p.223) says the following about a kind of descriptivism involving "Jack the Ripper," but the points apply to our discussion:

> The description needn't fit perfectly. 'Jack the Ripper' might take as referent the one who comes closest to doing this, that, and the other, if no better candidate is available. The intended interpretation of the augmented language, then, is to be that extension of the old interpretation that comes as close as can be to making the new Jack-the-Ripper theory come true. . . . There might be two candidates that both fit perfectly; more likely, there might be two imperfect candidates with little to choose between them and no stronger candidate to beat them both. If so, we end up with indeterminate reference (in addition to whatever results from indeterminacy of the old interpretation of the old language): the new term refers equally to both candidates.

In our case, we have (at least) two candidates that fit perfectly, and so by Lewis' lights, it's indeterminate which version of the worlds theory we ought to pick. Here, it's important to stress how Lewis construes step one in the plan. His main concern is that there are various (mostly) theoretical roles that we need some entity to play. The result of existentially quantifying over some subset of these role involving sentences (since Lewis is a propositional pluralist) will result in a kind of structuralism.[27] While the mathematical structuralist posits an abstract structure, this structuralism just is the functional roles that step one in the plan produces. It will be indeterminate which version of the worlds theory is *the* witness of the roles, but the point is to provide a class of entities, *tools* for further theorizing. What matters is that the propositions theorist provide anything with the right structure that satisfies step one of the plan. Thus, stipulating that, e.g., the set version is the

---

[27]I am indebted to Phillip Bricker for this point.

one that we'll employ to feed into our formal semantics, or formal epistemology, or whatever, is a benign kind of stipulation—one that the Benacerraf problem doesn't threaten.

CHAPTER 3

THE OBJECTS OF CREDENCE

Abstract

Chalmers (2011a) presents an argument against "referentialism" (and for his own view) that employs Bayesianism. He aims to make progress in a debate over the objects of belief, which seems to be at a standstill between referentialists and non-referentialists. Chalmers' argument, in sketch, is that Bayesianism is incompatible with referentialism, and natural attempts to salvage the theory, Chalmers contends, requires giving up referentialism. Given the power and success of Bayesianism, the incompatibility is *prima facie* evidence against referentialism. In this paper, I review Chalmers' arguments and give some responses on behalf of the referentialist.

3.1   Introduction

Roughly, referentialism is the view that the object of a belief about an individual and property is wholly determined by that individual and property. In a recent paper, Chalmers (2011a)[1] argues that referentialism is incompatible with a successful theory in formal epistemology, Bayesianism, which I explicate below. In particular, he argues that standard Bayesian reasoning breaks down if a bare version of referentialism is true, and any attempt to salvage the core of the bare version requires giving up referentialism. According to Chalmers, the referentialist, then, faces a dilemma: She must either reject Bayesianism or give up her theory—both undesirable options.

---

[1]Chalmers (2006b), which was presented for the Online Philosophy Conference in May 2006, is the predecessor of Chalmers (2011a). In what follows, I will also be citing Braun (2006), which he gave as comments on the Chalmers paper at that conference. I found Braun's paper only after writing the initial draft of this paper, but I have since worked in some of his comments.

Something like referentialism is a view familiar from the literature on Frege's puzzle, a literature that is vast and one in which referentialists[2] are well represented. The referentialist faces some old worries—that, for example, one believes the same proposition in believing that Hesperus is a celestial body and that Phosphorus is a celestial body, which seems false.

In response, some referentialists claim that the very same proposition can be taken in different ways—that agents bear various attitudes toward propositions only under a *guise*. A novelty in Chalmers' paper is his claim that the familiar appeal to guises raises *new* problems, new problems that can break what he views as a stalemate in the Frege's puzzle literature. As I understand him, Chalmers claims that we ought to make the three-way identification among the objects of belief, the objects of credence, and the objects that inhabit the domain of the credence function. Then, he argues that if the referentialist includes her guises in the domain of the credence function, then referentialism is false. My thesis is that Chalmers has not broken the stalemate. Specifically, I argue that Chalmers' second horn—that attempts to salvage referentialism requires giving the theory up—fails. For this thesis, I provide numerous arguments below, each with a similar structure: That while an unadorned version of referentialism indeed conflicts with Bayesianism, the core of the unadorned version can be recovered and built upon, resulting in a theory that does no worse than Chalmers' own theory and doesn't require giving up the core of referentialism. Given that I'm arguing for a stalemate thesis, my *tu quoque* objections are not fallacious. After arguing that the second horn fails, I focus on how I think the thesis of referentialism should properly be understood. Once properly understood, one sees that Chalmers' arguments don't target referentialism but instead conclusions that follow from referentialism packaged with auxiliary

---

[2]Or at least the theorists that Chalmers associates with referentialism. In what follows, I argue that Chalmers mischaracterizes how we should properly understand referentialism.

theses. Once those auxiliary theses are dropped, one sees that Chalmers' arguments miss their mark.

Before moving on, let me make a few terminological and structural points. *Terminological*: Given that one of the conclusions of this paper is that Chalmers mischaracterizes the theorists he intends to argue against, it will benefit us to make some terminological points to avoid confusion. I'll call the thesis that Chalmers calls "referentialism," which I'll quote shortly, "naive referentialism." I'll call the class of theories that Chalmers is attacking on the second horn, theories that employ something like guises, "sophisticated referentialism." And finally, I'll call what I think should properly be described as referentialism, which won't be revealed until the end of this paper, "true referentialism."[3] *Structural*: This paper is structured to mirror Chalmers' dilemma. In §3.3 I canvas the ways in which naive referentialism conflicts with Bayesianism. At this point, I want to stress that we're considering naive referentialism, which doesn't include the notion of a guise. Only in §3.4 does the notion of a guise come into play.

3.2   Background: Bayesianism, Referentialism

I take Bayesianism to comprise at least the following claims:

- A credence function, $cr(\cdot)$, defined over propositions, represents an agent's belief state. A credence function takes a proposition as argument and returns a real number in the interval $[0, 1]$. These real numbers represent an agent's subjective confidence in a proposition, with 0 representing absolute certainty of falsehood and 1 representing absolute certainty of truth.

---

[3] I am intentionally avoiding the term "Millianism" because many Millians employ something like guises, which won't surface until §4.

- A credence function represents both an agent's unconditional credence that a proposition p is true (notated $cr(p)$) and an agent's conditional credence that a proposition p is true given that a proposition q is true (notated $cr(p \mid q)$).

- An agent ought to update her credences, in light of evidence, via *conditionalization*: When an agent receives evidence e, her new credence in any proposition h should be her old credence in $(h \mid e)$.[4]

Probabilism—the constraint that an agent's credence function ought to be a probability function—is often included in the Bayesian package, but I take the constraint to be independent. I soon raise an argument that referentialism conflicts with probabilism, but there I will flag that the argument doesn't target the core of Bayesianism but a claim often conjoined with Bayesianism.

Referentialism, unfortunately, isn't as easily glossed. Here's what Chalmers (2011a, p.587–588) has to say about it:

> Loosely speaking, referentialism about belief says that in so far as beliefs attribute properties to individuals (e.g. the belief that Nietzsche is dead), the objects of these beliefs are wholly determined by those individuals (e.g. Nietzsche) and those properties (e.g. the property of being dead). On one version of referentialism, the objects of belief are Russellian propositions composed from the individuals and properties that one's belief is about. On another version, the objects of belief are sets of possible worlds in which the individuals in question have the relevant properties.

This is what we will call *naive referentialism*. Here are some familiar consequences of naive referentialism:

> If 'a' and 'b' are two names for the same object, then in believing that a has ϕ and in believing that b has ϕ (e.g. in believing that Hesperus is a planet and in believing that Phosphorus is a planet), one believes the same proposition. Likewise, in sincerely asserting 'a has ϕ' ('Hesperus is a planet') and in sincerely asserting 'b has ϕ' ('Phosphorus is a planet'), one expresses belief in the same proposition.

---

[4]For simplicity, and since nothing hangs on it, I set aside other versions of conditionalization, such as Jeffery conditionalization.

Propositions play many roles. We can start by thinking of referentialism about propositions generally and then move to the more specific thesis regarding propositions' role as the objects of belief in particular. I'll call the *core of referentialism* the thesis that a proposition about an individual and a property is wholly determined by that individual and property. As it stands, this thesis makes no predictions that are relevant to Bayesianism. We need to add auxiliary theses regarding the role these propositions play as the objects of belief and credence. Since the core makes no predictions relevant to Bayesianism, it is consistent with Bayesianism. To get *naive referentialism* we take the core, and we add the auxiliary claim that nothing else is relevant to whether an agent believes a proposition so described. In particular, what I have in mind here is the rejection of any non-referential component, like guises, that mediates belief in propositions. *Sophisticated referentialism*, as mentioned, includes guises. Note that naive and sophisticated referentialism agree on the core of referentialism but disagree on the grasping of propositions that satisfy the core. I want to stress that sophisticated referentialism isn't a mere extension of naive referentialism. For if it were, then if the conjunction of naive referentialism and Bayesianism is inconsistent, then so is the conjunction of sophisticated referentialism and Bayesianism. Rather, naive referentialism rejects, while sophisticated referentialism includes, a mediating component, though both agree on the core of referentialism. My thesis is that while naive referentialism is inconsistent with Bayesianism, sophisticated referentialism isn't. And since sophisticated referentialism retains the core of referentialism, it is a referentialist thesis.

3.3   Arguments from Bayesianism against Naive Referentialism

Chalmers sets up his argument as a dilemma: Naive referentialism conflicts with Bayesianism, and sophisticated referentialism's philosophical machinery can't be imported into a Bayesian framework without giving up the core of referentialism.

So the referentialist either has to accept that her theory conflicts with Bayesianism, a powerful and successful theory, or give up her own theory to gain harmony—both unpalatable options. Let's call this "Chalmers' dilemma."

On to the first horn—that Bayesianism and naive referentialism are incompatible. The simplest argument against naive referentialism involves probabilism. Since, in my view, probabilism isn't essential to Bayesianism, this argument is the weakest, but it's also the best warm up to the style of argument Chalmers advances. The conflict arises from an extension of Frege's puzzle from full to partial belief. Consider (1)—

(3.1) Hesperus is Phosphorus.

—which is necessarily true according to naive referentialism since the object of belief that (1) determines is the same as (2), which is necessarily true by anyone's lights:

(3.2) Hesperus is Hesperus.

It's taken as a datum that one's rational credence in (1) can be short of certainty. But, on one understanding of it, a probability axiom requires agents to be certain of necessary truths. Thus, naive referentialism runs afoul of probabilism. But this argument isn't weak only for targeting probabilism. As Chalmers notes, the typical formulation of the relevant probability axiom only requires certainty of logical truths, not all necessary truths, and (1) isn't a logical truth (Chalmers, 2011a, p.599).

Another serious problem is on the horizon, though. (1) and (2), according to naive referentialism, determine the same proposition. That is, insofar as we're concerned with a credence function, (1) and (2) are the same argument. But a rational agent can have differing credences in (1) and (2)—$cr(2) = 1$ and $cr(1) < 1$. This means that the agent wouldn't have a credence *function* at all since one argument is associated with two values.[5] Probabilism aside, naive referentialism conflicts with

---

[5]Braun (2006) makes similar points.

a central aspect of Bayesianism—representing credences with functions, which is essential to the first two bullet points in my characterization of Bayesianism above. Without the apparatus of credence functions, Bayesianism can't even get off the ground.

Chalmers' main grievance is that naive referentialism conflicts with the core of Bayesianism: conditionalization. Here's a quick example.[6] Suppose the facts are as follows: 1/10 of people with P have schizophrenia and 1/5 people with Q have schizophrenia, and 9/10 of people with P and Q have schizophrenia. Now suppose Utterson knows that Jekyll has P but doesn't know that he has Q. Since Utterson knows that Jekyll has P and knows that 1/10 people with P have schizophrenia, his credence in Jekyll having schizophrenia is as follows:

(3.3) cr(Jekyll has schizophrenia)=.1.

If Jekyll has Q, then since he has P and Q, and since 9/10 people with both have schizophrenia, then (4) would give one of Utterson's conditional credences:

(3.4) cr(Jekyll has schizophrenia|Jeckyll has Q)= .9.

We know that "Jekyll" and "Hyde" are two names for the same individual, but Utterson doesn't. Suppose Utterson comes to know that Hyde has Q. According to naive referentialism, the proposition that Hyde has Q is the same proposition that Jekyll has Q. Assuming naive referentialism is true, given the story so far, Utterson should conditionalize:

(3.5) cr_{new}(Jekyll has schizophrenia) = cr_{old}(Jekyll has schizophrenia|Hyde has Q)

$= cr_{old}$(Jekyll has schizophrenia|Jekyll has Q) = .9.

But intuitively, this is wrong. Since Utterson doesn't know that Hyde and Jekyll are the same person, seeing that Hyde has Q isn't evidence for Jekyll having

---

[6]This has been adapted from Chalmers' original example in Chalmers (2006b, §2)

schizophrenia. The culprit here is the claim that "Hyde has Q" determines the same object of belief as "Jekyll has Q," which naive referentialism entails. Thus, naive referentialism must be false if Bayesianism is true.

The second horn of the dilemma, and my responses to it, will concern us for the remainder of this paper.

3.4   Responses to the Second Horn

The second horn of the dilemma purports to block the sophisticated referentialist from importing her account of full belief to a Bayesian setting. This horn begins by identifying the objects of credence with the objects of belief. Then, as the arguments from the first horn show, the objects of credence are determined non-referentially. Then Chalmers invokes a crucial premise—that if the objects of credence, and so the objects of belief, are determined non-referentially, then referentialism is false. This is the premise I dispute, but let's see how this dilemma plays out for actual sophisticated referentialist theories.

It's important to note that all referentialists are sophisticated in some way or other. Naive referentialism has obviously untoward consequences. The dialectic in the Frege's puzzle literature has been to square the simplicity and theoretical considerations[7] that recommend referentialism with the seemingly absurd consequences that follow from it.

This is where guises enter the picture. To block these consequences, as I mentioned, sophisticated referentialists claim that an agent can take the same proposition in different ways—that she can bear an attitude toward a proposition under

---

[7]The theoretical considerations are the direct reference theses that Kripke (1980) advances for proper names, that Kripke (1980) and Putnam (1975/2007) advance for natural kind terms, and that Kaplan (1989/2013) advances for indexicals. The simplicity (and also theoretical) consideration that recommends referentialism is that it adheres to "semantic innocence": the constraint that an expression's reference stays stable across linguistic environments. For more about semantic innocence, see Crimmins and Perry (1989/2007).

one guise and fail to bear it under another. There isn't consensus regarding the nature of these guises, but all sophisticated referentialists see their role as mediating an attitude an agent bears toward a proposition.[8] Furthermore, these propositions, for sophisticated referentialists, are Russellian. A *Russellian* about propositions claims that propositions are structured and, insofar as the propositions are about individuals, those propositions are singular. A proposition is *structured* just in case it contains constituents. A proposition is *singular* just in case it contains objects as direct constituents. As an illustration, in outline, a sophisticated referentialist would characterize Lois Lane's (of the Superman story) situation as follows: Lois takes the very same Russellian proposition—$\langle flies, CK \rangle$, where CK is the person to whom both "Kent" and "Superman" refer and *flies* is the property of flying—in different ways, depending on whether she grasps it under the bespectacled reporter guise or under the superhero guise. Russellianism satisfies the core of referentialism.

Sophisticated referentialists understand this three-place relation of an agent bearing an attitude toward a proposition under a guise in different ways. According to Schiffer's (2007/1992) *hidden indexical theory*, this relation is just belief. Here is the logical form of belief reports according to this theory:

(HIT) $\exists m(\Phi^* m$ & Believes(a, p, m)).

Belief here is a ternary relation between a believer (a), a Russellian proposition (p), and a contextually determined property of propositional modes of presentation to which implicit reference is made (Φ^*),[9] which is the non-referentialist component.

---

[8]There is some ambivalence regarding the nature of guises. According to one specific class of views, guises are mental states (or more specifically sentences in the language of thought). See Braun (1998, 2006) and Crimmins and Perry (1989/2007). One may object to the sophisticated referentialist that the nature of guises is unclear, but that is an old objection, which I am not here to defend against.

[9]Implicit reference to a *property* of propositional guise (a guise that attaches to an entire proposition rather than to propositional constituents) rather than a particular guise is made because in many cases of belief attribution, a belief reporter will not be aware of the specific guise under which the person to whom the belief is being attributed believes the proposition.

According to another sophisticated referentialist, Salmon (1986, p.111), belief is two-place but is analyzed in terms of the ternary relation of something like disposition toward inward assent under a guise:

(SALMON) a believes that p iff[10] $\exists m(a$ grasps p by means of m & BEL$(a, p, m))$.

On both of the above accounts, the object of belief, p, is a Russellian proposition, which satisfies the core of referentialism.

There are important differences between Schiffer- and Salmon-style accounts.[11] For example, the accounts give radically different judgments of the truth values of propositional attitude ascriptions. For this paper, it is important that, as McKay and Nelson (2010) put it, accounts such as Crimmins and Perry (1989/2007)—who offer a version of the hidden indexical theory—"embrace a metaphysics of belief similar to the standard Naive Russellian's [accounts such as Salmon's]. The difference between the views concerns whether or not the information semantically encoded by utterances of belief sentences is sensitive or insensitive to differences in the way a proposition is grasped." Furthermore, referentialists of both stripes have said a great deal about numerous puzzles that might seem to carry over to Chalmers' Bayesian challenge. For example, writers that embrace a Salmon-style account in response to, among other things, Kripke's puzzle have made claims not only about the semantics and metaphysics of belief, but also about rationality. According to such writers, an agent can believe a pair of contradictory propositions and still be rational if she takes the proposition in different ways. It will benefit us to see how the appeal guises works out in the Bayesian machinery. To avoid complication in what follows, let us focus on theories like HIT.

---

[10]Salmon himself says that a believes that p "may be analyzed" as the material I have on the right of the biconditional.

[11]My labels "naive" and "sophisticated" are somewhat non-standard. "Naive" is often associated with Salmon-style accounts, and Schiffer-style accounts are sometimes referred to as "sophisticated" or "contextualist." What is important for me is that they both agree about the nature of propositions.

The natural route for the sophisticated referentialist is to associate credences with propositions and guises. According to this proposal, call it *proposal 1*, each proposition can be associated with a guise under which that proposition is grasped. The domain of the credence function comprises ordered pairs of Russellian propositions and guises under which those propositions are grasped. On this proposal, a credence function, characterized set-theoretically, would look as follows:

$$\{\langle\langle G_{p_1}, p_1\rangle, r_1\rangle, \langle\langle G_{p_2}, p_2\rangle, r_2\rangle, ...\},$$

where G_{p_i} is a guise, p_i a Russellian proposition, and r_i corresponds to a number in the interval $[0, 1]$. Since sophisticated referentialist accounts deliver (or aim to deliver) the correct conditions under which an agent believes a proposition, this route would also give the same results for partial belief.

Chalmers anticipates this strategy and gives a few objections. The first objection is that, on the route we're considering, the domain no longer consists of sets but rather ordered pairs: guise–proposition pairs. On the other hand, Chalmers' own semantics, he claims, fits seamlessly with the Bayesian machinery. In a reply to Braun, Chalmers (2006c) says the following:

> By contrast, on my proposal about the nature of guises [primary intensions], guises are naturally associated with sets of centered worlds . . ., and credences in enriched propositions mirror credences associated with these sets. This allows one to bring to bear the full set-theoretic power of the probabilistic apparatus in the analysis of rational belief. This is not a knock down argument for my view of guises, as a probability assignments can coherently be defined in the absence of the associated set-theoretic apparatus. But doing so loses some important explanatory structure. So if other things being equal, a view that allows set-theoretic analysis is preferable.

Soon I'll return to more of the details of Chalmers' own semantics, on which, over the course of numerous articles, he has spared no detail.[12] Suffice it to say here that

_____

[12]His fullest treatment can be found in Chalmers (2006a).

his objects of credence are non-referential, avoiding the first horn of the dilemma, and are set-theoretic, avoiding the current objection to proposal 1.

The first response to this objection is simple: The probability axioms can be stated in terms of a logic on sentences instead of a set-theoretically.[13] While the absence of set theory might put the proponent of proposal 1 at some disadvantage, the pressure is light. Chalmers (2006c) himself recognizes this, saying of the argument:

> This is not a knockdown argument for my view of guises, as a probability assignments can coherently be defined in the absence of the associated set-theoretic apparatus. But doing so loses some important explanatory structure. So if other things being equal, a view that allows set-theoretic analysis is preferable.

It would be preferable, however, for the referentialist to be at *no* disadvantage. This brings me to the second proposal.

According to *proposal 2*, the referentialist associates her guise-proposition pairs with non-referentially determined sets of worlds. On this proposal, the lines between Chalmers' own account and the referentialist's begin to blur. But this shouldn't surprise. The reason that this shouldn't surprise requires me to say more about Chalmers' semantics and in particular his view of propositional attitude reports, though I will be very brief. On Chalmers' two-dimensional semantics, sentences and sub-sentential components are associated with two different intensions,

---

[13]Weisberg (2011, pp.5–6) suggests, in passing, something similar to my approach I'm giving:

> On the other hand, philosophical considerations can make the sentence-based approach more perspicuous. Consider, for example, the sentences "Superman will save the world today" and "Clark Kent will save the world today". To Lois, these two sentences represent very distinct eventualities, though in fact they correspond to the same possible outcome. On the set-based approach, these two eventualities are represented by the same set, and hence must have the same probability. It seems reasonable, however, for Lois to think that they have different probabilities, which is allowed if we use different atomic [for my account, guise-proposition pairs], *A* and *B*, to represent them.

one referential (secondary intension), the other not (primary intension), where an intension is a function from a possibility to an extension. The referential intensions are functions from metaphysically possible worlds to the entity of the correct type. For the non-referential intensions, replace the metaphysically possible worlds with epistemically possible worlds. The object of belief incorporates both intensions, as Chalmers (2011b, p.6) describes here:

> [T]he enriched intension of a simple expression is an ordered pair of the expression's primary intension and extension. The enriched intension of a complex expression is a structure consisting of the enriched intension of its simple parts (including any unpronounced constituents), structured according to the expression's logical form. The enriched intension of a sentence is its associated enriched proposition.

As an example, "Hesperus is Phosphorus" expresses the following proposition— Where a h' is the primary intension of Hesperus, p' of Phosphorus, $='$ of identity, and where $=$ is the extension of identity, and v the extension of "Hesperus" and "Phosphorus":

$$\langle \langle =', = \rangle, \langle \langle h', v \rangle, \langle p', v \rangle \rangle \rangle.$$

And, as I said, in his theory of propositional attitude ascriptions, primary intensions play a similar role as the sophisticated referentialist's guises. So it shouldn't be surprising that the differences between sophisticated referentialism and Chalmers' account begin to disappear.

I'd like to pause to note that on either proposal, we escape the first horn. To remind ourselves, the problems of the first horn were that naive referentialism conflicted with three aspects of Bayesianism (and a related thesis): probablism, measuring credences with a function, and conditionalization. The first and third problems are dissolved on both responses I've given. Both problems arose from straightforward substitutivity maneuvers. "Hesperus is Phosphorus" is no longer associated with the same guise–proposition pair (or set of worlds) as "Hesperus is Hesperus." Similarly for "Jekyll has Q" and "Hyde has Q." Substitution of

co-referring names doesn't require identity of credences on either option that I've given. The second problem also dissolves: On either response I've given, $cr(\cdot)$ is still a function. Take the Hesperus–Phosphorus example, and consider proposal 1 for focus (proposal 2 solves the function problem as well). We have two sentences that express the same Russellian proposition, viz. $\langle =, \langle v, v \rangle \rangle$. But that proposition is not what is fed into the credence function. Rather, it is an ordered pair of a proposition and a propositional guise. If an agent's credence under a guise in the proposition that Hesperus is Hesperus and the agent's credence under a guise in the proposition that Hesperus in Phosphorus differ, and the agent is rational, then the agent grasps the same proposition under different guises. That means two different guises–proposition pairs are mapped to two different credences.

At this point, two natural objections arise. The first worry for both of my proposals is that the referential material is an idle wheel. For proposal 1, the credence function seems to operate only on the guise member of each ordered pair. The Russellian component seems to be irrelevant to the determination of credences. Call this the "idle wheel objection." Proposal 2 faces a similar worry, though it's not quite as vivid. The referential component seems not to play a role, or a major role, in determining the relevant non-referential set of worlds.

A second worry is that Chalmers can invoke the premise that the objects of credence are the objects of belief, and so the referentialist, it seems, has to face the fact that the objects of credence, and so the objects of belief, are either guise–proposition pairs or non-referentially individuated sets of worlds. Here is what Chalmers (2011a, p.601) says:

> But this view now says that the objects of credence, as we are understanding them, are such ordered pairs. If so, the objects of credence behave in a non-referential way. . ., and referentialism will be false of the objects of credence.

Even the sophisticated referentialists that we have canvassed claim that the object of belief is a Russellian proposition, though some non-referential component mediates that belief. With the premise that the objects of credence are the objects of belief, the sophisticated referentialist seems to be in trouble.

To these objections, Chalmers faces a *tu quoque* response. To the first objection, one of Chalmers' dimensions, the referentialist dimension, is similarly an idle wheel. As we've seen, Chalmers' objects of belief contain a referential component, the secondary intension, and a non-referential component, the primary intension. But the objects of credence for Chalmers are unstructured primary intentions that are associated with entire utterances. The referential dimension plays no role in the determination of the object of credence; thus, Chalmers faces the idle wheel objection as well. And Chalmers' premise that identifies the objects of credence with the objects of belief can similarly be reflected on him to cause trouble. Primary intentions are unstructured as well as non-referential while Chalmers' objects of belief are neither. The object of belief, as we've seen, consists of both referential and non-referential components, and furthermore, it's a structured entity whose constituents are primary and secondary intensions. So Chalmers' own account can't satisfy the premise that the objects of credence are the objects of belief.

One response that Chalmers could give here is to raise the fact that he is a "semantic pluralist." As he says: "It is natural for a two-dimensionalist to be a semantic pluralist, holding that there are many ways to associate expressions and utterances with quasi-semantic values, where different quasi-semantic values play different explanatory roles" (Chalmers, 2011b, p.5). This response is equally open to the sophisticated referentialist, but anyway the issue is orthogonal. The question at hand is whether Bayesian considerations put pressure on the sophisticated referentialist, not whether Bayesian considerations favor semantic pluralism.

The most plausible way to reject the foregoing argument is to notice its implicit premise: That the objects that inhabit the domain of the credence function are the objects of credence—and so the objects of belief. I accept the identification of the objects of credence and the objects of belief. When I believe that it's raining, doubt that it's raining, think it likely that it's raining, etc., I bear those various attitudes toward the same object. But I see no reason to grant the assumption that the objects of credence are the objects that inhabit the domain of the credence function. The domain of a credence function is a *technical* notion. For the sophisticated referentialists considered so far, the objects of partial and full belief are Russellian propositions, though a non-referentialist component plays a mediating role. That the domain of the credence function includes a non-referentialist component doesn't impugn the thesis of referentialism unless we grant what I see as an implausible assumption. There are various technical reasons for characterizing the domain of the credence function set theoretically, but that shouldn't commit us all to the view that the objects of credence, and so the objects of belief, are structureless.

A nearby premise, however, might raise trouble for the sophisticated referentialist: Though the objects of credence, the objects of belief, and the domain of the credence function needn't be identical, the objects of credence and the domain of the credence function should share important properties in common, including the property of behaving referentially or not. Thus, since the domain of the credence function behaves non-referentially, then the objects of credence behave non-referentially. Since I think the sophisticated referentialist shouldn't give up the identification of the objects of belief with the objects of credence, we ought to conclude that the objects of belief behave non-referentially. Chalmers (2011a, p.608) sums up the argument this way, where "surrogate" applies to both proposals I've given:

> If the surrogates behave referentially, then they are inadequate to serve as objects of credence. If the surrogates do not behave referentially,

then in so far as they are objects of credence, referentialism is false. It follows that if the objects of credence are surrogates, referentialism is false.

In my estimation, this version of the second horn of Chalmers' dilemma is its strongest form.

But the second horn rests on a crucial confusion. The confusion is the following:

(3.6) While it is true that Russellian propositions are individuated referentially, the conditions under which they are believed—and derivatively the conditions under which agents have partial belief in them—should behave non-referentially since a non-referential component mediates belief in these referentially individuated propositions. Another way of putting this is that whether an agent believes a proposition in a given context doesn't depend wholly on the referentially individuated proposition that is up for belief.[14]

Everyone working on Frege's puzzle is trying to square our Fregean intuitions with various philosophical results that have developed in the last few decades. The sophisticated referentialist has to explain, or explain away, how it can be that Lois believes that Superman flies while she believes that Kent doesn't. Both the referentialist and Chalmers are trying to account for the data, albeit through different routes—though how different the routes are is up to question.[15] Chalmers' second horn stems from confusing the individuation conditions for propositions and the conditions under which those propositions are partially and fully believed. The core of referentialism is a starting point on which the sophisticated referentialist builds to accommodate Fregean intuitions. Chalmers' requirement that sophisticated referentialists are committed to partial belief conditions that behave referentially amounts

---

[14]I have phrased this in terms of HIT-style accounts. For a Salmon-style account, the conditions under which agents have a disposition to inwardly assent to a proposition when taken a certain way—and so a disposition to partially inwardly assent—is determined non-referentially.

[15]This is especially clear when we compare Chalmers' account with HIT.

to the requirement that a referentialist theory must accept naive referentialism as the whole story.

To clarify this point, consider the analogy between Chalmers' second horn, which regards partial belief, with full belief. The compressed version of Chalmers' second horn is this one-premise argument: Partial belief behaves non-referentially; therefore, Referentialism is false. The analogous argument for full belief would be:

(3.7) Belief behaves non-referentially.

(3.8) Therefore, Referentialism is false.

But no sophisticated referentialist should accept this argument. As I noted, the sophisticated referentialist theories presented in this paper should explain our Fregean intuitions regarding belief and credence. That is, Lois will believe that Superman flies, she won't believe that Kent flies,[16] it will sometimes be rational for an agent to believe that Hesperus isn't Phosphorus while also believing that Hesperus is Hesperus, and so on.

While the sophisticated referentialist accepts the core of referentialism, she is in no way obligated to accept the absurd consequences that follow from this core plus the claim that nothing else is relevant to propositions as the objects of belief. Or more cautiously: Suppose you do think that the argument from (7) to (8) is a good argument, and you do think that any referentialist theory ought to be rejected because its only true form is naive referentialism, and naive referentialism is absurd. This is a position one may take, but one doesn't get any mileage from recasting the argument in terms of credences. For if you think it is enough to refute referentialism by showing that full belief behaves non-referentially, then one would naturally expect partial belief to also behave non-referentially. And

---

[16]To put it again in terms of a Salmon-style account: Lois is disposed toward inward assent to the proposition that Superman flies when that proposition is presented in the Superman way and not when it is presented in the Kent way.

recall that my goal has only been to show that recasting Frege's puzzle in terms of Bayesianism doesn't award either side of the debate ground. Chalmers' target seems to be sophisticated referentialists of the type I've considered, viz. versions that are committed to Russellian propositions, but as we've seen these arguments rest on a confusion. Therefore, if I'm correct in this, then the stalemate resumes.[17]

There is yet a larger confusion in Chalmers' argument, which brings me to my final point: The referentialist needn't even be committed to the claim that the objects of belief are individuated referentially. What should properly be described as referentialism—or *true referentialism*—needn't even be committed to the core of referentialism. As it turns out, the core of referentialism itself is a cluster of views that needn't come in a package. And once separated, we can see that Chalmers' arguments miss the mark.

Sophisticated referentialists tend to claim that the objects of belief are individuated referentially, though belief is in some way mediated by a non-referential component. But that is an artifact of the majority of sophisticated referentialists in the literature employing Russellian propositions. And indeed, Russellian propositions are individuated referentially: The proposition that Hesperus is a celestial body and the proposition that Phosphorus is a celestial body are one and the same proposition, $\langle c, v \rangle$, where c is the property of being a celestial body and v is Venus. But here is a simple yet crucial point: The referentialist needn't be a Russellian. And that fact brings out what referentialism properly concerns. True referentialism properly concerns how individuals relate to propositions that are about those individuals. Put semantically, true referentialism properly concerns the nature of the

---

[17]At this point, Chalmers might object that referentialism is, by definition, the thesis that belief and credence behave referentially. Indeed, he may hold this view: He classifies Salmon as a referentialist about language but a non-referentialist about belief (Chalmers, 2011a, fn.9). If Salmon ends up not being a thoroughgoing referentialist about both language and belief, then, although I understand the thesis that Chalmers is attacking, I'm less clear who holds it, for he gives no representative citations in his explication of the thesis.

contribution that certain pieces of language make to the proposition expressed. And for structured propositionalists, true referentialism is a thesis about the nature of some propositional constituents—that individuals are direct constituents. But that is, for at least the structured propositionalist, one piece of the account. On certain views of the nature of the *structure* of propositions, Chalmers' initial characterization of naive referentialism and the core of referentialism are straightforwardly false. To remind ourselves, Chalmers characterizes naive referentialism as follows: "referentialism about belief says that in so far as beliefs attribute properties to individuals (e.g. the belief that Nietzsche is dead), the objects of these beliefs are wholly determined by those individuals (e.g. Nietzsche) and those properties (e.g. the property of being dead)."

To illustrate my point, consider the following:

(3.9) Bob Dylan wrote "Blowin' in the Wind."

(3.10) "Blowin' in the Wind" was written by Bob Dylan.

According to the core of referentialism (9) and (10) determine the same propositions. And according to the way Chalmers has characterized naive referentialism, since nothing else is relevant to propositions in their capacity as the objects of belief, both (9) and (10) determine the same object of belief. And indeed, according to most sophisticated referentialists, being Russellians, this is true. It's plausible that both express the following Russellian proposition:

$$\langle w, BD \rangle,$$

where w is the property of having written "Blowin' in the Wind" and BD is the person that both "Bob Dylan" and "Robert Zimmerman" name. However, consider one the most detailed accounts of structured propositions on the market—Jeff

King's—a theory I consider for illustrative purposes, though I want to note that I'm not here endorsing the theory.[18] King adheres to what I think should properly be deemed true referentialism: *names, indexical, and demonstratives contribute the individuals they designate in a context to the proposition expressed*.[19] According to King, however, (9) and (10) express different propositions. King's account is detailed and rich, but for our purpose it will do to say that propositions are complex facts that involve semantic and syntactic relations. These complex facts are constituted by lexical items standing in semantic relations to their semantic values, and these lexical items stand in a complex syntactic relation. And that syntactic relation itself encodes a semantic function that instructs the semantics to interpret syntactic concatenation as the instantiation function. To illustrate, the following sentence—

(3.11) Rachel reads.

—expresses the proposition that is the following complex fact, though a mouthful to say: There are lexical items x and y of some language L occurring at the left and right terminal nodes, respectively, of the sentential relation R that in L encodes the instantiation function, where the semantic value of x is Rachel and the semantic value of y is the property of being a reader.[20] According to King, that complex fact that is the proposition that "Rachel reads" expresses is true just in case Rachel reads—that is, just in case the ordinary fact of Rachel reading obtains. Thus, a difference in the complex syntactic relation R is sufficient for a difference in the proposition expressed. That (9) and (10) differ syntactically is sufficient for each

---

[18]His fullest treatment can be found in King (2007).

[19]See King (2007, Ch1). To put the thesis non-semantically, propositions contain individuals as direct constituents, a thesis that Russellianism includes but is not exhausted by.

[20]I'm leaving off some details of King's account that have to do with indexicals because that aspect of his account isn't relevant to our discussion.

to express a different proposition.[21] Thus, the core of referentialism and naive referentialism is false for King.

For at least one other referentialist, Soames, similar considerations obtain.[22] On Soames' recent account, propositions' representational capacity derives from the representational capacity of agents. According to this account, to entertain a proposition is to engage in a cognitive activity—the mental act of predication, which for Soames is a primitive mental act, basic among others. When an agent sees o as red, there is an event token of her predicating redness to the object o. Now of course this event token can't be the proposition that o is red. Instead, Soames takes the cognitive event *type* of predicating redness to o to be the proposition that o is red expresses. And more complex propositions and differing propositional attitudes are built upon the mental event type of predication. Regarding other attitudes, some examples: To judge that o is red is to predicate redness to o and then to endorse that predication; to believe that o is red is to predicate redness to o and to be disposed to judge that o is red (and so to be disposed to endorse the predication of redness to o). Regarding more complex propositions, an example: To entertain the proposition that o isn't red is first to predicate redness to o and then to predicate not being true to the result of that first predication.

Soames still qualifies as a structured propositionalist in the sense that the cognitive event types that are propositions contain parts. And Soames, like King, is a true referentialist. But this doesn't commit him to the claim that propositions are individuated referentially. This is because the verb *predicate*, for Soames, is an intensional transitive verb in the way that *look for* is. Lois can be looking for Kent without looking for Superman even though Kent and Superman are the same

---

[21]This extreme fineness of grain that results from syntactic differences has been the source of some criticism. For a critical discussion, see Collins (2007).

[22]See Soames (2010) and Soames (2011).

person. Thus, on Soames' account, "Tully shaved Cicero," "Cicero shaved himself," and "Cicero shaved Cicero" all express different propositions even though the semantic contribution of the subject and object of each sentence is the same.

As I've said, I'm not here to evaluate Soames' and King's accounts. The discussion under way illustrates what referentialism properly concerns. Chalmers can insist that the thesis he's attacking is the one he describes in his quote that we have called "naive referentialism." But that would be a mistake: King and especially Soames are well-known supporters of what should properly be deemed referentialism. What Chalmers ended up attacking was a thesis that followed from referentialism plus other theses, specifically Russellianism. The two recent theories of propositions from King and Soames show that, once Russellianism is dropped, naive referentialism, as Chalmers describes it, can be separated from true referentialism.[23] And this brings to light the real nature of referentialism, which Chalmers' Bayesian arguments leave untouched.

Chalmers hasn't scored a point for the Fregean. However, his novel focus on an uncharted interface has brought to light issues that deserve further exploration.

---

[23]One may find these two recent theories of propositions implausible and think that true referentialism ought to be packaged with Russellianism, but that would require much additional argumentation that Bayesianism would likely be irrelevant to.

CHAPTER 4

FIRST- AND THIRD-PERSON EPISTEMIC MODALS

Abstract

I introduce an epistemic modal distinction—the first-person–third-person distinction—which, in addition to being interesting in its own right, helps explain epistemic modal disagreement. In particular, I claim that there are two different types of information sources involved in epistemically modalized propositions. One information source is a first-person epistemic state, or a group of epistemic states; another is a third-person, external source of information. This distinction helps make sense of felicitous and infelicitous responses in epistemic modal disagreement cases, which I go through in some detail.

4.1 Introduction

"Jones might be home. His light is on."—a typical use of the modal "might" in its epistemic sense. This small word has by now generated a vast literature in philosophy and linguistics.[1] Parties to the literature agree that such uses have something to do with knowledge and information, but writers dispute how these notions play out in the semantics. According to the orthodox theory of epistemic modals, contextualism, the proposition that an epistemically modalized sentence expresses can vary with context. Consider the lead sentence: I or we possess some information (including, among other things, that the light is on in Jones' apartment), and the sentence is true just in case this information doesn't rule out the proposition that Jones is home. Contextualism's motivating idea, then, is that context (often) supplies the information that an unmodalized proposition is compatible with if

---

[1]The literature on epistemic modals is vast. For a representative sample of the literature, mostly from philosophers but including some linguists, see Egan and Weatherson (2009).

the sentence is true.[2] We'll call an epistemically modalized sentence an *epistemic sentence*; the information involved in the epistemic sentence simply the *information*;[3] and the proposition that the information purportedly doesn't rule out the *prejacent*.

Heterodox theorists, relativists in particular, claim contextualism can't make sense of the ways we react to challenges to our epistemic sentences.[4] If someone with more information than I (or we) challenges my epistemic sentence regarding Jones' whereabouts, the natural response is for me to *fold*—i.e., to admit I was wrong instead of *digging in* and insisting what I said was true. Yet, as the objection goes, according to contextualism, my initial assertion was true—at the time I said it, the information was consistent with Jones' being home. Writers have taken such evidence to shed light on the *semantics* of epistemic modals.

As the literature developed, writers introduced numerous complications. To name a few: The knowledge that determines the information can be group or solipsistic, digging in can sometimes be felicitous, and the prejacent rather than the whole epistemic sentence can be the disagreement target. Regarding just the last member of that list, von Fintel and Gillies (2008, p.83) say that "[b]efore resorting to the CIA's [the relativist's] extraordinary measures, we would like to see it carefully argued that flexibility in the target of denials and acceptances does not explain what needs explaining."

In this paper, I give a contextualist friendly solution to the puzzle that will involve two pieces. The first is to take von Fintel and Gillies' point and carefully go through the data, but with respect to all of the complications. The second

---

[2]The version of contextualism I assume in this paper is simpleminded, mostly because it leaves out the ordering source. I choose this version to keep things simple since the additions to make contextualism sophisticated don't affect what I say in this paper.

[3]*Knowledge* is often used here, but as we'll soon see, epistemic modals involve more than just knowledge, which is why I use information instead.

[4]See, in particular, John MacFarlane's work (e.g. MacFarlane (2011)) and Egan et al. (2005) for influential relativist works that also critique contextualism.

is to introduce a novel distinction that I put to work in explaining some of the cases. I distinguish two different information sources that the prejacent is asserted against in an epistemic sentence. On the one hand, the source that determines the information may be first-person knowledge; on the other, the source that determines the information may be third-person external information.[5] This distinction, which I think is best understood as non-semantic, helps to make sense of the disagreement data in the recent epistemic modals literature. One way of thinking of (part of) my solution is that we may appeal to non-semantic features of disagreements to make sense of them. And who doesn't enjoy at least an attempted non-semantic solution to a first-blush semantic quandary?

This paper will be fairly non-technical. I'm not going to map out all the various semantic views on epistemic modality.[6] In what follows, I assume that epistemic sentences express propositions, and that these propositions determine or are truth conditions. In §4.2 I make the first–third-person distinction. In §4.3, I give a complete catalogue of the data and explain some of the easier cases. In §4.4, I focus on the harder cases and draw some big-picture lessons. To close, in §4.5, I consider the first–third-person distinction as such.

4.2 The First and Third Person

In this section, I distinguish the first and third person and draw some consequences of the distinction.

To start, the distinction. The first two sentences of this paper—"Jones might be home. His light is on"—constitute a typical epistemic-modal-involving situation. In such a situation, I'm not sure, or we're not sure, whether Jones is home. What

---

[5]I state the distinction in terms of different sources that determine the information rather than different information *simpliciter*. I address this difference at the end of the paper.

[6]See Schaffer (2011) for a nice taxonomy of truth-conditional views.

I or we know doesn't rule it out, so I say that Jones might be home. The truth conditions, according to the simple analysis that I'll assume in this paper are that the information source, in this case *knowledge*, leaves open the prejacent—the proposition that Jones is in his office (at the time in question). Let me highlight some features of the information source in this case in order to contrast with another shortly. First, the information source is epistemic—knowledge. Second, the information source is *dynamic*. What I mean by this is that the information source is part of a system that changes all of the time, which I update, in the *epistemic* sense, as I find my way through the world. Third, I have a kind of commitment to the information source such that it doesn't make sense to say that I could disbelieve or be indifferent to it. Fourth, the information source is *perspectival* because I consider it part of my perspective. Because of the perspectival, epistemic nature of the information source in this situation, as well as similar situations, we'll say that such epistemic sentences are *first person*, or *in* the first person.

There is a very different kind of epistemic-modal involving situation. Let's say a while ago I opened up the *Fox News* website. Various articles claim that the available information leaves open whether President Obama was born in the United States. I say that according to *Fox*, or given what *Fox* says, Obama might have been born outside the US. I say this while I of course know that President Obama was born in Hawaii. The information source in this case contrasts with the above case. First, the information source is external and non-epistemic. Second, the information source is *static*. What I mean by this is that the *Fox News* website, although it may change, contains static information that isn't updatable in the *epistemic* sense. Third, I don't believe the information source, and in cases like this there are a range of attitudes I can take: I may disbelieve, or be indifferent, or believe and know it. Such flexibility isn't available in the contrast case and in similar cases. Fourth, unless we're speaking loosely, the information source isn't perspectival. Of course we

can say things like "From Fox's perspective," but what I think we ultimately mean involves the perspectives of *people* at Fox, not the "perspective" of the content on the website. Because of the non-perspectival, non-epistemic nature of this information source in this situation, as well as similar situations, we'll say that such epistemic sentences are *third person*, or *in* the third person.[7,8]

We find an analogous situation to the above two, one we're perhaps more familiar with, with arguments. There are at least two different ways we view the premises, the conclusion, and the inferential relations between them in an argument. Sometimes the premises are *ours*—we consider them our information and as such we believe and often know them and we draw conclusions from that information. In such a situation, the information that serves as the premises is epistemic, it's dynamic, and it doesn't make sense to say that we're indifferent to or disbelieve it. This situation is analogous to the first-person. But we—and here I mean philosophers—often find ourselves in a different situation: that of mapping logical space. When we consider and give arguments amidst such mapping, we're most interested in the inferential relations between the premises and conclusion, and we're often indifferent toward the truth of the premises and conclusion (or perhaps

---

[7]von Fintel and Gillies (2011, p.112, fn. 9) note that epistemic modals can be relativized to any source of intensional content. Kratzer's (2012/1981) examples also make it clear that epistemic modals don't only involve epistemic states. I see this paper as working out the differences in commitment of these insights.

[8]The person distinction might bring to mind a notion that Lasersohn (2005) and Stephenson (2007) have put to work in their accounts of epistemic modals and predicates of personal taste—that of the "autocentric" and "exocentric" positions. These accounts invoke the notion of an index supplemented with a judge parameter. In one way of going about it, a judge parameter is added to the index of evaluation, and an epistemic modal is true (relative to a world and context and possibly more parameters) just in case the judge's beliefs leave the truth of the prejacent open. While it is often the case that the judge is the speaker—the autocentric stance—we can empathize with another's view of the truth of an epistemic modal or predicate of personal taste—the exocentric stance. This distinction can be put to use in various ways, e.g. shifting the judge to the subject of the matrix clause in an attitude report on someone else's epistemic modal or predicate of personal taste claim. Note that this distinction is not the same distinction that I am making. The first- and third-person involve different sources of information, whether of one's own or not, and is not limited to the tastes and information of agents. Also, their distinctions are explicitly semantic.

hostile). In this situation, the information that serves as the premises isn't epistemic, it's *static*, and we're, as noted, often indifferent about the information's veracity. E.g., as we read some paper, we consider what propositions a philosopher's premises or commitments leave open or closed. These premises and commitments aren't *ours* and they're not changeable. We don't "update" them in the same way we update our beliefs. This view is analogous to the second case that involved a third-person epistemic sentence.

It's no surprise that we find an analogy between arguments and epistemic modals since the standard account of epistemic modality is built on, as Kratzer (2012/1977, p.1) says, "a mechanism for drawing conclusions from premises." The upshot of this distinction, however, is surprising, and I think underappreciated: That there is nothing necessarily *epistemic* about epistemic modals since the information source can be knowledge or a non-epistemic information source. In other words, different types of information sources determine the information involved in an epistemic sentence, and while it seems that knowledge is the norm, non-epistemic sources can determine the information. What's important for this paper is that the first-person–third-person distinction tracks differences in the *commitment* we have to our assertions. I further consider the nature of the first-person–third-person distinction at the end of the paper.

The idea of commitment is familiar in the pragmatic notion of *speaker orientation*. A clause with some denotation in an utterance is *speaker oriented* just in case the speaker commits to that denotation in speaking that utterance—else, it's non-speaker oriented.[9] We're familiar with the idea's straightforward application: I say that according to Bush, Iraq has WMDs. The embedded clause is non-speaker oriented, so I don't commit to the proposition it expresses. On the other hand,

---

[9]This is a definition from Harris and Potts (2009, p.524).

if I say outright that Iraq has WMDs, I commit to the proposition that sentence expresses.[10]

Commitment is also involved in epistemic modals, but the difference in commitment between epistemic sentences in the first- and third-person is more subtle. I want to develop the idea over the course of the next section, but the basic idea is that the third person allows a kind of backing away from an assertion upon learning that the prejacent is false whereas, at least sometimes, that option is extremely odd in the first person. This is because, as mentioned, in the third person, we may only be interested in the compatibility of the information and the prejacent yet not the veracity of either. This difference in commitment will help explain the data in the next section.

We've already seen in the *Fox* example that we can use an adverbial modifier to dial in exactly what the information is. Thus, though contextualism is so called because context *often* supplies the information, it doesn't always. Complications arise here, though, with the choice of adverbial modifiers: *given. . .*, it might be that. *. .*; *according to. . .*, it might be that. *. .*; *in light of . . .*, it might be that . . .; etc. One of Kratzer's initial motivations for contextualism was to capture the neutral meaning of modals when an adverbial phrase makes the flavor of the modal (i.e. deontic, bouletic, circumstantial, epistemic, etc.) explicit.[11] Adverbial modifiers, with deontic modals, can reference this or that set of rules; with circumstantial modals, can reference this or that set of circumstances; and with epistemic modals, can reference this or that information source, *whether it be first or third person*.[12] But there

[10]The distinction would seem uninteresting if not for surprising applications in contexts involving, e.g., appositives and expressives in which material remains speaker oriented when we might expect a shift to non-speaker orientation. See Potts (2005) for the view that appositives and expressives are invariably speaker oriented and Amaral et al. (2007) for a critique.

[11]See Kratzer (2012/1977).

[12]This point militates against an objection that has been made to me regarding the first–third person distinction: that, perhaps, in the third person, the information involved is really the speaker's

is a complication that Kratzer notes in a new introduction to Kratzer (2012/1981): Different adverbial modifiers come with different strengths of commitment. I've chosen "given." What matters for my project is that there is a way to make salient a third-person information source, and that one can assert such third-person epistemic sentences while being agnostic about the prejacent (or even thinking it false). Kratzer (and others in conversation) seem to think that "given" commits the speaker to the veracity of the content of the information, but I don't think this is right. A quick example: I have three little cousins, Leila, Posie, and Alabama, ages five, four, and one, respectively. There is a vase broken in the room, and I'm trying to find out who, if anyone, did it. Leila and Alabama have been getting along, but Alabama has been getting on Posie's bad side. Neither Leila nor Posie know for sure who broke the vase. Posie first tells me that Alabama was in the room when the vase broke. Leila, coming to Alabama's defense, tells me that Alabama was in her crib. I then say: "Given what Posie said, Alabama might have broken the vase, but given what Leila said, Alabama couldn't have."[13] I didn't commit myself to the veracity of their sayings or to either modal.

---

knowledge, or the group's knowledge, *about* the third-person information source. But as I said, just as we've taken adverbial modifiers, in other flavors of modals, to reference some particular set of rules, circumstances, or desires, etc., I think it's reasonable that we can reference, in an epistemic modal, some particular intensional content as well, regardless of its internal or external status. And I think it goes without saying that we can speak directly about what propositions some external source of intensional content leaves open or closed without discussing second order claims about our knowledge of that content. Once we admit that, there is nothing in the way of external intensional content determining the information directly in an epistemic modal.

[13]Note that my assertion of the conjunction of epistemic sentences doesn't involve any kind of reporting. It's not a case in which I assert one person's report and then assert another's and then ask, Who am I to believe? Neither little cousin *said* that Alabama might have/couldn't have broken the vase. Rather, the content of their sayings left open/closed off Alabama's breaking the vase.

4.3   Zooming in

In this section, I clarify the target of my explanation, distinguish data and explanation, clarify my goal and strategy, and then consider the total number of cases to explain before proceeding, finally, to concrete cases.

My explanation's target. I narrowly aim to explain when and why it is either felicitous or infelicitous to dig in or fold when a conversational participant (in some sense) disagrees with one's epistemic sentence. My explanation's target simply consists of numerous dialogues in which a speaker issues an epistemic sentence, another speaker in the conversation (in some sense) disagrees, and the initial speaker either digs in or folds—along with a judgement about the felicity of that response, which I hope I am correct in and that the reader shares. The number of such dialogues depends on the features that we can dial in, which, in addition to the just-introduced first–third-person distinction, includes three complications (that I mentioned in the introduction) that writers in the literature have introduced. First, there is the so-called group versus solipsistic distinction. When an epistemic sentence is first person, the information can belong only to the speaker of that sentence, the solipsistic case, or, in the group case, to some group, which may include the conversational participants and possibly others. Thankfully, we can clearly distinguish between the two readings with adverbial modifiers, either "Given what I know, . . ." or "Given what we know, ..." Second, as I've quoted von Fintel and Gillies as noting above, there are complications regarding in what sense a conversational participant disagrees with an epistemic sentence. It seems in some cases, one may disagree with an entire epistemic sentence, and in others one may only target the prejacent. Third, although the focus in the literature initially was on the infelicity of digging in, we now recognize that digging in is sometimes felicitous. Last—and this is just an obvious feature that I appeal to rather

119

than an interesting complication—either the initial's speaker's epistemic sentence will be true or false, according to orthodoxy, when she initially uttered it.

To sum up and to be clear, let me state the parameters exactly. Each dialogue will be of the following form:

(4.1) *Disagreement Dialogue Form*

- (a) (Speaker 1:) An epistemic sentence

- (b) (Speaker 2:) Disagreement, in some sense, with that epistemic sentence

- (c) (Speaker 1:) Either (i) or (ii) along with a felicity judgement:

  - (i) Digging in

  - (ii) Folding

In these dialogues, we'll always assume that the disagreer doesn't have the facts wrong.

Here are the parameters along with their respective abbreviations that I'll use throughout the paper in parentheses:

- First vs. third person (1 vs. 3): Classify a disagreement dialogue as first/third person just in case the sentence (1a) is in first/third person.

- Group vs. solipsistic (G vs. S): Classify a disagreement dialogue as group/solipsistic just in case sentence (1a) is in the first person and group/solipsistic.

- Prejacent vs. modal (P vs. M): Classify a disagreement dialogue as prejacent/modal just in case the disagreer's sentence (1b) targets the prejacent/modal.

- True vs. false (T vs. F) initial epistemic sentence: Classify a disagreement dialogue as true/false just in case the initial epistemic sentence in (1a) is true/false by the contextualist's lights.

120

- Infelicitous vs. felicitous ($\mathscr{I}$ vs. $\mathscr{F}$): Classify a disagreement dialogue as felicitous just in case (1ci) is felicitous (and (1cii) is infelicitous) and infelicitous just in case (1cii) is felicitous (and (1ci) is infelicitous).

Data vs. explanation. It's important to note that, inevitably, I mix data, explanation, and neutral features of the dialogues in the upcoming cases. Data, I take it, are features of the dialogues that we agree about but call for explanation. The felicity judgements are the sole data point for this paper: *That* a response is felicitous/infelicitous will be uncontroversial (again, I hope), but *why* it is so is puzzling, at least given orthodoxy. Then there are aspects of the dialogues that are neutral in that they are neither puzzling nor in need of explanation. The other four parameters fit into this category in the sense that I take a dialogue's membership in a given category to be neither controversial nor puzzling. Last, there are the explanatory aspects of the dialogues, which include the *relationship* between the first four parameters and the last. Showing how the first four parameters bear on the last is my main argumentative task.

The goal and strategy. The goal ultimately is to explain when and why one can felicitously reiterate an epistemic sentence in the face of challenge and when and why one cannot. An utterance's status as infelicitous doesn't mean that that sentence is false—just that it is defective in some way. One reason that it's infelicitous to dig in is simply because one's epistemic sentence was false at the initial time of utterance. It should be uncontroversial that you cannot utter a falsehood, have it shown as such, and then insist that the sentence was true when you first said it. But another reason for infelicity, as we'll see, is that one's epistemic sentence at the initial time of utterance was true, yet for some reason it's odd to reiterate the sentence upon challenge. On the other hand, it may be that one's initial epistemic sentence was true at the time of utterance, yet it's not odd to reiterate the sentence upon challenge. I'll argue that all of the cases in which it's felicitous to dig in will

fit into this category. Thus the goal is to explain these felicity intuitions, and the strategy is to use the resources above to show that each dialogue fits into one of these three categories:

- False initial epistemic sentence + infelicitous digging in because of falsehood; or

- True initial epistemic sentence + infelicitous digging in because of inappropriateness; or

- True initial epistemic sentence + felicitous digging in because of lack of inappropriateness.

The number of cases. As we've noted, the group–solipsistic parameter doesn't combine with the third-person parameter. Taking that into account, there are twenty-four cases—an enormous amount. Fortunately, we can make some observations that will rule out certain cases. Specifically, two observations rule out certain parameter combinations: first, the stipulation that the challenger isn't mistaken about whatever it is that she is challenging, and second, that one cannot reiterate a false epistemic sentence once one has been shown it is false.

- Rule out the M–T combination: If the challenger targets the entire epistemic sentence, then that sentence couldn't have been true since the challenger isn't mistaken. In other words, because the challenger isn't wrong, M must go with F, so M can't go with T.

- Rule out the F–$\mathscr{F}$ combination: If the challenger shows the initial speaker's modal to be false, then she can't reiterate that modal. In other words, $\mathscr{F}$ must go with T, so $\mathscr{F}$ can't go with F.

- Rule out the M–$\mathscr{F}$ combination: If the challenger targets the modal, then that modal must be false, and, as we've just seen, if the modal is false, it can't be reiterated. In other words, M must go with $\mathscr{I}$, so M can't go with $\mathscr{F}$.

That cuts down the number of possible cases to twelve.

The distinctions grow almost unmanageable, and it may help the reader, as it helps me, to keep track of them visually. At this point, I refer the reader to the end of the chapter, which contains a table with all of the possible combinations of the parameters. The top row contains the parameters and the left-most column numbers the rows. The white rows are the cases we just ruled out, and the last column contains the parameter combination that rules that kind of case out. The grey rows are the live case kinds that we need to explain. Those cases have names for concrete instances that we provide in the next section.

4.3.1 The cases

We can present the cases is different orders. Herein, I present them in terms of ease of argument, starting with easy cases and ramping up. This is in some ways less orderly—e.g., we could go through all the first-person cases and then the third—but that would require some heavier before light lifting. When a case is important—an illustrative case that combines the above parameters—I'll name it and put the parameter combination and row next to the name. We'll start with the F cases and then move to the T cases since the F cases have an easy explanation—false initial utterance leads to infelicity. Going through the cases isn't just fastidious completeness, though. It's illustrative to see, against what one might think, how sparse such cases in fact are.

4.3.1.1 F cases

The easiest cases are those that fit the paradigm of disagreement wherein one asserts p, another challenges p, and then, since we've stipulated the challenger is

correct, the initial speaker retracts p. In terms of our parameters, this means that we set the M/P dial to M so that the disagreer targets the initial speaker's entire initial assertion. Since the disagreer doesn't have the facts wrong, other parameters will fall into place—the initial assertion will be false, and because of that, digging in will be infelicitous. Again, adverbial phrases will come to our aid to assure that the entire utterance, and not the prejacent, is the disagreement target. Let's start with the third person case:

(4.2) Row 22—*Logic-Modal*: 3$\mathscr{I}$MF

Context: A logic student and teacher are discussing what is left open by a set of sentences. In particular, they're looking at the following list:

$$p \rightarrow q; q \rightarrow \neg p$$

(a) (Student:) [Pointing at the list] Given this list, p might be the case.

(b) (Teacher:) [Gives some logic explanation showing how the list rules out p's being true, then says] So it's not true that, given the list, p might be the case.

(c) (Student:)

(i) I guess I was wrong.

(ii) #But what I said was still true./#But still, *given this list*, p might be the case.

Here, the information is third person—sentences from propositional logic. The challenger, the teacher, targets the entire epistemic sentence. That means that the student's initial assertion was false at the time of utterance, which results in infelicity if the student digs in.

Logic, being third person, is quite natural and almost not worth mention. It's my estimation that the corresponding first-person cases are quite rare, especially if they are solipsistic. For what would such a case look like? It would be one in which you claim that your knowledge leaves open some claim and then I object that your knowledge doesn't. But it's often the case that I'm not privy to all of what you know. This would suggest that robust, first-person–solipsistic disagreement is rare. We can, however, construct cases that aren't outlandish. Here's one.

(4.3) Row 14—*Solipsistic Tutoring*: 1⚔SMF.

Context: I'm tutoring my little cousin, Leila, in math. I'm trying to show her that whenever you multiply two even numbers, you get an even product. We're looking at the numbers 2 and 4. I say to her, "you have two numbers, 2 and 4, and you know that both are even (which we established earlier that she knows). Given what you've learned,[14] what do you know about the product?" She then says,

(a) (Leila:) Well, given what I learned, It might be odd.

(b) (I reply:) No, Leila, that's not right. It's not true that, given what you learned, it [the product] might be odd.

(c) (Leila:)

(i) I guess I was wrong.

(ii) #But what I said was still true./#But still, *given what I learned*, it might be odd.

The explanation here is essentially the same as the last case: I target Leila's entire assertion, and since I'm not mistaken about the facts, then she was wrong that her

---

[14]I'm using "Given what you learned. . . " because I actually used that. It would be more precise to say "Given what you know. . . ."

knowledge left open a possibility. So her assertion was false when she said it, so she can't felicitously reiterate it.

It is interesting to note that a more straightforward example (a *1𝒥SMF* case) without a lot of context, strikes us as odd:

(4.4) Context: Jones and Smith wonder where Bob is, but they haven't been searching together for him and don't know much about what each other knows.

    (a)   (Jones:) Given what I know, Bob might be in his office.

    (b)   (Smith:) That's not true.

    (c)   (Jones:) What's not true?

    (d)   (Smith:) # What you said—it's not true that, given what you know, Bob might be in his office.

We're rarely so privy to another's knowledge, making such a challenge so preposterously presumptuous as to result in infelicity. Again, this suggests that what we might think of actual instances of archetypical modal disagreement are somewhat rare.

We can turn the group–solipsistic knob to group, all else being the same, including the explanation of the infelicity, although again the case is somewhat atypical:

(4.5) Row 10—*Group Tutoring:* 1𝒥GMF.

Context: I'm tutoring my little cousins, Leila and Posie, in math. I'm trying to show them that whenever you multiply two even numbers, you get an even product. We're looking at the numbers 2 and 4. I say to them, "you have two numbers, 2 and 4, and you know that both are even (which we established earlier that they know). Given what you've learned, what do you two know about the product?" Speaking on behalf of the group, Leila then says,

126

(a) (Leila:) Well, given what we learned, it might be odd.

(b) (I reply:) No, that's not right. It's not true that, given what you two learned, it [the product] might be odd.

(c) (Leila:)

    (i) I guess we were wrong.

    (ii) #But what I said was still true./#But still, *given what we learned*, it might be odd.

The $1\mathscr{I}GMF$ case type raises some interesting issues that the corresponding solipsistic case doesn't, supposing we avoid recherché epistemological issues involving agents unaware of their own knowledge. In a group modal, a speaker can mistake what the information is by somehow mischaracterizing it, either including propositions that don't belong or by having the information incomplete. We can generate a plausible case with a group in some official capacity to pin down what counts as the group's knowledge with respect to the prejacent:

(4.6) *Investigation*

Context: A police committee is investigating a death. The committee's official information—i.e. their knowledge that they've all decided should count as the information in the case—is determined by the propositions e_1, e_2, and e_3, which, let's suppose, rule out accidental death. One member of the committee—let's call him the representative—for some reason mischaracterizes the information, either by including some extra information or by not having the information complete. This mischaracterization allows for the possibility of accidental death. The representative for the committee gives a press conference, but a reporter has already talked with a more careful member of the committee:

(a) (Representative:) Given what the committee knows about the death, it might have been an accident.

(b) (Reporter:) No, that's not right. The official committee information has ruled out accidental death.

(c) (Representative:)

   (i) Well, I guess I was wrong.

   (ii) #But what I said was still true./#But still, *given what the committee knows about the death*, it might have been an accident.

In *Group Tutoring* Leila had the group's information correct, but she bungled the inference. Here, the investigator bungles the inference because he's mischaracterized the group's official information. He has it wrong that the information leaves open the prejacent in virtue of having the information wrong, a possibility that is more salient in group rather than solipsistic cases.

What ties all of the M cases together is that the challenger targets the entire modal, and that amounts to targeting an *inference*. In *Logic-Modal*, *Solipsistic Tutoring*, *Group Tutoring*, and *Investigation*, the speaker of the initial epistemic sentence somehow has the inference wrong, either simply or by mischaracterizing the information. And having the inference wrong is another way of saying that the epistemic sentence is false at the time of utterance, and this explains our data point—the felicity judgement. We've now taken care of rows 10, 14, and 22.

Before moving onto the more interesting T cases, three rows remain, the first two of which are first person and the last of which is third: 12, 16, and 24, which we'll call *Group Prejacent Tutoring*, *Solipsistic Prejacent Tutoring*, and *Logic-Prejacent*, respectively. The explanation for the infelicity of digging in in these cases will be the same as the above—you can't reiterate a false sentence—but they can be messier because the challenger in these versions doesn't show the initial speaker that her

entire epistemic sentence is false, though in these cases we've stipulated that it is. In each case type, the challenger targets the prejacent of the initial speaker's false epistemic sentence. The initial speaker can then, by finding out the prejacent is false, come to know that the epistemic sentence is false, ruling out its reiteration. This is the cleaner option. On the messier option, the initial speaker may, despite coming to know the prejacent's falsehood, still think that the information at the time of the initial utterance of her epistemic sentence was compatible with the prejacent. The participants in this kind of case, specifically the initial speaker, however, are too confused for the case to be interesting. The related case in which the initial speaker is *correct* about the initial epistemic sentence's truth are quite interesting, and are a subset of the T cases, to which, having covered the F cases, we now turn.

4.3.1.2 T cases

The cases that most interest us, those that puzzle the most, are those in which the initial epistemic sentence was true at the time of utterance. This interest arises from the fact that contextualist truth is easy to come by—we only need the weak logical relation of compatibility to hold between the prejacent and the information— and reiterating a true sentence shouldn't be a problem. Yet it often is. I focus on these cases, the T cases, in this section, and then expand on big-picture themes that the T cases involve subsequently. Instead of starting with the third-person cases and moving to first, as in the last subsection, I'll move from the $\mathscr{F}$ cases to the $\mathscr{I}$ cases. Again, the organizing principle will be progressive argumentative difficulty.[15] Explaining why and when, in the subset of the T cases, digging in is felicitous or infelicitous is the most important and challenging part of the paper. So in this section, I lay out the cases and quickly explain the source of the felicity

---

[15]The $\mathscr{F}$ cases are more surprising since it seems that the $\mathscr{I}$ cases are the norm, so the opposite order may seem more natural. However, I think the norm cases are more difficult to explain.

judgment, then in the next section, I give a more detailed, big-picture argument to explain the felicity judgements.

To reiterate the dialectic in the literature to this point: It seemed at first that digging in was always infelicitous, but then we realized that sometimes it isn't odd, given the right context, to dig in—though infelicity, it seems, is the norm. The few cases above were all infelicitous cases, and we presently introduce felicitous cases. We can take two views on the complicated data. First, we may say that we don't need to explain the data because they're unclear. Or, as I prefer, as it is the point of this paper, we may say that the data are more difficult to explain since we need to explain why the felicity can go one way or the other. This is how things will go for the rest of the paper.

We'll start with the surprising felicitous cases first, starting with the one third person case, row 19, which I think is the clearest example of felicitous digging in:

(4.7) Row 19—*Bible*: 3$\mathscr{F}$PT.

Context: An atheist Bible scholar is explaining his findings to a lay theist he just met. The findings involve a newly discovered manuscript, which we'll assume leaves open the possibility that Judas was a good man.

(a) (Atheist Bible scholar:) Given what this newly discovered manuscript says, Judas might have been a good man.

(b) (Theist:) Judas was a treacherous man!

(c) (Atheist Bible scholar:) Look *I* have no opinion on Judas. I'm merely saying: *Given what this newly discovered manuscript says*, Judas might have been a good man.

Here, the digging in sounds fine—surely not infelicitous. This is because, as is sometimes the case in scholarly pursuits, the Atheist Bible scholar doesn't really care much about how Judas in fact was. She is only interested in the compatibility of

the information from the scroll and Judas being a good man, and about the veracity of the information or the truth of the prejacent, she may be hostile or indifferent. The theist, on the other hand, is concerned with the actual truth of the prejacent, which explains her response.

The corresponding first person cases are next, which I name after von Fintel and Gillies (2008, p.81, my numbering), who first showed (as far as I know) that digging in is fine. As we'll see, we can take this case to be either group or solipsistic. We'll call the group version *Group vF&G* and the solipsistic version *Solipsistic vF&G*:

(4.8) Rows 3 & 7—*vF&G*. 1$\mathscr{F}$-G/S-PT

Context: Alex and Billy are looking for keys.

(a) (Alex:) The keys might be in the drawer.

(b) (Billy, Looks in the drawer, agitated:) They're not. Why did you say that?

(c) (Alex:) Look, I didn't say they were in the drawer. I said they might be there—and they might have been. Sheesh.

The point of this case, as von Fintel and Gillies originally presented it, was to show that the data were mixed—that digging in is sometimes felicitous. I agree that this is a data point, but I'm going to explain the case in terms of our parameters. The case sets the $\mathscr{F}/\mathscr{I}$ parameter to $\mathscr{F}$, which, once in place, automatically sets the T-F dial to T—because one can't reiterate a sentence that the challenger has shown false. Setting that dial automatically sets the M/P dial to P since, if the dial were set to M, then the initial epistemic sentence would be false. Since there's no obvious third-person information source in the context, it's clear that this is a first-person case. Last, the G-S parameters is a little more difficult to set because both are compatible. The information involved in Alex's initial epistemic sentence, given the facts of the case, can either be Alex's knowledge alone or both Alex and Billy's. It's clear that Alex's knowledge is compatible with the prejacent when she

makes her initial assertion, and Billy looked for the keys *after* Alex's initial assertion, so Billy didn't know where they were at the time of the utterance. So both G and S are compatible with the case, but whichever option it is will be immaterial since what I say below is compatible with either option. So we'll count this case for both rows 3 and 7. I do, however, think it's much more plausible that this is an S case. What one wants to say about Alex's digging in, which I discuss more below, is that she seems like she's defending her intelligence, and I think the idea of defending one's intelligence is more plausible on a solipsistic rather than a group reading.

In the cases up until now, the setting of the non-$\mathscr{F}/\mathscr{I}$ parameters explained the the $\mathscr{F}/\mathscr{I}$ parameter for that case. For example, being shown that one's sentence is false is enough to explain why one can't reassert that sentence. We need to say more in this case, though, to explain what's going on with Alex's digging in, and here the first-person–third-person distinction transforms from an interesting distinction to an explanatory one.

What exactly is Alex saying when she says "[the keys] might have been [in the drawer]"? The epistemic modal operator tends to out-scope the past-tense operator, so we tend to read sentences that involve the two operators as referencing the *current* information about a *past* possibility. While "might have been" is often read as non-epistemic, an epistemic interpretation of "might have been" appear in sentences such as the following:

(4.9) Jones isn't a philosopher, but he might have been.

This sentence still favors, it seems to me, the non-epistemic interpretation under which Jones was never a philosopher, but it is counterfactually possible that Jones was. The speaker in this kind of context probably thinks that although Jones ended up (say) becoming a lawyer, his nature was compatible with the life of a

philosopher's.[16] The epistemic interpretation of (9) is that I know Jones is not a philosopher, but the *current* information leaves open whether—at some time before now—he was. The corresponding interpretation isn't what we're looking for in *vF&G*. This would mean that Alex, in (8c), is asserting that the keys aren't in the drawer, but the current information leaves open whether, at some time before the assertion time, the keys were in the drawer. However, the context makes clear that the keys weren't in the drawer at some relevant past time—maybe they were days ago, but an assertion with that meaning would be irrelevant.

The intended reading, then, is that the keys' being in the drawer was compatible with the speaker's information at the past utterance time. If this is indeed the reading, then I suggest that Alex's digging in is a past-looking, third-person epistemic sentence. Here, the information source is third-person; it's Alex's knowledge frozen at a past time, and Alex claims that her frozen knowledge state at that time left open the key's being in the drawer.

Here's another way to put the point. As noted, what we want to say about Alex's response, because of its defensive air, is that she is defending her intelligence. But we shouldn't stop there, for what does defending one's intelligence consist in in this case? It consists in, in this case, looking *back* on one's knowledge and asserting that it was compatible with some proposition at that time, which, crucially, involves a *different* take on the information. The information source is now static and non-updatable.[17] I'll leave this case here, but we'll pick it up in the next section, comparing it with the corresponding infelicitous case—for these two cases are what I consider the most important for the paper.

---

[16]Indeed, I naturally read *vF&G* as counterfactual—that although the keys were never in the drawer at the relevant time, events could have unfolded such that they were. A natural continuation on the counterfactual reading would be "[the keys] might have been in the drawer, but since someone didn't put them where they were supposed to go they weren't!"

[17]For the group version, Alex would be looking back at the group knowledge, frozen in time.

We now have three $\mathscr{I}$ rows left: 11, 15, and 23. We'll go in reverse order. Case 23 is a case in which the initial epistemic sentence is third person and true, the challenger targets the prejacent, and then digging in is infelicitous:

(4.10) Row 23—*Exploring*: 3$\mathscr{I}$*PT*

Context: Magellan and Drake are trying to find Atlantis with a scroll with some vague references to Atlantis' location. Magellan is holding the scroll, and both Magellan and Drake believe the scroll to be accurate. The scroll suggests that Atlantis is in the location that Magellan is pointing.

(a) (Magellan, pointing, but not looking:) Given what the scroll says, Atlantis might be right there.

(b) (Drake:) Well obviously it's not. We're looking at landless ocean!

(c) (Magellan, pointing still:) ? But still, *given what the scroll says*, Atlantis might be right there!

The adverbial phrase makes clear the third person reading, and we've stipulated that the scroll leaves open the prejacent. Drake targets not the inferential relation but rather simply the proposition that Atlantis is in front of them (or something to that effect). So what's wrong with Magellan's digging in? Here, I want to say that, given their joint activity—actively searching for an island (and not studying, say, maps and scrolls)—Magellan's response is bizarre, perhaps infelicitous. The irrelevance of the digging in to their joint task explains the bizarreness. Again, I say more about this in the next section.

Next, row 15, the kind of dialogue that started it all—the kind, as mentioned, involved in the relativist's anti-contextualist arguments. But we must be careful with how we construct these cases, which are supposed to go something like this: Jones says "Bob might be in his office," Smith says "No, he told me he's going to be out of town this week," and Jones says "I guess I was wrong" instead of

the infelicitous "Sure, but what I said was true." This case is first person (since there's no salient third-person information source), and the felicity intuition is robust enough. But the G-S parameter is tricky: Unlike *vF&G*—where Billy's key searching shows she's ignorant with respect to the prejacent of the initial epistemic sentence—in this case, Smith knows the truth of the prejacent (i.e. he knows that Bob isn't in his office) at the time of Jones' initial utterance. This means that if this is a group case, then we have an easy explanation of the felicity intuition: The initial epistemic sentence was false since the group knowledge did *not* leave open the truth of the prejacent. In other words, if we're not careful, we'll end up with a case like *Investigation* again. Thus, we need a case in which the challenger is ignorant with respect to the prejacent. In order to have a solipsistic case, we need a context in which the participants aren't involved in any kind of joint search, where it's much more plausible that the initial speaker isn't referencing joint knowledge, one in which it would be very strange to prepend "Given what we know, . . ." to the epistemic sentence. Here is such a case:

(4.11) *Row 15–Solipsistic Usual Relativist*: 1⟦SPT

Context: Jones is looking for Bob, and Smith is nearby. Jones is the searcher here—it's clear that Jones and Smith aren't jointly searching for Bob, and Smith, though not so much an eavesdropper, is just hanging around.

(a) (Smith:) Hey Jones, How you doing?

(b) (Jones:) Good, I was just looking for Bob. Oh yeah, I think he might be in his office.

(c) (Smith, glances at his office door, on which there's a note:) Oh, no, he's not—there's a note here that says he's left for the day.

(d) (Jones:)

    (i) Huh, I guess I was wrong.

(ii) #Sure, but what I said was true.

For the group version—row 15, our last row, we change the context:

(4.12) *Row 11–Group Usual Relativist*: 1$\mathscr{I}$GPT

Context: Jones and Smith are looking for Bob to ask him an important question.

(a) (Jones:) Oh, Bob might be in his office.

(b) (Smith, glances at his office door, on which there's a note:) No, he's not—there's a note here that says he's left for the day.

(c) (Jones:)

(i) I guess I was wrong.

(ii) #Sure, but what I said was true.

In both of these cases, digging in, unlike in *Group* and *Solipsistic vF&G*, is infelicitous. Distinguishing these two sets of cases is crucial. What I want to maintain is that the digging in would be irrelevant enough to constitute infelicity in these cases, or at least deserve two question marks, whereas in the *vF&G* cases it wouldn't. I explain this crucial difference in the next section.

4.4   Zooming Out

Recall the narrow goal: to explain the felicity data in these dialogues. And, recall, the strategy: to classify each dialogue as one in which

- reiterating a challenged epistemic sentence is infelicitous because the sentence was false;

- reiterating a challenged epistemic sentence is infelicitous because, although the sentence was true, it's too irrelevant to reiterate felicitously; or

136

- reiterating a challenged epistemic sentence is felicitous because the epistemic sentence is true and, at the reiteration time, is relevant.

The following table classifies each dialogue:

False	True + Odd	True + Not Odd
Group Tutoring, Group Prejacent Tutoring, Solipsistic Tutoring, Solipsistic Prejacent Tutoring, Logic-Modal, Logic-Prejacent	*Group Usual Relativist, Solipsistic Usual Relativist, Exploring*	*Group vF&G, Solipsistic vF&G, Bible*

Explaining the difference between the odd + true cases versus the not odd + true cases is our current challenge. In other words, we have a handful of cases that are alike in that the initial epistemic sentence is true and the challenger targets the prejacent. Yet, in some of these cases—the *vF&G* cases and *Bible*—it's not odd to dig in whereas in others—the *Usual Relativist* cases and *Explorer*—it is. What gives?

To begin, let's visualize the difference in the way that the actual world relates to first- and third-person epistemic sentences. Consult figure 1. The *I* circle represents the worlds that are compatible with the information, and the *P* circle represents the worlds at which the prejacent is true. The "@" symbol represents the actual world; the rectangle represents logical space, which "Ω" denotes.

The semantics for simple-minded contextualism—i.e. contextualism minus the stuff about the ordering source—says that a sentence of the form $\ulcorner$It might be that $\phi\urcorner$ is true so long as the information and the prejacent are compatible, i.e. so long as some of the information and prejacent worlds overlap. Erasing the actual world symbols, Figure 4.1 pictures this semantics. The information source that determines the information involved in the semantics will determine the possible locations for the actual world. Figure 4.1(a) pictures what logical space *looks* like to someone when she asserts a first-person epistemic sentence. Since the information, either

(a) First Person   (b) Third Person
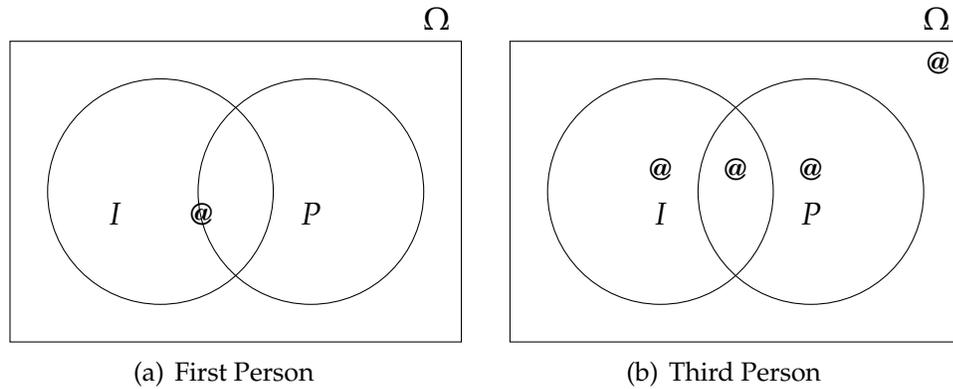
Figure 4.1. Epistemic modals

group or solipsistic, is knowledge, and since knowledge is factive, the actual world is a member of it. One asserts an epistemic sentence when one is unsure whether to locate the actual world in $I \setminus P$ or in $I \cap P$, though, *in fact*, it will be on one side or the other if the proposition that the epistemic sentence expresses is true. On the other hand, depending on the type of third-person epistemic possibility modal, there are no restrictions on the location of the actual world, as Figure 4.1(b) pictures. If the static content is past knowledge (or some other kind of factive content), then the actual world will be somewhere inside of I. But if the content is some other kind of non-factive content, then the actual world could be anywhere. Saying that the assertor of a third-person epistemic sentence is indifferent to the information's veracity, as the atheist Bible scholar is in *Bible*, is another way of saying that speaker is indifferent about the location of the actual world in Ω. Of course, *in fact*, the actual world will be somewhere or other on the diagram.

The first-person naturally pairs with an interest in locating the actual world in logical space. Cases like *Usual Relativist*, either group or solipsistic, are paradigmatic here. Let's just focus on the solipsistic version for concreteness. When I say "Jones

might be in his office" amidst searching for Jones, I'm trying to locate the actual world in logical space. When I say that sentence, what I know leaves open my world being a Jones-in-office world. But given my activity—searching for Jones—I'd like to further narrow down Jones' location—i.e., to further narrow down the location of the actual world, finding out whether or not Jones is in his office. Applying Figure 4.1(a) to this example, where P is the proposition that Jones is in his office (at some time), I'm not sure which side of the line on which to place the actual world, but I'm actively trying to find out. The third-person, on the other hand, can pair with merely filling out logical space. In other words, in the third person, one may only be interested in whether the prejacent and the information are compatible regardless of where the actual world may be in logical space. To give an example, consider the police shooting of Walter Scott, in which the police initially claimed that the fatal use of force might have been justified, but later video contradicted that. Suppose I utter the following epistemic sentence: Given the initial local news coverage on the Walter Scott shooting, the fatal use of force might have been justified. Now further suppose that I say that *after* I've seen the video that shows that the use of force was clearly unjustified. In this context, I'm only interested in the compatibility of the worlds compatible with the initial local news coverage and the worlds at which the use of force in this incident is justified—not in the location of the actual world. Applying Figure 4.1(b) to this example, I'm not actively trying to place the actual world in either $I \cap P$ or in $I \setminus P$—I know that the actual world isn't even in I. I'm just interested in whether the I and the P circles overlap.

Just as epistemic sentences pair with a certain attitude toward the location of the actual world, so can conversations: some revolve around locating the actual world in logical space; others, around filling it out. But when we mismatch the conversational attitude with the wrong kind of epistemic sentence, a certain kind of tension results because the epistemic sentence isn't relevant to the conversation.

And now we're in a position to finally say what distinguishes the true-yet-odd epistemic sentences from the true-and-not-odd ones: *If the tension is great enough, we get infelicity*. Tension here means a difference between the point of an assertion and the point of a conversation. Note that, when I say "tension," I'm not introducing new or interesting philosophical terms here that are to do major work. I'm merely saying, in other words, what irrelevance is: Conversations tend to revolve around a given topic and if someone says something off topic enough, infelicity ensues. The point I'm making, interesting if I succeed, is to say first that irrelevance explains the typical confounding infelicity of digging in, and second, to point out the source of the irrelevance. To put the second point differently: To say why a given instance of digging in is irrelevant enough to result in infelicity, I need to say why such an assertion is off topic, and here's why: The topic is locating the actual world; digging in isn't on that topic.

Let's see how this works out with our cases. First, let's explain the different felicity data in the *Usual Relativist* cases verses the *vF&G* cases, which I see as the greatest challenge for this paper. In the *vF&G* cases, the context makes clear that the participants are actual-world hunting. In other words, they're trying to figure out which side of the line to place the actual world on Figure 4.1(a), which is how logical space looks at the time of the initial utterance. After Billy's challenge, the actual world moves to the left, in $I \setminus P$. But the challenger introduces a special feature of this context: her assertion is somewhat heated and accusatory. This prompts Alex's response, which in terms of our diagram says that, although she *currently* knows that logical space is such that $@ \in I \setminus P$, it was reasonable for her to place the actual world on the line when she uttered her initial epistemic sentence. Alex's assertion in her digging in is in the third person: She's defending the reasonableness of I and P overlap. Being third person, Alex's digging in also has the effect of abandoning the hunt for the actual world. I said that we get infelicity if the tension is great

enough. Here, the point of the conversation is initially to find keys, and Alex's digging in doesn't contribute to that. But there is less tension because of the special context—there is a reason for Alex's response, viz. Billy's need to relax a bit.

What I'm going to say about the *Usual Relativist* cases applies to both, so I won't differentiate them. I've given some context to *Usual Relativist*, but I think that it's no accident that similar cases are glossed with little to no context. Like *vF&G*, *Usual Relativist* starts out in a context in which the participants attempt to locate the actual world. But unlike, *vF&G*, the challenger doesn't take an accusatory tone. Because of this, there is no reason for Jones to defend his intelligence—doing so would be bizarre. In other words, asserting the third-person epistemic sentence implicated in digging in would introduce an intolerable tension between the point of the conversation (locating the actual world—i.e. finding Bob) and the point of the epistemic sentence (not locating the actual world). In yet other words, digging in would be so irrelevant as to be infelicitous.

Before turning to closing remarks, we have two third-person cases to explain: *Bible-Prejacent* and *Explorer*. *Bible* is quick to explain: The Atheist Bible Scholar only cares about the overlap between *I* and *P*, not in locating the actual world, which the Theist is interested in, as her reply shows. The Atheist Bible Scholar's digging in is a way of stressing that she's only interested in *I–P* overlap, not in locating the actual world. If there is any tension in this conversation, it's that the Theist is somewhat confused about the point of the Atheist Bible Scholar's assertion, but the confusion is mild, and the Atheist Bible Scholar simply let's the Theist know what she is up to. Another way to put this is that there's no clear point, and some confusion, over what the conversation revolves around, and the digging in resolves the confusion.

Explorer is interesting in that it somewhat mimics *Usual Relativist*. There are two reasons for this: The context is one in which the conversational participants are attempting to locate the actual world, and the conversational participants *take* the

information to be factive. Drake and Magellan presume the map to be veracious, and the feel of the context is that it's somewhat high stakes—they're on the high seas searching for a lost island. If Magellan were to dig in, he would abandon his search for the actual world to consider how logical space looked at some past time. The result would be a tension between a conversational context that revolves around locating the actual world and a sentence that would abandon that purpose—a sentence whose point would be to consider the content of a map that they currently know is faulty. Such tension would be too great, and so the digging in would be irrelevant to the point of infelicity.

In summary, here is the diagnoses of both the true-and-odd cases and the true-yet-not-odd cases:

True+Odd

- Both *Usual Relativist* cases and *Exploring*. Context involves locating actual world. Digging in involves consideration of logical space at some past time with no special prompt to do so.

True+Not Odd

- *Bible*. Context is somewhat unclear. Digging in resolves the unclarity.

- Both *vF&G* cases. Context involves locating actual world. Digging in involves consideration of logical space at some past time but there is a special prompt to do so.

To finally close this section, let me recast the person distinction in such a way that will highlight its importance. With the narrow focus on knowledge as the base for epistemic modality, one might have thought that epistemic modality groups with metaphysical, nomological (or circumstantial), and analytic modality in being *alethic*. A modality is alethic if, but only if, necessarily, the actual world is a member of the base. If must-*p* or might-*p* are true, and the modality is alethic, then, necessarily, the

actual world will be a *p* world. And the following inferences are characteristically valid for alethic modalities:

$$\Box p \vDash @p; @p \vDash \Diamond p.$$

Unlike the just-mentioned modality flavors, deontic modalities are, paradigmatically, non-alethic.

So, Is epistemic modality alethic? If we were to confine our attention to an epistemic modality whose base knowledge determines—i.e. first-person epistemic modality—the answer is yes. It's hard to overstate how many critical papers in at least the philosophy literature suggest that knowledge is the exclusive base for epistemic modality, further suggesting that epistemic modality plays double duty. Egan and Weatherson (2009), in their introduction to the only philosophical anthology on epistemic modality, define contextualism in terms of knowledge and expound on considerations that only make sense in terms of knowledge—e.g., the so called "speaker inclusion constraint," which requires that the speaker's knowledge is a part (proper or improper) of the knowledge that determines the base. Such a consideration only makes sense, obviously, if something epistemic determines the base. And contextualists aren't alone in their narrow focus. Both content and truth relativists speak in terms of an exclusive knowledge base.

But with the person distinction in hand, the answer is sometimes yes, sometimes no, depending on the information source. If the information source is factive, then the modal will be alethic; otherwise, not. Epistemic modality, as far as I know, stands alone in housing both alethic and non-alethic versions of itself. And the alethic–non-alethic distinction gives us another angle on some of what we've explained. When a modality is non-alethic, it comes as no surprise that one can "dig in" in the face of a prejacent challenge. For example, if I point to a crosswalk on a new road and say that pedestrians can cross there, and someone informs me that it's not true

that pedestrians cross there (since none have), then I can felicitously reiterate my deontic sentence. The same goes for an epistemic sentence that's initially asserted in the third person.

4.5   Closing remarks: The nature of the person distinction

If my arguments up to this point have succeeded, then I've achieved my narrow goal of explaining the felicity data within one kind of modal disagreement. To close, I want to quickly consider the nature of the first–third-person distinction. As I said in the intro, I think we best understand the distinction as non-semantic, and I've appealed to non-semantic notions, viz. relevance, in my main argument. I've also tried to speak in terms of different information *sources* in distinguishing first- and third-person epistemic sentences rather than different information *simpliciter*. To close the paper, let me briefly say why I think we should understand the difference between first- and third-person epistemic sentences as a difference in the way information is determined rather than in terms of different *kinds* of content.

The kind of case that convinces me is one in which agents commit differently to the same content, assuming the simple-minded version of contextualism that I've been assuming in this paper. The difference in commitment arises from different information sources, either first or third person. In other words, we can construct a case in which an agent's knowledge and some external information source determine the same information that is involved in an epistemic sentence, yet the commitment facts are different. Here is such a case.

Suppose I come into existence, *tabula rasa*, and read the entire *Encyclopedia Britannica*, which exhausts the entirety of my knowledge and belief. Suppose that *Britannica* strongly suggests that Shakespeare wrote all of the plays that are attributed to him but also suggests that there's a modicum of evidence that Francis Bacon wrote a few of the plays. Now suppose I assert the following:

(4.13) (I:) Bacon might have written *Romeo and Juliet*.

My information leaves open whether Bacon wrote Romeo and Juliet. Now suppose
that you have also read *Britannica,* but it doesn't exhaust your knowledge and belief.
You are in fact a Shakespeare scholar and have considered all the information, and
you have ruled out that anyone but Shakespeare wrote *Romeo and Juliet*. You say:

(4.14) (You:) Given what *Britannica* says, Bacon might have written Romeo and
Juliet.

Here the same intensional content, *Britannica*, determines the information that (it
is asserted) leaves open the possibility that Bacon wrote *Romeo and Juliet*. For me,
however, that content exhausts my epistemic state. For you, it is the content of
a set of encyclopedias. According to a very simple version of contextualism, (13)
and (14) determine the same propositions, yet these different assertions come with
different commitments. In making your assertion, you play the role of arbiter.
You're considering the inferential relation between some static information and a
proposition. This being the case, you can follow up your claim with "but of course
Shakespeare wrote Romeo and Juliet." On the other hand, I, issuing a first-person
epistemic sentence, cannot give the same follow-up to my assertion.

Finally, let's put this point in another, general way. We've repeatedly made
the point that we can reference particular information sources that determine the
information involved in the proposition an epistemic sentence expresses. And
though I haven't spoken in terms of the formal semantics of epistemic modality
much up to now, in closing, I bring it in because I think it makes this point especially
clear. According to standard possible worlds semantics (leaving out the ordering
source), we associate the information with a set of possible worlds compatible with
that information, and the possibility modal particles *might* or *may* semantically
contribute an existential quantifier over that set of worlds. We state the truth

conditions in terms of existential quantification over the information worlds: An epistemic sentence is true just in case there is at least one world in the information worlds at which the prejacent is true. If we identify these truth conditions with the proposition an epistemic sentence expresses, then, at the level of the proposition, different information sources can determine the same set of worlds, as in the above case. And we can't, by inspecting this or that proposition that an epistemic sentence expresses, determine how a speaker or hearer may relate to that proposition vis-à-vis how they relate to the actual world. There is, so to speak, no road backwards from the proposition an epistemic sentence expresses to the information source involved in that proposition. I'm not sure what to make of the equivalent point, *mutatis mutandis*, for other flavors of modality—that, e.g., different rule sources can determine the same quantificational domain for deontic modals. But for epistemic modality, it's significant.

Table 4.1. Cases (gray rows are possible cases; white, impossible)

	1 or 3	$\mathscr{I}$ or $\mathscr{F}$	G or S	M or P	T or F	Case name	Bad combo
1	1	$\mathscr{F}$	G	M	T		f/M, M/T
2	1	$\mathscr{F}$	G	M	F		f/F, f/F,
3	1	$\mathscr{F}$	G	P	T	*Group vF&G*	
4	1	$\mathscr{F}$	G	P	F		f/F
5	1	$\mathscr{F}$	S	M	T		f/M, M/T
6	1	$\mathscr{F}$	S	M	F		f/M, f/F
7	1	$\mathscr{F}$	S	P	T	*Solipsistic vF&G*	
8	1	$\mathscr{F}$	S	P	F		f/F
9	1	$\mathscr{I}$	G	M	T		M/T
10	1	$\mathscr{I}$	G	M	F	*Group Tutoring*	
11	1	$\mathscr{I}$	G	P	T	*Group Usual Relativist*	
12	1	$\mathscr{I}$	G	P	F	*Group Prejacent Tutoring*	
13	1	$\mathscr{I}$	S	M	T		M/T
14	1	$\mathscr{I}$	S	M	F	*Solipsistic Tutoring*	
15	1	$\mathscr{I}$	S	P	T	*Solipsistic Usual Relativist*	
16	1	$\mathscr{I}$	S	P	F	*Solipsistic Prejacent Tutoring*	
17	3	$\mathscr{F}$	X	M	T		f/M, M/T
18	3	$\mathscr{F}$	X	M	F		f/F, f/M
19	3	$\mathscr{F}$	X	P	T	*Bible*	
20	3	$\mathscr{F}$	X	P	F		f/F
21	3	$\mathscr{I}$	X	M	T		M/T
22	3	$\mathscr{I}$	X	M	F	*Logic-Modal*	
23	3	$\mathscr{I}$	X	P	T	*Exploring*	
24	3	$\mathscr{I}$	X	P	F	*Logic-Prejacent*	

BIBLIOGRAPHY

Amaral, Patricia, Roberts, Craige, and Smith, E. Allyn. 2007. "Review of the Logic of Conventional Implicatures by Chris Potts." *Linguistics and Philosophy* 30:707–749.

Armour-Garb, Bradley and Woodbridge, James A. 2012. "The Story About Propositions." *Noûs* 46:635–674.

Armstrong, D. M. 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

Armstrong, David. 2004. *Truth and Truthmakers*. Cambridge: Cambridge University Press.

Bealer, George. 1998. "Propositions." *Mind* 107:1–32.

Benacerraf, Paul. 1965. "What Numbers Could not Be." *The Philosophical Review* 1:47–73.

Braddon-Mitchell, David and Jackson, Frank. 2007. *The Philosophy of Mind and Cognition*. Blackwell.

Braun, David. 1998. "Understanding Belief Reports." *Philosophical Review* 107:555–595.

—. 2006. "Comment on David Chalmers's "Probability and Propositions"." Manuscript.

Bricker, Phillip. 2006. "David Lewis: On the Plurality of Worlds." In John Shand (ed.), *Central Works of Philosophy, Vol. 5: The Twentieth Century: Quine and After*. Acumen Publishing.

—. 2015. "Truthmaking: With and Without Counterpart Theory." In Barry Loewer and Jonathan Schaffer (eds.), *The Blackwell Companion to David Lewis*, chapter 11, 159–187. Blackwell.

Camp, Elisabeth. 2007. "Thinking with Maps." *Philosophical Perspectives* 21:145–182.

Caplan, Ben and Tillman, Chris. 2013. "Benacerraf's Revenge." *Philosophical Studies* 166:111–129.

Chalmers, David. 2006a. "The Foundations of Two-Dimensional Semantics." In Manuel Garcia-Carpintero and Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*, chapter 4. Oxford: Oxford University Press.

Chalmers, David J. 2006b. "Probability and Propositions." Manuscript.

—. 2006c. "Reply to David Braun's Comments on "Probability and Propositions"." Manuscript.

—. 2011a. "Frege's Puzzle and the Objects of Credence." *Mind* 120:587–636.

—. 2011b. "Propositions and Attitude Ascriptions: A Fregean Account." *Noûs* 45:595–639.

Choe, Yoonsuck. 2002. "Second Order Isomorphism: A Reinterpretation and Its Implications in Brain and Cognitive Sciences." *Proceedings of the Cognitive Science Society* 24.

Collins, John. 2007. "Syntax, More or Less." *Mind* 116:805–850.

Craik, Kenneth James Williams. 1967. *The Nature of Explanation*. Cambridge, Cambridge U.P.

Crimmins, Mark and Perry, John. 1989/2007. "The Prince and the Phone Booth: Reporting Puzzling Beliefs." In Matthew Davidson (ed.), *On Sense and Direct Reference*, 356–379. Boston: McGraw-Hill.

Cummins, Robert C. 1989. *Meaning and Mental Representation*. Cambrdige: The MIT Press.

—. 1996. *Representations, Targets, and Attitudes*. Cambrdige: The MIT Press.

Dixon, T. Scott and Gilmore, Cody. 2016. "Speaks's Reduction of Propositions to Properties: A Benacerraf Problem." *Thought: A Journal of Philosophy* 5:275–284.

Egan, Andy, Hawthorne, John, and Weatherson, Brian. 2005. "Epistemic Modals in Context." In Gerhard Preyer and Georg Peter (eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, 131–170. Oxford: Oxford University Press.

Egan, Andy and Weatherson, Brian. 2009. *Epistemic Modality*. Oxford University Press.

Field, Hartry. 1986. "Stalnaker on Intentionality: On Robert Stalnaker's Inquiry." *Pacific Philosophical Quarterly* 67:98–112.

Fletcher, Logan. 2014. "Why it isn't syntax that unifies the proposition." *Canadian Journal of Philosophy, forthcoming* .

Fodor, Jerry. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambrdige: The MIT Press.

Fodor, Jerry A. 2007. "The Revenge of the Given." In Brian P. McLaughlin and Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, 105–116. Blackwell.

Harris, Jesse A. and Potts, Christopher. 2009. "Perspective-Shifting with Appositives and Expressives." *Linguistics and Philosophy* 32:523–552.

Heck, Richard. 2007. "Are There Different Kinds of Content?" In Brian P. McLaughlin and Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*, 117–38. Blackwell.

Jacobson, Anne Jaap. 2013. *Keeping the World in Mind: Mental Representations and the Sciences of the Mind*. New Directions in Philosophy and Cognitive Science. New York: Palgrave Macmillan.

Jubien, Michael. 1991. "Could This Be Magic." *Philosophical Review* 100:249–267.

—. 2001. "Propositions and the Objects of Thought." *Philosophical Studies* 104:47–62.

Kaplan, David. 1989/2013. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In A.P. Martinich and David Sosa (eds.), *The Philosophy of Language*. Oxford: Oxford University Press, 6th edition.

Keller, Lorraine Juliano. 2017. "Against Naturalized Cognitive Propositions." *Erkenntnis* 82:929–946.

King, Jeffrey C. 2007. *The Nature and Structure of Content*. Oxford: Oxford University Press.

—. 2009. "Questions of Unity." *Proceedings of the Aristotelian Society* 109:257–277.

—. 2013. "On Fineness of Grain." *Philosophical Studies* 163:763–781.

—. 2014. "Naturalized Propositions." In Jeffrey C. King, Scott Soames, and Jeff Speaks (eds.), *New Thinking about Propositions*. Oxford University Press.

King, Jeffrey C., Soames, Scott, and Speaks, Jeff. 2014. *New Thinking About Propositions*. Oxford: Oxford University Press.

Kratzer, Angelika. 2012/1977. "What 'must' and 'can' must and can mean." In *Modals and Conditionals*, Oxford Studies in Theoretical Linguistics, chapter 1, 1–20. Oxford: Oxford University Press.

—. 2012/1981. "The Notional Category of Modality." In *Modals and Conditionals: New and Revised Perspectives*, Oxford Studies in Theoretical Linguistics, chapter 2, 21–69. Oxford: Oxford University Press.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Lasersohn, Peter. 2005. "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy* 28:643–686.

Lewis, David. 1970a/1983. "General Semantics." In *Philosophical Papers*, volume 1, chapter 12, 189–232. Oxford: Oxford University Press.

—. 1970b/1983. "How to Define Theoretical Terms." In *Philosophical Papers*, volume 1, chapter 6, 78–95. Oxford: Oxford University Press.

—. 1972/1999. "Psychophysical and Theoretical Identifications." In *Papers in Metaphysics and Epistemology*, chapter 16, 248–261. Cambridge: Cambridge University Press.

—. 1975/1983. "Languages and Language." In *Philosophical Papers*, volume 1, chapter 11, 163–188. Oxford: Oxford University Press.

—. 1979. "Attitudes De Dicto and De Se." *The Philosophical Review* 88:513–543.

—. 1980/1998. "Index, Context, and Content." In *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.

—. 1984. "Putnam's paradox." *Australasian Journal of Philosophy* 62:221–236.

—. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

—. 1988. "Vague Identity: Evans Misunderstood." *Analysis* 48:128–130.

Lloyd, Elisabeth. 1988. *The structure and confirmation of evolutionary theory*. New York: Greenwood Press.

MacFarlane, John. 2011. "Epistemic Modals Are Assessment-Sensitive." In Andy Egan and B. Weatherson (eds.), *Epistemic Modality*, chapter 5, 145–178. Oxford University Press.

McGinn, Colin. 1989. *Mental Content*. Blackwell.

McKay, Thomas and Nelson, Michael. 2010. "Propositional Attitude Reports." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2010 edition.

Millikan, Ruth G. 2002. *Varieties of Meaning: The 2002 Jean Nicod Lectures*. Cambrdige: The MIT Press.

Moore, Joseph G. 1999. "Numbers, and the Problem of Arbitrary Identification." *Synthese* 120:229–263.

Paul, L. A. 2012. "Metaphysics as Modeling: The Handmaiden's Tale." *Philosophical Studies* 160:1–29.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Putnam, Hilary. 1975/2007. "The Meaning of 'Meaning'." In Matthew Davidson (ed.), *On Sense and Direct Reference*, 126–179. Boston: McGraw-Hill.

Ramsey, William. 2007. *Representation Reconsidered*. Cambridge: Cambridge University Press.

Rescorla, Michael. 2009. "Cognitive Maps and the Language of Thought." *British Journal for the Philosophy of Science* 60:377–407.

Salmon, Nathan. 1986. *Frege's Puzzle*. Cambrdige: The MIT Press.

Schaffer, Jonathan. 2011. "Perspective in Taste Predicates and Epistemic Modals." In Brian Weatherson and Andy Egan (eds.), *Epistemic Modality*, chapter 6, 179–226. Oxford: Oxford University Press.

Schiffer, Stephen. 2003. *The Things We Mean*. Oxford: Oxford University Press.

—. 2007/1992. "Belief Ascription." In Matthew Davidson (ed.), *On Sense and Direct Reference*, 402–420. Boston: McGraw-Hill.

—. 2016. "Cognitive Propositions." *Philosophical Studies* 173:2551–2563.

Shepard, Roger N and Chipman, Susan. 1970. "Second-Order Isomorphism of Internal Representations: Shapes of States." *Cognitive Psychology* 1:1–17.

Shields, Christopher. 2016. "Aristotle's Psychology." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2016 edition.

Soames, Scott. 2010. *What is Meaning?* Soochow University Lectures in Philosophy. Princeton: Princeton University Press.

—. 2011. "Propositions as Cognitive Event Types." In Scott Soames Jeffrey C. King and Jeff Speaks (eds.), *New Thinking about Propositions*, chapter 6. Oxford: Oxford University Press, *forthcoming*.

—. 2016. "Propositions as Cognitive Acts." *Synthese* .

Stalnaker, Robert. 1970/1999. "Pragmatics." In *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford Cognitive Science Series, chapter 1, 31–46. Oxford: Oxford University Press.

—. 1987. *Inquiry*. Cambridge: The MIT Press.

Stalnaker, *Robert. 1999/1978. "Assertion." In *Context and Content*, chapter 4. Oxford: Oxford University Press.

Stephenson, Tamina. 2007. "Judge dependence, epistemic modals, and predicates of personal taste." *Linguistics and Philosophy* 30:487–525.

Swoyer, Chris. 1991. "Structural Representation and Surrogative Reasoning." *Synthese* 87:449–508.

Swoyer, Chris and Orilia, Francesco. 2014. "Properties." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

van Inwagen, Peter. 1986. "Two Concepts of Possible Worlds." *Midwest Studies in Philosophy* 11:185–213.

—. 2004. "A Theory of Properties." In Dean W. Zimmerman (ed.), *Oxford Studies in Metaphysics*, volume 1, 107–138. Oxford: Oxford University Press.

von Fintel, Kai and Gillies, Anthony S. 2008. "CIA Leaks." *Philosophical Review* 117:77–98.

—. 2011. "Might Made Right." In Brian Weatherson and Andy Egan (eds.), *Epistemic Modality*, chapter 3, 108–130. Oxford: Oxford University Press.

Weatherson, Brian. 2014. "David Lewis." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

Weisberg, Jonathan. 2011. "Varieties of Bayesianism." In Stephan Hartmann Dov Gabbay and John Woods (eds.), *Handbook of the History of Logic*, volume 10. Elsevier.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Williamson, Timothy. 2016. "Reply to Sullivan." *Canadian Journal of Philosophy* 46:759–765.