Fine-grained semantics for attitude reports

Harvey Lederman*

December 15, 2020

Abstract

I observe that the “concept-generator” theory of Percus and Sauerland [2003], Anand [2006], and Charlow and Sharvit [2014] does not predict an intuitive true interpretation of the sentence “Plato did not believe that Hesperus was Phosphorus”. In response, I present a simple theory of attitude reports which employs a fine-grained semantics for names, according to which names which intuitively name the same thing may have distinct compositional semantic values. This simple theory solves the problem with the concept-generator theory, but, as I go on to show, it has problems of its own. I present three examples which the concept-generator theory can accommodate, but the simple fine-grained theory cannot. These examples motivate the full theory of the paper, which combines the basic ideas behind the concept-generator theory with a fine-grained semantics for names. The examples are also of interest independently of my own theory: two of them constrain the original concept-generator theory more tightly than previously discussed examples had.

Keywords: attitude reports, Frege’s puzzle, names, impossible worlds

1 Introduction

Let Millianism be the thesis that names which intuitively name the same thing have the same compositional semantic value. Since “Hesperus” and “Phosphorus” both intuitively name the planet Venus, Millians say that these names have the same compositional semantic value. Accordingly, they also say that the two sentences

1. Plato believed Hesperus was visible in the evening; and

2. Plato believed Phosphorus was visible in the evening

*Thanks to audiences at Philosphical Linguistics and Linguistical Philosophy (PhLiP) 5, and the Princeton Talks in Epistemology and Metaphysics for their questions, to Chris Barker, Seth Cable, Mike Caie, Daniel Hoek, Wes Holliday, Sarah Moss, Paolo Santorio and Yael Sharvit for conversations and correspondence, and to Kevin Dorst, Ben Holguín, Matt Mandelkern, Daniel Rothschild, three anonymous referees, and Josh Dever in his role as editor for comments on the paper. I’m especially grateful to Kyle Blumberg and Cian Dorr, each of whom read two lengthy drafts and gave insightful detailed comments on both of them. I’ve learned a huge amount about these issues from my joint work with Jeremy Goodman, and the paper is heavily indebted to the many many many conversations we have had about them.
have the same compositional semantic value. This consequence of Millianism has been a key source of resistance to the theory. If Plato nightly pointed to Venus and said (the Greek translation of) “Hesperus is visible now, but Phosphorus never is”, many judge that 1 would be true, while 2 would be false.

But Millians too can respect this pattern of judgments, provided they hold that attitude reports are context-sensitive in the right way (Schiffer 1979, Crimmins and Perry 1989; see also Crimmins 1992, Dorr 2014, Goodman and Lederman forthcoming). Millians may hold that in the right circumstances, uttering 1 naturally suggests a context in which both 1 and 2 are true, while uttering 2 naturally suggests a different context, in which both 1 and 2 are false. The two sentences are true in exactly the same contexts – and “Hesperus” and “Phosphorus” have the same compositional semantic value – but typical uses of 1 in such circumstances are true in the contexts they suggest, while typical uses of 2 are false in the (different) contexts they suggest.

As it stands, of course, this idea is more of a wish-list than a theory. How should we think about the different contexts in which 1 and 2 are supposedly interpreted? The most prominent Millian theory of attitude reports in semantics today, first published by Percus and Sauerland 2003, and developed by Anand 2006 and Charlow and Sharvit 2014, can be seen as implementing a natural answer to this question. Very roughly, on this theory context supplies a set of salient descriptions of each object, and 1 and 2 are true in exactly the contexts where one of the contextually salient descriptions for Venus, δ, is such that \[ \text{Plato believed } \delta \text{ is visible in the evening} \] is true. The idea is then that using the word “Hesperus” often suggests a context where “the planet visible in the evening” is a salient description of Venus, while using the word “Phosphorus” often suggests a context where this description is not salient (but “the planet visible in the morning” is). This theory offers a simple and intuitive account of the contrast between 1 and 2. But unfortunately, as I will argue, it is not sufficiently flexible to handle closely related examples. Suppose again that Plato nightly pointed to Venus and said “Hesperus is visible now but Phosphorus never is” and consider:

3. Plato did not know that Hesperus is Phosphorus;

4. Plato was not sure that Hesperus is Phosphorus;

5. Plato did not believe that Hesperus is Phosphorus.

These sentences, as Frege 1892 observed, are naturally interpreted as true in this scenario. But, as I show in section 2, a straightforward application of the theory of Percus and Sauerland 2003 to these sentences predicts that none of them has an intuitive true reading.

---

1 Some take a passage in Laws 821c, where the character Kleinias describes the paths of “Hesperus and Phosphorus and other stars”, as evidence that the historical Plato did not know that the planet called “Hesperus” was the planet called “Phosphorus”.

2
One response to this argument – and one I consider near the end of the paper, in section 8 – would be to develop a different Millian theory – perhaps a variant of the theory of Percus and Sauerland [2003] – which avoids this prediction. Here, however, I first explore a more radical response. Millianism drove the need for a contextualist account of the contrast between 1 and 2 and it is one of the assumptions which leads to the problem with 3-5. It is therefore natural to wonder whether we might have been better off rejecting Millianism from the start. Motivated by this line of thought, I develop a semantics for attitude reports based on a fine-grained theory of the semantics of names, according to which names which intuitively name the same thing may nevertheless have different compositional semantic values. (Given my definitions, a theory of the semantics of names is fine-grained if and only if it is not Millian. 2 In Section 3 I present an abstract model for a fine-grained theory, and illustrate how it allows for reasonable true interpretations of 3-5. In section 4 building on ideas from Kaplan [1986] and Aloni [2005], I show how the theory can be extended to handle generalized quantifiers.

So far, it might seem, so good. But the basic fine-grained theory has some new problems of its own. In section 5 I present three examples which the basic theory cannot handle, but which the theory of Percus and Sauerland [2003] handles smoothly. These examples motivate developing a theory which combines the key ideas behind Percus and Sauerland’s theory with a fine-grained semantics for names. Section 6 presents such a theory, and section 7 shows how the theory accounts for the examples. In that section, I discuss how the examples impose independent constraints on the shape of my theory, as well as how they go beyond examples in the literature (in particular those given in Anand [2006]) designed to motivate particular features of Percus and Sauerland’s theory (most notably, the use of existential quantification over concept-generators).

With the new theory before us, we face an important question: should we prefer this fine-grained theory, or a more conservative, Millian variant on the theory of Percus and Sauerland [2003]? In section 8 I present a Millian theory which is sufficiently flexible to handle 3-5 as well as the examples from section 5. But I argue that the fine-grained semantics should be preferred over this alternative. Section 9 then concludes.

2Many theories of names can be natural developed in a way which is fine-grained according to my definitions. Typical “being called” predicativist theories, whether “that-” predicativist (Burge [1973]) or “the-” predicativist (Larson and Segal 1995, Elbourne 2005, Matushansky 2008, Fara 2015), as well as descriptivist accounts of the kind often associated with Frege (1892) and Russell (1905) are often intended as fine-grained theories. It is also natural to see “variabilist” theories (Dever 1998 §2.3, Cumming 2008, Pickel 2015, Schoubye forthcoming) as fine-grained theories, since even though two differently-indexed variables may have the same value relative to one assignment, relative to a different assignment function they may not. Many non-descriptivist variants on the ideas of Frege (1892) are also fine-grained, since they predict that when “Hesperus” and “Phosphorus” are embedded in attitude reports, they will not have the same compositional semantic value (because they will have different referents).

3I will speak throughout the paper as though any satisfactory theory must accommodate intuitive true readings of 3-5 and related sentences. But I am in fact open to the idea that the best overall theory may predict that these sentences do not have true readings at all (see Goodman and Lederman forthcoming §11). This paper can be read as exploring what the best theory would be on the assumption that a satisfactory theory
Two appendices discuss some further issues. Appendix A offers some more concrete ways of understanding the abstract models I employ in the paper, and appendix B shows how my theory can be extended to handle some puzzling potential data about negative quantifiers due to Charlow and Sharvit 2014.

2 A problem for the concept-generator theory

In this section I present the concept-generator theory (which I will refer to as CG-theory) first published in Percus and Sauerland 2003 (building on notes of Irene Heim), and further developed by Anand 2006 and Charlow and Sharvit 2014. I argue that this theory cannot account for relevant true readings of 3-5. This problem with the CG-theory provides some motivation for a non-Millian alternative, which I develop in the following sections.

The CG-theory is stated in a standard possible-worlds semantics for attitude verbs in the tradition of Hintikka 1962. We assume in the background a set of worlds $W$ and a set of individuals $X$, along with a function $\text{DOX} : X \rightarrow (W \rightarrow \mathcal{P}(W))$ which delivers for each individual $x$ and each world $w$ where $x$ has beliefs, the set of worlds that are consistent with $x$’s beliefs at $w$. The theory is distinctive in its use of “concept-generators”. A concept-generator is a function from individuals to individual concepts, where an individual concept is in turn a function from worlds to individuals. I will use @ throughout for the actual world, and I will require in my discussion that for any individual $x$ and concept-generator $G$, $G(x)(@) = x$. The basic idea of the CG-theory is that when names occur within the scope of attitude verbs, they are the arguments of covert pronouns whose denotations are concept generators. These concept generators operate on individuals (the type of denotations of names) to produce individual concepts (the type of denotations of definite descriptions).

Let’s first see how this theory works for 1. A simplified syntax of this sentence, at an appropriate level of abstraction, would be:

will deliver an intuitive true reading of these sentences, leaving it open that broader theoretical considerations could lead us ultimately to reject that assumption.

While writing Goodman and Lederman forthcoming §9, Jeremy Goodman and I recognized a version of this problem for versions of our own theory. At the time I did not appreciate that the problem arose also for Percus and Sauerland 2003.
Plato believes λG₅

λs₁

t₅ is visible in the evening

First, I will show how (relative to an assignment function) the whole phrase below λs₁ computes to a function from worlds to truth-values. Then I’ll turn to the whole sentence.

We assume that the usual assignment function is extended to the class of indices Gᵢ and the class of indices sᵢ, where i is a natural number, and that on the former class of indices it takes concept generators as values, while on the latter it takes worlds as values. In the syntax above, a bound pronoun which (relative to an assignment) denotes a concept generator occurs as sister to the name “Hesperus”. These two elements compose by function application to produce an individual concept: for instance if the denotation of t₅ is a concept generator G on an assignment g then the value [t₅ Hesperus] would be G(Hesperus), a function from worlds to individuals. This value composes in turn with the (bound) world-pronoun tₕ to produce an individual; if the value of g(s₁) is a world w, then [t₅ Hesperus-tₕ] is G(Hesperus)(w). Given a straightforward semantics for the predicate “is visible in the evening” (on which it denotes a function from worlds to functions from individuals to truth-values), [t₅ Hesperus tₕ is visible in the evening tₕ] will evaluate to a truth-value: 1 if G₅(Hesperus)(w) is visible in the evening at w, and 0 otherwise. And given this, the whole clause beneath λs₁ evaluates to λw.[t₅ Hesperus-tₕ] is bright-tₕ[w/s₁], that is, a function from worlds to truth-values. (In general I use g[x/α] for the assignment function such that for every index β ≠ α g[x/α](β) = g(β), and g[x/α] = x.)

In the simplest, standard setting, we would be done at this point: the first argument of “believe” would be a function from worlds to truth-values, and we could give a flatfooted lexical entry for “believe” in terms of DOX. But in the syntax above, the first argument of “believe” will not simply be a function from worlds to truth-values; it will instead be a function from concept-generators to functions from worlds to truth-values. So the new syntax requires a more complex lexical entry for “believe”. This lexical entry needs to accommodate not just the case where there is exactly one abstraction over concept-generators just below “believe” (as in the syntax above); it also needs to accommodate cases where there is no such
abstraction, or where there is more than one. The following clause is designed to deal with the variation in the type of the argument of “believe”:

**CG-Believes** $[\text{believes}] = \lambda p. \lambda x. \lambda w. \text{either for all } w' \in DOX(x)(w), p(w') = 1, \text{ or for some } n \geq 1, \text{ there are } G_1, ..., G_n \text{ which are salient for } x \text{ and such that for all } w' \in DOX(x)(w), p(G_1)...(G_n)(w') = 1^5$

We assume informally that for each individual $x$, context supplies a set of concept-generators which are salient relative to $x$. Here and throughout, I associate function application to the left, so $p(G_1)(G_2)(w)$ is properly $((p(G_1))(G_2))(w)$. The first disjunct (“either...”) covers the case where there is no abstraction over concept-generators below “believe”, so that the complement of “believe” is simply a function from worlds to truth-values. The second (“or...”) covers the more interesting cases, where $p$ is a function from concept-generators to functions from concept-generators...to functions from worlds to truth-values. In effect, the lexical entry introduces a sequence of existential quantifiers over concept-generators, of the exact length needed to saturate the first argument of “believes” to produce a function from worlds to truth-values.

Let us suppose that Plato believes that the planet which he calls “Hesperus” is visible in the evening and that Plato does not believe that the planet which he calls “Phosphorus” is visible in the evening. Given this background, in contexts where a concept-generator $G$ which when applied to the planet Venus yields the individual concept expressed by the definite description “the planet which Plato calls ‘Hesperus’” is salient relative to Plato, $1^1$ will be true. By contrast, in contexts where the only concept-generators which are salient relative to Plato are such that when applied to the planet Venus they yield the individual concept expressed by the definite description “the planet which Plato calls ‘Phosphorus’”, $1^2$ will be false. Since the names “Hesperus” and “Phosphorus” have the same semantic value, in a given context substituting one name for the other makes no difference to the final computation: the contexts in which $1^1$ is true are exactly those in which $1^2$ is true. But if we assume that typical utterances of $1^1$ against our background story suggest contexts of the first kind just described, while typical utterances of $1^2$ suggest contexts of the second kind, we can explain the apparent contrast between these sentences.

The CG-theory offers a promising story about the contrast between $1^1$ and $1^2$. But, as I will now show, this natural story cannot be extended to $1^5$.

According to the CG-theory, the following is the natural syntax for the VP of $1^5$ at an appropriate level of abstraction:

---

$^5$Strictly speaking, this clause only governs the case where $DOX(x)(w)$ is defined; for the case where it is undefined, we assume that the entry returns 0 regardless of the complement. This issue won’t be important for the remainder of the section, so I won’t mention it again, but subsequent lexical entries for attitude verbs should be understood to be restricted to the case where $DOX(x)(w)$ is defined.
Given our lexical entry for “believe” we compute the following property as the denotation of the VP of 5 with the above syntax:

6. \( \lambda x. \lambda w. \) there are concept generators \( G_1 \) and \( G_2 \) which are salient relative to \( x \) such that for all \( w' \in DOX(x)(w) \), \( G_1(\text{Hesperus})(w') = G_2(\text{Phosphorus})(w') \).

This property will be satisfied by any \( x \) and \( w \) whatsoever, provided there is a single concept-generator \( G^* \) that is salient for \( x \). For by instantiating the existential quantifiers over concept-generators \( G_1 \) and \( G_2 \) in 6 with \( G^* \) we obtain:

7. \( \lambda x. \lambda w \) for all \( w' \in DOX(x)(w) \), \( G^*(\text{Hesperus})(w') = G^*(\text{Phosphorus})(w') \).

Since Hesperus=Phosphorus, for any world \( w' \), \( G^*(\text{Hesperus})(w') = G^*(\text{Phosphorus})(w') \). And since this holds for all worlds \( w' \), it follows that for any \( x \) and any \( w \), it will hold for all \( w' \in DOX(x)(w) \). So the VP will be satisfied by any individual and any world in any context where some concept-generators are salient relative to that individual.

These very weak satisfaction conditions for the VP give very demanding satisfaction conditions for the negated VP, and for the whole sentence: on this theory, 5 will be true only in contexts in which no concept-generators are salient relative to Plato (and similar points apply to 3 and 4). But contexts of this kind yield bizarre readings of attitude ascriptions. In such a context, “Plato did not believe Athens was a city”, and “Plato did not believe Socrates was a philosopher”, would be true, as would variants with “did not know” or “was not sure” in place of “did not believe”. Since the CG-theory predicts that 3-5 are true only in such contexts, it fails to do justice to the idea that these sentences have the intuitive true readings that they seem to have: readings on which the sentences describe Plato’s specific ignorance or lack of opinion about a particular astronomical fact.\(^a\)

\(^a\)Allowing one of the names to take a world argument which is bound outside the scope of “believe”, while the other is bound underneath “believe”, would allow a somewhat intuitive true reading of our sentence. But this approach does not generalize to closely related sentences, which would also naturally be taken to be true.
In working this example, I have assumed that the copula “is” can express the relation of identity. But the argument does not depend essentially on this assumption. I could have run the argument with the following sentences instead:

8. Plato did not believe that Hesperus shares its center of mass with Phosphorus.

9. Plato did not believe that Hesperus has matter in common with Phosphorus.

10. Plato did not believe that Hesperus is coextensive with Phosphorus.

On the natural assumption that relative to every relevant way of thinking about Venus, Plato believed that the planet shares its center of mass with itself, believed that it has matter in common with itself, and believed that it is coextensive with itself, the CG-theory would predict that none of these sentences have the intuitive true readings they seem to have. In the rest of the paper, I will continue to discuss the problem I’ve developed here in terms of the examples 3-5. The main reason for this is that my own theory will involve a non-standard treatment of identity, which is highlighted by the way it handles these examples. But the reader who is (rightly) concerned about the behavior of the copula when it occurs in the scope of attitude verbs may understand my references to these sentences as references to 8-10 instead; my formal treatments of sentences featuring identity can be extended straightforwardly to these other sentences as well.

There is a tradition, often associated with Quine [1956] and Kaplan [1968], of distinguishing between “de re” and “de dicto” readings of reports like 5. In light of this tradition, one might see 3-5 not as posing a problem for Percus and Sauerland’s theory, but instead as showing that their theory of the de re readings of such reports must be supplemented with a further theory of the de dicto readings of them. But this response solves one problem only by creating a new, different one. For the traditional distinction between de re and de dicto readings of sentences like 5 is not in good standing. There is strong evidence for such a distinction between readings of ascriptions which feature overt definite descriptions or quantifiers. It is easy to feel a difference between two ways of understanding sentences like “Plato thought the star which rises in the evening did not rise in the evening” or “Plato thought every planet was not a planet”. Moreover, the same kind of ambiguity is evident in sentences where definites and quantifiers interact with modal and temporal operators (e.g. “it could have been that the stars which rise in the evening did not rise in the evening”, “in ancient times, the star which rises in the evening did not rise in the evening”). But there is no similar felt change of

7 Indeed, in correspondence Percus and Sauerland have said that their theory should be supplemented in this way; see also Sauerland [2015, p. 77].
perspective between readings of 5 and referential uses of names signally do not exhibit such an ambiguity when they interact with modal or temporal operators. More generally, I am not aware of any direct evidence that sentences like 5 exhibit this ambiguity (for some further discussion see e.g. Cumming [2016]). So if this argument shows that Percus and Sauerland’s theory must distinguish \textit{de re} and \textit{de dicto} readings of such sentences, it is still an argument against that theory: it shows that the theory requires postulating an ambiguity for which there is no direct evidence. We should prefer theories – like the ones I will develop below – which do not require postulating such an ambiguity.

The argument of this section narrowly targets what I have called the “CG-theory”, that is, the main theory found in Percus and Sauerland [2003], Anand [2006] and Charlow and Sharvit [2014]. It does not apply to all Millian theories, or even all Millian theories which use the machinery of concept-generators. One way of responding to the argument would be to develop a Millian variant on the CG-theory which avoids the problem with 3-5. I will consider and argue against such a response much later in the paper, in section 8. But first, I will explore a different response. While the argument above relies on features of the CG-theory which go beyond Millianism itself, Millianism is one assumption of the CG-theory which helps to create the problem. As I will show in the next section, non-Millian, fine-grained theories can smoothly account for reasonable true readings of 3-5. This makes it natural to wonder whether it was Millianism itself that led us astray, so that a fine-grained theory might in the end be preferable to a Millian one. I explore this question by developing a fine-grained theory in detail in Sections 4-7. Only after this theory has been presented, will I return to the question of whether it should be preferred to a Millian variant on the CG-theory (section 8).

The CG-theory was developed as a way of compositionally implementing an earlier theory, due to Cresswell and Von Stechow [1982], who in turn were developing ideas of Lewis [1979]. But there are important differences between the CG-theory and these earlier theories. Cresswell and von Stechow assume that attitude-verbs express relations to structured entities, consisting of sequences of objects and properties. (In this regard they see themselves as building on ideas of Lewis [1970]; cf. Cresswell [1975].) This aspect of their theory is key to how they would handle 5. They would say that this sentence is structurally ambiguous between a reading in which the prejacent of “believe” expresses a structured complex consisting of a binary relation and two individuals (or, really, the planet Venus taken twice over), and a reading on which this prejacent expresses a structured complex consisting of a property (being identical to Venus) and an individual (Venus). On the latter reading, they would predict that the sentence 5 has a reasonable true interpretation, since there is no requirement that the contextually salient ways of thinking about the property of being identical to Venus correspond to the contextually salient ways of thinking about Venus itself. While this flexibility allows Cresswell and von Stechow to deal with the present problem, it is an unattractive feature of their system. It is implausible that the sentence exhibits a structural ambiguity of this kind, and, in any case, the approach generates a vast array of readings of attitude reports, without obvious principled ways of narrowing them down. For the rest of the paper, I will focus on approaches to these reports which do not rely on structured propositions.

I focus in this paper on variations on the CG-theory, but there are other Millian theories which predict a true reading of “Plato did not know that Hesperus is Hesperus” (and of 4 and 5 as well). Cable [2018] is one such approach; I discuss it in fn. 54. Others are Crimmins and Perry [1989], Crimmins [1992] and the theories described in Goodman and Lederman forthcoming §9.1 and §9.2.
3 A basic fine-grained semantics

In this section I present a simple model of a fine-grained theory, which allows a reasonable true reading of 5.

In presenting my model, I’ll re-use some notation from the less formal presentation of the CG-theory; from now on the notation should be taken to have the meanings I give it here. Our basic class of models has the following ingredients:

- $W$, a non-empty set, thought of as the set of worlds;
- $D_e$, a set;
- $DOX : D_e \to (W \rightarrow \mathcal{P}(W))$, a function which, for each element of $D_e$, returns a partial function which maps each world where the individual corresponding to that element of $D_e$ has beliefs to a nonempty set of “doxastically possible” worlds for that individual at that world;
- $R \subseteq W \times W$, an equivalence relation on $W$, thought of as representing relative possibility, as used in the semantics for the modal “it’s necessary that”;
- $E : W \to \mathcal{P}(D_e \times D_e)$ a function from worlds to equivalence relations on $D_e$, used to give the semantics for the “is” of identity, and such that if $wRw'$, then $E(w) = E(w')$.

For readability in what follows, I will often subscript world-arguments, so for example, I will write $E_w$ for $E(w)$. I will use 2 for the set of truth values $\{0, 1\}$ and $D_p$ for the set of functions from worlds to truth values, i.e. $2^W$. I sometimes call these “propositions”.

Two aspects of this model will be unfamiliar. First, in not requiring $R$ to be the universal relation on $W$, we allow $W$ to contain some worlds which are intuitively “impossible” relative to others. Second, the “is” of identity is interpreted not by model-theoretic identity, but by possibly non-trivial equivalence relations $E_w$ on $D_e$, which can vary across impossible worlds. The elements of $D_e$ are thus not to be thought of as individuals; instead we should think of individuals as standing in a natural bijection with equivalence classes under $E_{@}$. I will sometimes say that individuals “are represented by” or “correspond to” such equivalence classes. By this I mean no more than that there is this natural bijection between individuals and these equivalence classes.

I’ll return to these aspects of the model theory in a moment, but first, let’s see how the semantics allows us to deliver a reasonable trivial true reading of 5. Consider the following toy model from our class of models, in which $D_e = \{h, p, p1\}$, $W = \{@, i\}$, $R$ is the identity

---

10For simplicity in the formal treatments in the paper I won’t consider variability across times; I will pretend that the only dimension of variability for these relations is world-variability. I also won’t consider issues connected to contingent existence or non-denoting names. (I don’t think either of these presents any real challenge; to handle the latter, we can use variable domains in a straightforward way.)
relation on $W$, $E_i$ is the smallest equivalence relation which relates $h$ and $p$, $E_i$ is model-theoretic identity on $D_e$, and finally for all $w \in W$, $DOX(pl)(w) = \{i\}$. Here and throughout, I will use @ to denote the actual world. Here then is a simple fragment interpreted on this model, with a flatfooted entry for “believe” that I will revise later on (the entries here are all insensitive to the assignment function $g$):

- $[\text{Hesperus}]^g = \lambda w. h,$
- $[\text{Phosphorus}]^g = \lambda w. p,$
- $[\text{Plato}]^g = \lambda w. pl$
- $[\text{is}]^g = \lambda w. \lambda x. \lambda y. x E_w y;$
- $[\text{it’s not the case that}]^g = \lambda w. \lambda x \in 2.1 - x.$

I will assume that not just negation, but also conjunction and disjunction behave classically at all worlds, both possible and impossible. The impossible worlds I employ will thus differ very little from standard possible worlds, as I’ll discuss further in a moment.\[11\]

**Believe (Preliminary)** $[\text{believe}]^g = \lambda w. \lambda p. \lambda x. \forall w' \in DOX(x)(w) \ p(w') = 1^{12}$

We can now give a straightforward treatment of 5. The set $DOX(pl)(@) = \{i\}$, and it is not the case that $hE_i p$. So “Plato does not believe Hesperus is Phosphorus” is true at all worlds in our model (as is “Plato believes Hesperus is not Phosphorus”). More generally, in any model in which the set $DOX(pl)(@)$ contains any (impossible) worlds $w$ such that $\neg hE_w p$ then “Plato does not believe Hesperus is Phosphorus” will be true. Relative to our toy model, not only 5 is true, but so are other attitude reports, such as “Plato believes that Hesperus

---

\[11\] Throughout the the paper I assume an extensional treatment of modality, in which covert world-pronouns occur in the syntax of sentences, and abstraction over world-pronouns is used to produce a proposition (a function from worlds to truth-values) when one is required. A sentence which features an intensional operator like “necessarily” or “believe”, which takes a proposition (and not a truth-value) as its argument, will be uninterpretable unless an abstractor over worlds occurs below the intensional operator. Arguments that something like this theory is needed can be found in Fodor [1970], Bäuerle [1983], Percus [2000], Keshet [2008]. For the sake of familiarity, I use the simplest version of the theory, where the semantic value of each expression is a function defined on worlds, and thus every expression is assumed to have a world-pronoun as its sister. As is well known, this simple theory overgenerates; further assumptions about the distribution of world-pronouns are needed. In the rest of the paper, I won’t discuss this issue further, and in working examples, will simply cherry-pick my preferred syntax. But for the record, my favorite theory which uses world-pronouns is that of Schwarz [2012] (where the only constituents which take world-pronouns are determiners); everything I do below could be defined in Schwarz’s more restrictive setting. My basic approach also does not require the extensional treatment of modality: one could instead develop it using alternative approaches to the “de re” / “de dicto” or “transparent” / “opaque” ambiguity, for instance, a “split intensionality” theory (Keshet [2008] [2010] [2011]).

\[12\] Again, technically, this only governs the case where $DOX(x)(w)$ is defined; the sentence should be taken to be false regardless of its complement if $DOX(x)(w)$ is undefined. But this issue won’t matter at all below, so I won’t mention it again.
was Hesperus” and “Plato believes Hesperus is not Phosphorus”. Unlike the CG-theory, then, the present theory allows for a true reading of [5] without appealing to a reading of “believe” on which Plato does not believe (basically) anything at all.

This simple, abstract model thus allows us to make reasonable predictions about [5]. But have we given a semantics for this sentence? Fittingly, the word “semantics” means different things to different people. Philosophers often use it to mark a contrast between model theory and semantics (Burgess [2008]). Here, a model theory is a formal tool for studying properties of the language (or a logic defined for the language), while a semantics is a theory of the meanings of the expressions of a language, which is typically taken to require making claims about what there is. For instance, if a semantics in this sense says that the meaning of an expression is a set of possible worlds, then the person who advances the semantics is taken to claim that sets and possible worlds exist. In this paper I do not take myself to be giving a semantics in this sense. I do not claim that there are impossible worlds, or that there are elements of $D_c$. But there is another way the word “semantics” is used, for instance in one common use of “formal semantics”. In this different sense, a semantics is a model which is used to make predictions both about the truth and falsity of sentences in context and about entailment relations among sentences (Yalcin [2018]). A model of this kind does not say what the meanings of expressions really are. It is just a model, to be judged on the basis of its simplicity, tractability and predictive strength. When I call my theory a “fine-grained semantics”, I mean only that it is a semantics in this latter sense.

As the basis for a semantics in this second sense, my models are comparable to possible-worlds models: they differ only slightly from more standard models in their simplicity, tractability and predictive strength. As in standard possible-worlds models, at every world in every model I consider, Boolean connectives such as “it’s not the case that” behave completely standardly. As a result the propositions I will consider themselves form a Boolean algebra under the usual set-theoretic operations. The only non-standard feature of the models is that identity is interpreted by a non-trivial equivalence relation on $D_c$, an equivalence relation which can vary from world to world. This small deviation from the assumptions in possible worlds semantics is precisely what allows us to deliver a reasonable true reading of [5].

Still, one might wonder: is there any way of viewing these models as the basis for a semantics in the philosopher’s sense? I do not see answering this question as crucial for my purposes here: my goal is simply to provide an adequate formal model, not to answer

---

13This way of using impossible worlds thus avoids some standard arguments against the utility of more deviant impossible worlds (see Bjerring [2013], and Bjerring and Schwarz [2017]).

I will furthermore require that at all worlds, possible or impossible, identity is a congruence with respect to the denotation of intuitively extensional predicates. For example, I will assume that at every world $w$, the semantic value of “is bright” applied to $w$ and $x \in D_c$ is 1 if and only if for every $y$ such that $xE_w y$, the denotation of “is bright” applied to $w$ and $y$ is 1. This constraint means that for intuitively extensional predicates $F$, we will also have the law: if anyone believes that $x$ is $y$ then they believe that $x$ is $F$ if and only if they believe that $y$ is $F$. 
these hard questions about the real meanings of expressions. But, as I show in appendix A, a wide array of fine-grained theories of names can develop what they might take to be a semantics in the philosopher’s sense using models that are isomorphic to a subclass of mine. Descriptivists and predicativists can view elements of $D_e$ as individual concepts (functions from worlds to individuals), variabilists can take them to be the indices of variables, and theorists of many different stripes can see them as mental representations of some kind, for instance, names in a language of thought. Although I will be officially neutral on how to think about elements of $D_e$ for most the main text of the paper, in light of the first and third of these more concrete versions of the theory, I will sometimes speak of elements of $D_e$ heuristically as “ways of thinking about” individuals. For the most part this locution is just meant as a synonym for “element of $D_e$”, though there are some times when it bears more weight in motivating a particular way of developing the theory. In addition to providing ways of reconstruing the elements of $D_e$, appendix A also discusses how particular fine-grained theorists might reconstrue my impossible worlds. For instance, in section A.1 I show that a subclass of my models can be constructed only from resources available within the CG-theory (or for that matter the theory of Aloni [2005]). The appendix thus shows that someone could develop a semantics based on my model theory which would commit them only to entities that a proponent of the CG-theory or of Aloni’s theory should already accept.

4 Basic Surrogatism

In this section I consider how to extend the fine-grained theory from the previous section to sentences featuring quantifiers.

Consider first the following example:

**Context** Mercury and Venus are the only interior planets (i.e. planets closer to the sun than earth). Suppose that Venus is visible in the evening, but that Mercury is not.

11. At least two interior planets are visible in the evening.

This sentence should be false: there is only one interior planet, Venus, which is visible in the evening. But our semantics will not obviously deliver this result, since there are two elements

---

14 A similar abstract methodology is pursued by Thomason [1980] and Pollard [2015], though their approaches differ in other significant ways from mine. (Thomason treats names as having the type of monadic quantifiers, while Pollard does not consider the issues with determiners I discuss in the next section.)

15 A related idea that I don’t explore in detail might be that the element of $D_e$ denoted by a name is the customary Fregean sense associated with that name. Such a theory would correspond to a “Fregean” view according to which the compositional semantic value of an expression is its sense (and not its reference), and the senses of sentences are modeled by sets of impossible worlds. But Fregeans typically think that the arguments of attitude verbs are structured like the sentences which express them, so while one could see my account as a version of this form of Fregeanism, the fact that this version of the theory would predict that, for example, “John is happy and John is happy” has the same sense as “John is happy” would make it an unusual form of Fregeanism at best.
of $D_e$, the semantic value of “Hesperus”, and the semantic value of “Phosphorus”, which satisfy the predicate “is visible in the evening”.  

The basic problem is clear: we do not want “at least two” to count elements of $D_e$, but instead to count individuals, which correspond to equivalence classes of elements of $D_e$ under $E_w$. A simple way of solving the problem – and the one I will adopt here – is to assume a mandatory and stringent form of domain restriction, on which the only admissible domains for the quantifier at a world draw exactly one element from each (relevant) equivalence class at that world. This element of $D_e$ then acts as a “surrogate” or “proxy” for the equivalence class to which it belongs; we can count equivalence classes (and thus individuals) by counting their surrogates.  

Formally, a function $S : W \rightarrow \mathcal{P}(D_e)$ is a surrogate domain restriction if and only if for every $w \in W$ and every $X \in I_w$ there is exactly one $x \in X$ in $S(w)$. (Recall that $I_w$ is the set of equivalence classes of $D_e$ under $E_w$.) We assume that context supplies a surrogate domain restriction $S$, and then use the following lexical entry for the quantifier “at least two”:

Two

$[\text{at least two}]^{9:S} = \lambda w. \lambda F. \lambda G. \text{at least two } x \in S(w) \text{ are such that } F(x) = 1 \text{ and } G(x) = 1$.

The requirement that quantifiers be restricted by a surrogate domain restriction eliminates the problem with 11. For any $S$, the proposition expressed by an utterance of that sentence (assuming the most natural syntax) would be:

- $\lambda w. \text{ for at least two } x \in S(w) \text{ } x \text{ is an interior planet at } w \text{ and visible in the evening at } w$.  

Regardless of what surrogate restriction is chosen, this proposition will be false. For the equivalence class corresponding to Mercury does not have an element which is visible in the evening.
evening, and no equivalence class other than the ones corresponding to Mercury and Venus have elements which are interior planets. Any element of the equivalence class corresponding to Venus will be an interior planet at \( @ \) and also be visible in the evening at \( @ \), but there is only one such entity in the domain of the quantifier. Since the proposition is true only if there are at least two such entities in the domain of the quantifier, the proposition is false.\(^{18}\)

Surrogatist domain restrictions are similar in important ways to Maria Aloni’s conceptual covers (Aloni [2005]). In fact, there is a class of my models in which the conceptual covers are simply a subclass of the surrogatist domain restrictions (see n. 52 and surrounding text). One could see the remainder of the paper as presenting problems for Aloni’s theory and showing one way the theory could be extended to solve those problems. Indeed, for some readers, this may be a helpful perspective on the project of the paper more generally: as arguing that the best overall theory of attitude reports combines key elements of Aloni’s proposal with key elements of the CG-theory.\(^{19}\)

I’ll call the proposal that all determiners are mandatorily restricted by surrogate domain restrictions, while attitude verbs are given the simple semantics from section 3 Basic Surrogatism. In the next section I’ll present three problems for this theory and go on to propose a refinement of it.\(^{20}\)

\(^{18}\)Here I’ve used locutions like “\( x \) is an interior planet at \( w \)” as a shorthand for “the denotation of ‘is an interior planet’ applied to \( w \) and then \( x \) is 1”, and I’ll continue to do this throughout. But the denotations of predicates operate on elements of \( D_e \), not on individuals (which stand in bijection not with elements of \( D_e \) but with equivalence classes of them). So while I will say that such elements “are interior planets at \( w \)”, we should remember that the relevant entities do not correspond to individuals but stand for the semantic values of names.

\(^{19}\)Throughout the paper I will assume that a surrogate domain restriction is supplied by context and can change from context to context. But on some more concrete ways of viewing my model theory, a single surrogate domain restriction may be singled out as distinguished, and it may be natural to see it as imposed in every context. For instance, on the descriptivist proposal I present in appendix A.1 one could take the surrogate domain restriction in every context to be the set of functions which return the same individual at every world (and in fact this will be a conceptual cover). This idea can be extended to a variable domain model, by using the set of functions which return the same individual at every world where the individual exists, and some dummy value otherwise. Since I’m interested in providing a general theory, I won’t assume that some surrogate domain restriction is distinguished.

\(^{20}\)There are two salient alternatives to Surrogatism; in this note I’ll say briefly why I disprefer them. Alternative “Existentialist” and “Universalist” proposals are as follow:

**Existentialist Two**

\[
[\text{at least two}]^9 = \lambda w. \lambda F. \lambda G. \lambda w’ \text{ for at least two } Z \in I_w, \exists x \in Z \text{ such that } F(x)(w) = 1, \text{ and } \exists y \in Z, \text{ such that } G(y)(w’) = 1.
\]

**Universalist Two**

\[
[\text{at least two}]^9 = \lambda w. \lambda F. \lambda G. \lambda w’ \text{ at least two } Z \in I_w \text{ are such that } \forall x \in Z \ F(w)(x) = 1 \text{ and } \forall x \in Z \ G(w’)(x) = 1.
\]

The generalizations of these proposals to all determiners are unattractive because they allow that in a single context, there could be false instances of “not every \( F \) is \( G \) if and only if some \( F \) is not \( G \)” (where the conditional is interpreted as material); i.e. they predict that the universal and existential quantifiers would not be duals. For instance, the Existential proposal allows that an instance of “Every \( x \) is \( F \)” could be true in a context while the corresponding instance of “Some \( x \) is not \( F \)” is also true in the same context. They also lead to problematic results with determiners which involve counting e.g. an instance of “exactly half of the \( F \)s are \( G \)” could be true in a context where the corresponding instance of “it’s not true that exactly half of the \( F \)s are not \( G \)” is also true.
In Basic Surrogatism, the world-argument of a determiner has an important new role: it controls which equivalence-classes stand as proxy for the domain of individuals for the determiner (reflected in the fact that $S_w$ is defined with respect to $I_w$, i.e. equivalence classes with respect to the identity relation as interpreted at that world). We can motivate this feature of the proposal (and see how it works in more detail) by considering two further examples:

**Context** Suppose Plato believed that earth was the planet closest to the sun, so that there were no interior planets. Suppose furthermore that he believed that Hesperus and Phosphorus were two distinct exterior planets, believed that they were bright, and believed that Mercury was not bright.

12. Plato believed at least two exterior planets were bright.

13. Plato believed exactly one interior planet was bright.

Each of these sentences has a true reading in this context. The second may be easier to access by considering the dialogue “Venus and Mercury are the interior planets, Plato believed that Venus was bright and Plato did not believe that Mercury was bright. So Plato believed exactly one interior planet was bright.”

The salient true reading of 12 results from an “opaque” or *de dicto* interpretation of “at least two”, that is, an interpretation on which its world argument is bound below the attitude verb “believed”. For instance, the relevant syntax might be represented as “$\lambda w.\text{Plato-}w\text{ believed-}w\lambda w'.\text{at least two-}w'\text{ exterior planets-}w'\text{ were bright }w'$.” Using Surrogatist Two, the sentence on this regimentation would express the following proposition:

- $\lambda w.\text{for all }w'\in DOX(\text{Plato})(w)\text{ for at least two }x\in S(w') x\text{ is an exterior planet at }w'\text{ and }x\text{ is bright at }w'$.

Since exactly one $x$ is chosen from each equivalence class in $I_{w'}$, (which correspond to the individuals there would be if this world were the actual one), this proposition requires us

---

The universalist proposal leads to odd results in other cases as well. Suppose that Plato thought that Mercury rises in the evening, although it never does, and suppose he thought no planets other than Mercury and Hesperus rose in the evening (in particular he did not think that Phosphorus rose in the evening). Even so it seems true to say:

- At least one planet which Plato thought rose in the evening, does rise in the evening.

The Universalist proposal predicts that the sentence would be false.

Many hybrid proposals (using a blend of Existentialism or Universalism or Surrogatism for different quantifiers) will lead to similarly undesirable results about entailment patterns between quantifiers. One proposal that I don’t know of such a problem for would require surrogatist restrictions for quantifiers which are sensitive to number e.g. “most”, “both”, “at least two”, but not require such restrictions for quantifiers which are not, e.g. “every”, “some” “no”. This proposal could be a way of reconciling data like 11 with arguments in Cai et al. [2020] for treating “every” and “some” as unrestricted even in a fine-grained setting.
to count individuals at Plato’s belief-worlds. And the proposition will be true. For in this scenario, it is clear that the denotations of “Phosphorus” and “Hesperus” occupy different equivalence classes at Plato’s belief-worlds (Plato thinks they are distinct planets). Since these elements of $D_e$ satisfy the restrictor predicate (they are exterior planets) and the nuclear scope predicate (they are bright) at Plato’s belief-worlds, every element of their equivalence classes at those worlds must also satisfy both the restrictor and the nuclear scope property at those worlds. (Recall that we are assuming that intuitively extensional predicates are congruences with respect to $E_w$ at every world $w$; see n. [13]) So, regardless of the choice of surrogate from these equivalence classes, there will indeed be two distinct equivalence classes with elements which satisfy these properties.

The salient true reading of 13, by contrast, results from a “transparent” or de re interpretation of “exactly one”, that is, an interpretation on which its world argument (and the world argument of “exterior planets”) is bound outside the scope of the attitude verb “believed”. For instance, the relevant syntax might be represented as “$\lambda w. \text{Plato}_w \text{believed}_w \lambda w'. \text{exactly one}_w \text{interior planet}_w \text{were bright}_w$.” Using the obvious Surrogatist entry for “exactly one”, the sentence would express the following proposition

- $\lambda w. \text{for all } w' \in \text{DOX(Plato)}(w) \text{ exactly one } x \in S(w) \text{ is an interior planet at } w \text{ and is bright at } w'$.

Note here that the world arguments of $S$ and of “interior planet” are bound by the highest-scope binder over worlds, not by a binder under “believe”. As a result this proposition will also be true. There are two $Z \in I_@$ such that all of their elements are interior planets at @: the classes corresponding to Venus on the one hand, and Mercury on the other. By assumption one and only one of these classes has elements which are bright at $w'$ for all $w' \in \text{DOX(Plato)}(\@)$ (and we may assume that all of the elements of this equivalence class, including the denotations of “Hesperus” and of “Phosphorus” satisfy this condition). So, regardless of our choice of surrogate for these equivalence classes, the proposition expressed will be true.

Surrogate domain restrictions help us to solve the problem with 11. They also give rise to a constrained way of determining which domain a quantifier ranges over, based on its world-argument. This second feature allows us smoothly to account for varying domains in iterated reports, as in the different readings of “John thinks Mary hopes two people are coming for

---

[Barker 2016] develops a rich theory which is in some important ways related to mine. But, as Barker acknowledges, his theory cannot produce opaque (i.e. de dicto) readings of quantifiers inside attitude reports, so he cannot produce the relevant true reading of 12.

[22] It does not seem possible to separate the transparent/opaque interpretation of the restrictor of a determiner from the choice of which domain is used in counting by a determiner, suggesting that the world-pronouns of these two constituents should be coindexed. In my preferred setting, that of [Schwarz 2012], only determiners take world-arguments in the syntax, so the desirable requirement that the restrictor and the determiner are assessed at the same world is imposed essentially automatically.
dinner”. Since the treatment of such iterated reports is straightforward, I won’t describe it in detail. But since many fine-grained theories become very complex when they attempt to handle such iterated reports, it is an important feature of the present account that this generalization is so straightforward.23

5 Three Problems for Basic Surrogatism

In this section I present three problems for Basic Surrogatism, which the CG-theory handles smoothly. In the next section I will respond to the problems by presenting a theory which combines some key ideas from the CG-theory with the fine-grained semantics I’ve developed to this point.

As I discuss in more detail later, in section 7, the examples I will present go beyond and sharpen examples which have previously been used to argue for various aspects of the CG-theory (for instance, its use of existential quantification over concept-generators).

5.1 Beyond double vision

A traditional argument against theories like Basic Surrogatism, which hold that names and variables are associated with (only) a single precise way of thinking about an individual, is based on what is often called “double vision”. This argument, usually attributed to Quine

23 The system to this point (and also the final system of the paper) is naturally seen as predicting that the following are false:

(i) There is an \( x \) and there is a \( y \) such that \( x \) is \( y \) but Plato did not know that \( x \) was \( y \);
(ii) There is an \( x \) and there is a \( y \) such that \( x \) is \( y \) but Plato did not know that \( x \) was coextensive with \( y \).

On the (desirable) assumption that the surrogate domain restriction is the same for each occurrence of “there is a” in (i), that sentence will express the same proposition as “There’s an \( x \) such that Plato did not know that \( x \) was \( x \)”, which has no intuitive true interpretation in my system, since every element of \( D_e \) bears \( E_w \) to itself at every world. A similar point holds for (ii). There is thus an interesting difference between the way the system handles distinct coreferring names (as in (i) and (ii)), and the way it handles distinct variables governed by quantifiers which are assessed at the same world ((i) and (ii) are essentially the existential generalizations of (i) and (ii)) respectively). Neither (i) nor (ii) is an English sentence, and I don’t know of convincing English examples that tell against this prediction of my theory. The system does not make analogous predictions if distinct pronouns are simply bound by an abstractor which is not in turn operated on by an overt quantifier (e.g. “John and Jim are such that Mary didn’t know he was him”), or if two coreferential pronouns used referentially occur in the complement clause of an attitude report. The system handles e.g. “John doesn’t know that is that” (where the two demonstrations pick out the same object) straightforwardly, by assigning the two occurrences of “that” different elements of \( D_e \) which are related by \( E_\alpha \) (for the example, see Perry


The theory to this point also predicts that the following are false:

(iii) There’s an \( x \) and there’s a \( y \) such that \( x \) is \( y \) but Plato believed \( x \) wasn’t \( y \);
(iv) There’s an \( x \) and there’s a \( y \) such that \( x \) is \( y \) but Plato believed \( x \) wasn’t coextensive with \( y \).

But, as I will discuss in sections 7.3 and 8, my final theory will treat sentences with negation over the relevant attitude verb (as in (i) and (ii)) quite differently from sentences with the negation inside the scope of the attitude verb (as in (iii) and (iv)), and the final theory allows both (iii) and (iv) to be true.
starts from the following example:

**Context** Ralph sees Ortcutt by the docks. Ralph concludes on the basis of what he sees that Ortcutt is a spy. Later, Ralph watches Ortcutt’s mayoral inauguration address on TV. Ralph thinks that no mayor could possibly be a spy; the background checks are simply too rigorous. So he concludes that Ortcutt the mayor is not a spy.

14. Ralph believes that Ortcutt is a spy.

15. Ralph believes that Ortcutt is not a spy.

Since there is no precise way of thinking about Ortcutt relative to which Ralph both thinks that Ortcutt is a spy and thinks that Ortcutt is not a spy, if [14] and [15] are true in the same context, then “Ortcutt” cannot be associated with a single precise way of thinking about that individual.

This argument has played a central role in the development of semantic theories of attitude reports. But it is not as strong as it might seem. One can readily deny that [14] and [15] are true in the same context, while nevertheless maintaining that both sentences are typically true when uttered. In the fine-grained setting, for instance, one might hold that names are context-sensitive, and can denote different elements of $D_e$ in different contexts.

Someone who endorsed this form of contextualism could allow for the truth of [14] and [15] by claiming that typical utterances of these sentences occur in different contexts. Although there are no overt lexical cues in the sentences that might tip a hearer off that they are to be assessed in different contexts, the mere fact that they are uttered, along with hearers’ tendencies to charitably interpret utterances as true, might be thought to suffice for them to be interpreted in different contexts.

I have presented this contextualist response not to endorse it, but to illustrate a weakness of the traditional argument. I will now argue that the contextualist response is not workable.

---

24It might seem that for all I have said $D_e$ could contain “relaxed” or “disjunctive” ways of thinking about individuals as well as precise ones, so that “Ortcutt” could be associated with a single element of $D_e$ even if it is not associated with a precise way of thinking about this individual. For instance, perhaps there could be a single element $o \in D_e$, such that if one comes to believe that Ortcutt is a spy by seeing him at the docks, one believes the proposition $\lambda w. o$ is-a-spy-at-$w$, and if one comes to believe that Ortcutt is not a spy by seeing him on TV, one believes the proposition $\lambda w. o$ is-not-a-spy-at-$w$. But the existence of such an $o$ is ruled out by the fact that negation is interpreted classically at all worlds in the model theory. Provided a person has any belief-worlds (and we may assume that Ortcutt does) they will not believe the proposition $\lambda w. x$ is-a-spy-at-$w$ while also believing the proposition $\lambda w. x$ is-not-a-spy-at-$w$ for any $x$ in $D_e$. Of course we could relax this assumption about negation in the model theory, but doing so would come at the cost of a significant loss in predictive power.

25An alternative form of contextualist might treat the “double vision” case as an example of what Blumberg and Lederman [2020] call “revisionist reports” (for discussion, see Blumberg and Lederman [2020, §7]. This treatment is also not sufficiently general to handle my example below.

26As others have also noticed (see Anand [2006, p. 24-5]) a similar point applies to the use of double-vision to motivate the use of existential quantification over concept-generators within the CG-theory. I discuss this point in detail in section 7.1.
As I will discuss later, my next examples suggest that while the traditional argument based on double vision was too quick, the conclusion that was drawn from it was nevertheless correct.

**Context** John has four pictures in front of him, two pictures each of two teachers. The teachers are Alice and Bo; we think think of the photos of Alice as $A_1$ and $A_2$, and the photos of Bo as $B_1$ and $B_2$. John thinks that the photos are of four distinct people. He points at $A_1$, $A_2$ and $B_1$ and says as he points to each of them “this person is Italian”. He then points at the last picture, $B_2$, and says “this person is French”. As a matter of fact teacher Alice is Italian and Bo is French.27

16. Someone John thinks is French is French.

17. ?Everyone John thinks is Italian is Italian.

18. Someone John thinks is Italian is French.

19. ?No one John thinks is French is French.

The sentences 16 and 18 are naturally heard as true, whereas the sentences 17 and 19 are naturally heard as false. (They all have true, and false, readings in this scenario; I am only claiming that there is a contrast in immediate acceptability between these pairs.) Given very natural assumptions Basic Surrogatism – which in effect assigns traces, like names, single elements of $D_e$ – predicts that 16 is true in a context if and only if 17 is true in that context, and that 18 is true in a context if and only if 19 is true in that context. Moreover, it predicts that 16 is true in a context if and only if 19 is false in that context.

Since Basic Surrogatism makes these predictions, it is hard to see how a contextualist version of it could explain the above judgments about 16-19. According to the contextualist, 14 and 15 are interpreted as true because hearers are charitably inclined to interpret them in contexts where they would be true. This story could be extended to explain how both 16 and 18 are interpreted as true, in spite of the fact that Basic Surrogatism predicts that there are no contexts in which both are true. But the theory cannot explain why 17 and 19 are false. If the contextualist is right, we would expect that here, too, hearers would work to interpret this sentence in a context in which it is true (e.g. by interpreting the first in the natural contexts in which 16 is true, and the second in the natural contexts in which 18 is true). But this is not what we observe. A flatfooted appeal to hearers’ charity wrongly predicts that each pair of sentences would be equally acceptable.28

---

27This style of “pictures” case was introduced by Charlow and Sharvit [2014]. The example sentences used here are new.

28One might think that the different words “French” and “Italian” in the complement clauses of the reports above suggest different contexts for the relevant reports. But this feature of the examples is inessential. If we substitute “is not Italian” for the relevant occurrences of “is French” in 16 and 19 and substitute “is not French” for the relevant occurrences of “is Italian” in 17 and 18 the modified examples lead to the same
Simpler examples could have been used to show that a blanket appeal to charity to explain \([14]\) and \([15]\) leads to incorrect predictions. But in section 7.1 I will explain in more detail how the extra structure in my example contributes to a more general argument against subtler contextualist accounts of the truth of \([14]\) and \([15]\) as well as the flatfooted one I’ve considered here.

For now, the moral is that Basic Surrogatism cannot accommodate these data. But the CG-theory can. And, as I will show below, a theory which adapts the key insights of the CG-theory to a fine-grained setting can get the best of both worlds, accommodating these data, while also allowing a true reading of \([35]\).

5.2 Problems with plural subjects

The following example, due to Cian Dorr, presents a different kind of problem for Basic Surrogatism:

**Context** (Based on Dorr, p.c.) Eve knows that the heavenly body she sees in the evening and calls “Hesperus” is a planet and not a star, but she thinks that the heavenly body she sees in the morning and calls “Phosphorus” is a star and not a planet. Dawn knows that the heavenly body she sees in the morning and calls “Phosphorus” is a planet and not a star, but she thinks the heavenly body she sees in the evening and calls “Hesperus” is a star and not a planet. Neither has encountered this heavenly body in any other way than via their evening and morning sightings. On Monday at noon, Eve learns that Phosphorus is a planet, while Dawn learns that Hesperus is a planet, so

20. On Monday at noon, Eve and Dawn learned that Venus is not a star.

21. There’s a heavenly body which Eve and Dawn learned is not a star on Monday at noon.

These sentences have true readings in this scenario. But this fact poses a problem for Basic Surrogatism. It is natural to think that if a person stands in the relation expressed by “learns” in a context at a time \(t\) to a proposition \(p\), then (i) the person did not stand in the relation expressed by “knows” in that context to \(p\) in an interval between some \(t’\) earlier than \(t\) and \(t\), which is open at \(t\), and (ii) the person does stand in the relation expressed by “knows” in that context to \(p\) at \(t\) itself. The problem is that, to the extent that we have a grip on when different names are assigned different element of \(D_e\) and how those elements compose with the denotations of predicates, it is hard to see how there could be an element \(x\) of \(D_e\) pattern of judgments of acceptability and unacceptability. The difference also can’t be attributed merely to the use of the universal quantifier and negative universal rather than the existential, since “Every teacher John thinks is French is French” is acceptable, while “Some teacher John thinks is French is Italian” is not.

22 Jeremey Goodman and I discussed a related example in earlier versions of Goodman and Lederman [forthcoming].
that composes with the denotation of “is not a star” (given the appropriate abstraction over
world-pronouns) to produce a proposition p such that (i) neither Eve nor Dawn stood in the
relation expressed by “knows” to p before Monday at noon, and (ii) both Eve and Dawn
stood in the relation expressed by “knows” to p on Monday at noon. For example, if there is
a \( p \in D_e \) such that every occurrence of “Phosphorus” in the vignette above expresses \( \lambda w.p \),
and similarly an \( h \in D_e \) (where \( h \neq p \)) such that every occurrence of “Hesperus” expresses
\( \lambda w.h \), then s and c will both fail (i): at all times on Monday morning, Eve knew that Hesperus
was a planet and not a star, and Dawn knew that Phosphorus was a planet and not a star.

Once again, although Basic Surrogatism cannot handle this example, I will show that,
like the CG-theory itself, a theory which adapts elements of the CG-theory to a fine-grained
setting can. Moreover, in section 7.2 I’ll show that the constraints imposed on the CG-theory
by this example are in an important sense independent of those imposed by 16-19.

5.3 The bound \textit{de re}

A final problem for Basic Surrogatism comes from examples discussed by Soames [1989-90,
p. 198f.] (cf. Higginbotham 1991, p. 362 ex. 42) and, more extensively, Soames [1994],
which have recently been brought back into the spotlight by Sharvit 2010 and Charlow and
Sharvit 2014:

\textbf{Context} John knows that Jupiter is bigger than Mars, and that Mars orbits the sun faster
than Jupiter. He believes no planet is bigger than Jupiter, and no two planets are
exactly the same size. He thinks that Hesperus is Jupiter and thinks that Phosphorus
is Mars.

22. There’s something John thinks is Jupiter and is Mars.

23. There’s a planet which John thinks is as big as Jupiter and orbits the sun as fast as
Mars.

Intuitively these sentences are true. But Basic Surrogatism cannot predict this result. There
isn’t any way of thinking about Venus such that, relative to that way of thinking about it,
John thinks Venus is Jupiter and Venus is Mars. For John knows that Mars and Jupiter are
distinct. Similar points hold for 23.

\footnote{A different solution than the one I will adopt would be to allow type-raising of the semantic value of
“Venus”, so that instead of denoting a constant function from worlds to an element of \( D_e \), it denotes a function
from worlds to a function \( f : D_e \rightarrow D_e \). This function could be interpreted as “\( x \)’s Venus”; it would map
attitude holders to ways of thinking about Venus, allowing variability from attitude-holder to attitude-holder.

As far as I can see this proposal would handle the present example, but it does not give us a way of
handling examples like 16-19: these examples seem to require some kind of existential quantification over
ways of thinking. My way of handling these examples makes it natural to handle Dorr’s data by allowing the
domain of permutations to vary from attitude-holder to attitude-holder, rendering this kind of type-shifting
unnecessary.}
Charlow and Sharvit [2014] show that the CG-theory naturally predicts true readings of these examples. I’ll show below that a fine-grained theory which takes over ideas from the CG-theory can handle them too. Moreover, in section 7.3 I’ll discuss how the constraints imposed on the CG-theory by this example are in an important sense independent of those imposed by the other examples in this section.

6 Fine-grained Semantics

Although the CG-theory makes incorrect predictions about 3-5, it smoothly handles all of the data presented in the previous section. Basic Surrogatism smoothly handles 3-5, but it makes incorrect predictions about all of the data in the previous section. In this section I show how one can enrich Basic Surrogatism using ideas from the CG-theory to produce a theory which handles both sets of data.

A bijection \( \pi : D_e \to D_e \) is a permutation. A permutation \( \pi \) is \( w \)-admissible if and only if for all \( x, \pi(x) \in E_w \); for short I’ll call \( w \)-admissible permutations \( w \)-permutations. A \( w \)-permutation can map different values within the same \( E_w \) equivalence class to different values, but it can only map elements of an equivalence class to other elements of the same equivalence class. For example, there are an \( \alpha \)-permutations which map the semantic value of “Hesperus” to the semantic value of “Phosphorus”. But there are no \( \alpha \)-permutations that map the semantic value of “Hesperus” to the semantic value of “Mars”.

By analogy to the CG-theory, I will assume that any occurrences of names or e-type variables in the scope of attitude verbs are “wrapped” by variables denoting permutations, which are obligatorily bound by an abstractor. To account for these new variables \( t_{\pi_i} \), I assume that the assignment function \( g \) is extended to be defined on new indices \( \pi_i \) for all \( i > 0 \) and that these indices are assigned permutations. Thus for instance, imitating the syntax of the CG-theory, the syntax for the VP of 5 will be:

---

31There are no data I’m aware of that motivate using permutations rather than arbitrary functions from \( D_e \) to \( D_e \) (including those which are not bijections). But since there are also no data I’m aware of that require using functions that are not permutations, it seems preferable to use the more restrictive notion (and readers have found it easier to work with, as well).
We assume that context supplies a function $f$ which, for each person and world, returns a set of permutations which are salient relative to that person and admissible at that world. If we think heuristically of elements of $D_e$ as “ways of thinking” about individuals, we can see this $f$ as induced by contextually supplied equivalence relations among ways of thinking about individuals, which are defined relative to a person and a world. In some contexts, speakers take certain ways of thinking about objects to be equivalent relative to certain thinkers, while others are not. For instance in some contexts the way of thinking about the planet Venus associated with the name “Hesperus” is taken to be equivalent with the way of thinking about Venus associated with the name “Phosphorus” relative to Plato and the actual world; the conversational participants might be indifferent to how Plato thinks of the planet at the actual world, and hence choose to disregard the difference between whether Plato believes (for instance) the proposition typically expressed by “Hesperus is bright” or the proposition typically expressed by “Phosphorus is bright”. But in other contexts, the relevant ways of thinking about Venus may not be taken to be equivalent relative to Plato and the actual world; the conversational participants might care a great deal about whether Plato thinks about Venus in one way as opposed to another, and they might care a great deal about the difference in mental state between someone who believes (for instance) the proposition typically expressed by “Hesperus is bright” as opposed to the proposition typically expressed by “Phosphorus is bright”. Assuming that context supplies one such equivalence relation among ways of thinking about things for each person and each world, this equivalence relation gives rise to a natural set of permutations for each person and world, namely, the set of permutations which map every way of thinking to a way of thinking that is contextually equivalent relative to that relevant person and world. If we take this set of permutations as the value of $f$ relative to that person and world, then contexts where differences between “Hesperus” and “Phosphorus” are unimportant relative to Plato and the actual world will be associated with an $f$ such that $f(Plato, @)$ contains a permutation which maps the denotation of one to the other. By contrast, contexts like the second, where this difference is important,
will be associated with an $f$ such that $f(\text{Plato}, @)$ contains no permutation which maps one to the other.\(^\text{32}\)

Given this background, in the case of “believe” my proposal will be:

**Believe** $\langle \text{believe} \rangle_{g,S,f} = \lambda w. \lambda p. \lambda x. \text{either for all } w' \in \text{DOX}(x)(w), \ p(w') = 1,$ or for some $n \geq 1,$ and some $\pi_1 \ldots \pi_n \in f(x, w), \forall w' \in \text{DOX}(x)(w), \ p(\pi_1) \ldots (\pi_n)(w') = 1.$

As above, the first disjunct (“either...”) covers the case where the relevant argument of “believe” is just a proposition, while the second disjunct (“or...”) covers the more interesting case, where $p$ is a function from permutations to functions from permutations...to functions from worlds to truth-values.\(^\text{33}\)

If we assume that there are $h, p \in D_e$ such that $\langle \text{Hesperus} \rangle_{g,S,f} = \lambda w. h$ and $\langle \text{Phosphorus} \rangle_{g,S,f} = \lambda w. p$, then using this lexical entry (and after a series of simplifications), the displayed clause computes to:

- $\lambda x.$ there are $\pi_1, \pi_2 \in f(x, [s_2]_{g,S,f})$ such that for all $w \in \text{DOX}(x)([s_2]_{g,S,f}), (\pi_1 h)E_w(\pi_2 p)$.

This denotation of the VP is not trivially satisfied, as one can see by considering a context where for all $x$ and $w$, $f(x, w)$ is the singleton set consisting of the identity function on $D_e$. (This permutation is $w$-admissible for all $w$.) Under this assumption the clause will reduce to

- $\lambda x.$ for all $w \in \text{DOX}(x)([s_2]_{g,S,f}), hE_w p$,

which as we saw in section 3 is not trivially satisfied. The reader may readily verify that less restrictive assumptions about $f$ will also yield the result that the property expressed is not trivially satisfied, so that \(^5\) (as well as \(^3\) and \(^4\)) will have reasonable true readings in a range of contexts.\(^\text{34}\)

---

\(^{32}\)The informal discussion using “contextual equivalence” makes it natural to impose further constraints on the values of $f(x, w)$ for each $x$ and $w$. In particular, we should require that the set of permutations supplied for any world and individual by $f$ form a *group*: they should contain the identity permutation, and be closed under composition and inverses. Moreover, they should satisfy the further constraint that if a permutation $\pi$ is such that for each $a \in D_e$ there is a $\pi' \in f(x, w)$ such that $\pi(a) = \pi'(a)$, then $\pi \in f(x, w)$. For ease of exposition I won’t discuss these constraints further in what follows, but I think of the official theory as imposing both of them.

\(^{33}\)The extra parameter $w$ in $f$ is needed to handle iterated attitude reports. When an attitude verb is embedded in another intensional operator, the chosen permutations should be admissible relative to the worlds at which the embedded attitude verb is assessed; they should not (oddly) be required to be admissible in the worlds of the speaker’s context.

\(^{34}\)In n. 31 noted that there was no particular motivation for using permutations rather than the general class of functions from $D_e$ to $D_e$, but that I preferred the more restrictive use of permutations, and $w$-permutations. An incommensurable way of restricting the class of all functions from $D_e$ to $D_e$ would be to require that the values of $f(x, w)$ be $w$-collapse functions, where a function $\kappa : D_e \rightarrow D_e$ is a *$w$-collapse function* if and only if for all $x, y$ such that $x \in W_y, \underbrace{\kappa(x) = \kappa(y)}$. But there is a clear reason to reject this proposal, namely, that it would rule out our earlier treatment of \(^3\)\(^5\). Since $[\text{Hesperus}]E_w([\text{Phosphorus}])$, every $@$-collapse-function $\kappa$ has $\kappa([\text{Hesperus}]) = \kappa([\text{Phosphorus}])$, so that given the syntax above \(^3\) would be synonymous with “Plato didn’t know that Hesperus is Hesperus”, precisely the result our fine-grained semantics was designed to avoid. The restriction to permutations does not have this problematic result.
Given the assumption that when names occur inside attitude reports, permutation pronouns take them as arguments, the exact semantic values of names within a given equivalence class of $E_@$ no longer have real significance: these values are simply place-holders. Provided “Hesperus” and “Phosphorus” have distinct semantic values, our permutations can map them to (different) distinct values, and it is not important what the starting values are, so long as they are distinct. Still, although formally there is nothing important about the exact values we assign to names, it is natural to require that the identity function will always be an element of $f(x, w)$ for all $x$ and $w$ (as discussed in n. 32). If we make this assumption, then the choice of semantic values for names does matter.

35

To produce a fully predictive theory, we need an account of how features of speakers’ psychology and surroundings make particular permutations and surrogates salient. In this regard, my theory is on a par with the CG-theory: the CG-theory similarly stands in need of an account of why particular concept-generators are salient in particular conversations. (The notion of “acquaintance”, which proponents of the CG-theory typically appeal to, has yet to receive a sufficiently substantive characterization to yield a predictive account.) How to fill this lacuna is an urgent question both for my theory and for the CG-theory. But I will follow proponents of the CG-theory in setting it aside for now. My hope is that, once we have a model which makes reasonable predictions about truth, falsity and entailment among relevant sentences, we will be in a better position to fill in this gap.

36

One might wonder how the theory handles sentences which involve quantification into sentence position, like:

- John believes everything Mary believes.

I’ll show how by showing that we can extend the usual model theory to allow quantification over all of the variable-type arguments of “believe”. (I set aside world-pronouns for simplicity; adding them in is mechanical.) Let $\pi$ be the type of permutations. Then as usual we have base types $e, p$ for names and sentences respectively, and in addition $\pi$ for permutations. The simple variable type is $\chi$. The types are then the members of the smallest set containing $e, p, \pi, \chi$, and such that if $\sigma$ and $\tau$ are members of the set, so is $\sigma \rightarrow \tau$. A simple abnormal type is anything of the form $\pi_1 \rightarrow (\ldots (\pi_n \rightarrow t)))$. Expressions of type $\chi \rightarrow \tau$ combine with an expression of a simple abnormal type to produce an expression of type $\tau$. For instance, the type of attitude verbs (again, ignoring the world-pronouns we had above) is $\chi \rightarrow (e \rightarrow p)$. $\pi$ can then be regimented as $\forall^X(\lambda p^X, \lambda x^X)$, if Mary believes $P^X$ then John believes $P^X$). We can assume a set $W$ of worlds and let $D_p$ be $P(W)$, $D_e$ be an arbitrary set, and $\Pi$ be the set of permutations on that set. Domains for higher types are defined as usual, and booleans and the quantifiers can be interpreted in the standard way, with quantifiers for the simple variable type ranging over all elements of domains for simple abnormal types. Thanks to Peter Fritz here.

36

One might wonder whether in a fine-grained setting, as opposed to a Millian one, we could use a lexical entry which appeals to transformations of the whole embedded complement clause, and not just of the denotation of names within it. More precisely, say that a proposition $p'$ is a $w$-variant of a proposition $p$ if and only if for all $w'$ such that $w \in DOX(x)$, $p'(w') = 1$. This entry is essentially the lexical entry of Richard [1990], transposed to the present unstructured setting. The entry can accommodate all of the data we have considered to this point, and has the great advantage over the official theory of not requiring a complex syntax with permutation variables. This theory allows us to use the simple syntax of Basic Surrogatism; the arguments of attitude verbs are straightforwardly propositions.

The idea behind this simpler theory is attractive. But the theory itself allows very many readings which
7 Solving the problems

In this section, I’ll show how this extended fine-grained theory solves the three problems I described for Basic Surrogatism (sections 7.1-7.3). In each subsection I will also discuss more formally how my examples constrain the theory I’ve developed, as well as the CG-theory itself.

7.1 Beyond double vision

I suggested that on their most salient readings, 16 and 18 are true, while 17 and 19 are false. I’ll now show how my theory accounts for this contrast.

In spelling out the predictions of my theory formally, I’ll call the equivalence classes corresponding to each teacher A and B, and I’ll call the elements of these classes corresponding to the four pictures, \(a_1, a_2, b_1, b_2\). I will assume that the relevant elements of A are exactly \(a_1\) and \(a_2\) and similarly that the relevant elements of B are exactly \(b_1\) and \(b_2\). This assumption is very natural, given that we have not supposed that John knows about these individuals in any way other than the pictures, and this is all that is made salient about those individuals in our vignette. Finally, I will also suppose that every @-permutation of the domain is salient relative to John at the actual world (i.e. that \(f(\text{John}, @)\) is the set of all @-permutations. This assumption is not strictly required to produce the results I’ll describe, but it is a natural assumption which gives rise to the contrast.

Relative to any choice of \(f\) and \(S\), our lexical entry for “believe” predicts that on the most natural syntax 16 expresses:

- \(\lambda w. \text{there is an } x \in S_w \text{ such that for some } \pi \in f(\text{John}, w), \text{ for all } w' \in \text{DOX(John)}(w), \pi(x) \text{ is French at } w' \text{ and } x \text{ is French at } w\).

Given our assumptions about \(f\) and the domain of quantification, this proposition will be true. Regardless of the choice of surrogate of B (whether it is \(b_1\) or \(b_2\)), there is an @-permutation which maps this surrogate to \(b_2\), which is French at John’s belief-worlds. Regardless of the choice of surrogate of B, that surrogate is French at @ (since every element of B is). So the surrogate of B witnesses the existential “there is an \(x \in S_w\)”.

are not observed. Suppose that if a person is female, they are necessarily female, and that John mistakenly believes that Queen Elizabeth is male. Given these assumptions, the theory allows a true reading of “John believes 2 + 2 = 5”. For provided the proposition that Queen Elizabeth is male is salient relative to John, that proposition would be a proposition which is true at all the same possible worlds (i.e. none) that 2 + 2 = 5 is. This prediction seems absurd: a mistake about Queen Elizabeth’s sex does not amount to a mistake about simple mathematics.

The simpler theory requires further constraints on \(w\)-variants for it to be a viable, predictive alternative to the theory in the main text. I haven’t been able to find reasonable such constraints, so I will continue to work with the more complex lexical entry and the more complex syntax with permutation variables.

Note that an alternative way of simplifying the lexical entries for attitude verbs, by complicating the lexical entry for complementizers (as in Cresswell and Von Stechow [1982]), is available in both the Millian and the fine-grained setting.
But under the same assumptions, we predict that [17] will be false. Relative to any $S$ and $f$, our lexical entry for “believe” predicts that on the most natural syntax [17] expresses:

- $\lambda w. \text{for every } x \in S_w \text{ if some } \pi \in f(\text{John}, w) \text{ is such that for all } w' \in DOX(\text{John})(w)$, $\pi(x)$ is Italian at $w'$, then $x$ is Italian at $w$.

Regardless of the choice of surrogate of $B$, there is an @-permutation which maps this surrogate to $b_1$. So the surrogate of $B$ satisfies the antecedent of the conditional. But, again, regardless of the choice of surrogate of $B$, that surrogate is French at @ (and hence not Italian at @). So the surrogate of $B$ is a counterexample to the universal “for every $x \in S_w$”, and the proposition is false.

The reader may readily verify that [18] will similarly be predicted to be true, and [19] be predicted to be false, under the same assumptions.

We can see how this example constrains the official theory by comparing it to an alternative. A functionalist theory assumes that the range of $f$ consists only of singleton sets of permutations (or, equivalently that the range of $f$ is just the set of permutations, not the set of sets of permutations). By contrast, existentialist theories allow that non-singleton sets may be in the range of $f$. For example, here is a functionalist lexical entry for “believe”:

**Functionalist Believe**

$[\text{believe}]^{g,S,f} = \lambda p. \lambda x. \lambda w. \forall w' \in DOX(x)(w'), p(f(x, w))(w') = 1$ \textsuperscript{37}

In this entry I’ve assumed that the values of $f$ are just permutations, rather than singleton sets of permutations. I’ve also left out complications required to deal with cases where there are different numbers of permutation variables bound by the verb, since they won’t matter in my discussion. The official theory of the paper is existentialist, not functionalist.

Under the natural assumptions I made at the start of this section, functionalist theories cannot explain the contrast between [16] and [17] or between [18] and [19]. They predict that [16] is true in a context if and only if [17] is true in that context, and that [18] is true in a context if and only if [19] is true in that context. \textsuperscript{38}

This point holds not just for functionalist theories in my fine-grained setting, but also for analogous theories which use concept-generators in a Millian setting (think of $f$ in the entry above as supplying a single concept-generator for each individual and world). Such a proposal – which \cite{Anand2006} p. 25] calls the “Skolemized” proposal, but I will call “functionalist CG-theory” – will also predict that [16] is true in a context if and only if [17] is, and similarly [18] will be true in a context if and only if [19] is (at least, given analogues of the natural assumptions above). So again the proposal cannot explain the contrast between the members of these pairs.

\textsuperscript{37}A nice feature of functionalist proposals is that if extended to modals they would preserve the duality of “must” and “might”; if extended to modals my theory would fail to do this. Thanks to John Hawthorne here.

\textsuperscript{38}See above n. \textsuperscript{28} for arguments against a radical contextualist theory.
I will now discuss in some detail how this argument against the functionalist CG-theory complements earlier arguments against this theory from Anand [2006]. In section 5.1 I noted that Basic Surrogatism predicts that 14 and 15 cannot be true in a single context. But, as I said, if the theory is coupled with a form of contextualism about names, it can accommodate the judgment that these sentences are both true. Anand [2006, p. 24-5] (citing Zimmerman [1991] and Heim [1998]) begins his discussion by making a related point about the functionalist CG-theory. This theory predicts that 14 and 15 cannot be true in a single context, but it allows that each is true in some context, and so can accommodate the judgment that both are heard as true.

Anand [2006, p. 32-33] then develops two new arguments against the functionalist CG-theory. My diagnoses of his arguments are quite similar, so I will only discuss one of them in the main text (for the second, see n. 42). The argument is based on the following case (which I quote):

**Context** Ralph, John, and Bill all see Ortcutt in the same locales, and all come to the dual belief that Ortcutt is a spy and that he’s not a spy.

24. Each man thinks that Ortcutt is a spy.

25. # No man thinks that Ortcutt is a spy.

A functionalist theory (whether Millian or fine-grained) will predict that there are (different) contexts in which both 24 and 25 are true. Anand takes this point to be evidence against the functionalist theory, and ends his argument there. I agree that the example brings out an important challenge for the functionalist theory, but I think more has to be said about what the exact challenge is. Consider the following elaboration of Anand’s case:

**Context** Ralph, John and Bill are three independent investigators working to root out corruption in the town, and that they have all come to suspect that Olson, the police chief, is a spy. One night, while watching over the docks, they all saw someone – as it happens, Ortcutt the mayor – in shady circumstances, and concluded that the person is a spy. But none has correctly identified this person yet, and in fact they all suspect the person they saw was Olson. They drew this conclusion in part because they believe on strong evidence that Ortcutt the mayor is not a spy. Thus, although each man thinks Olson is a spy, 25 no man thinks Ortcutt is a spy.

This example is just a more detailed version of Anand’s case: as in Anand’s case the three men all know Ortcutt in two different ways; relative to one, they believe he is a spy, and relative

---

39 For some of my consultants the final sentence is improved if one adds “yet” before “thinks” or changes “no man” to “no investigator”, but they agree that the sentence is true in this setting.
to another they believe he is not a spy. But, while after hearing Anand’s underspecified story it is most natural to hear \(25\) as false, after hearing mine it is most natural to hear this same sentence as true. So the fact that functionalist theories predict that \(25\) has a true reading is not on its own evidence against that theory. On the contrary, everyone – whether functionalist or existentialist – should agree that \(25\) can used truly to describe Anand’s case. We should of course hope for a predictive account of how these two ways of telling the story lead us to understand this sentence in different ways. But everyone needs an account of this kind, not just the functionalist.

Still, as I have said, Anand’s case does provides evidence against the functionalist theory. As I discussed in section 5.1, the most obvious way for the functionalist to account for the change in context between \(14\) and \(15\) is to say that hearers charitably search for readings of these sentences on which they are true. But Anand’s example shows that a flat-footed application of this idea overgenerates: there are true readings of \(25\) but hearers do not always naturally access those true readings. So Anand’s case shows that functionalists need a more nuanced story about how \(14\) and \(15\) are both heard as true, one which does not also predict that \(25\) will be heard as true whenever there is a true reading of it.

We can now see how my example strengthens Anand’s case against the functionalist. A functionalist might attempt to account for the difference between Quine’s examples and Anand’s by holding that certain readings are “easier” to access in response to different stories, and that hearers interpret a sentence as true if and only if there is a reading on which it is true that is sufficiently easy to access. The idea would then be that in the original Ralph story it is sufficiently easy to access both a true reading of \(14\) and a true reading of \(15\), but after hearing Anand’s story it is only sufficiently easy to access a true reading of \(24\) and not of \(25\). This blueprint of a story does not pretend to be explanatory or predictive, but let’s suppose it could be spelled out in a more satisfying way. The problem is that, even if it could, it would still fail to account for my examples. Since \(16\) and \(17\) are true in the same relevant contexts, the functionalist should hold that a true reading of \(17\) will be just as easy to access as a true reading of \(17\) (and similarly for \(18\) and \(19\)).

This particular statement of the problem might make it seem that the issue arises only for

---

\[40\]It might seem that even simpler arguments could be given against the functionalist position by focusing on “Ralph does not think that Ortcutt is a spy” (which, unlike \(15\) i.e. “Ralph thinks Ortcutt is not a spy”, has a negation over the main verb), but there are good reasons to focus instead, as Anand does, on ascriptions with quantified subjects. First, “think” tends to exhibit what is often called “neg-raising”, that is, main-clause negations (“does not think”) are readily interpreted as negating only the complement clause of the verb (“thinks it is not the case that”). Second, as Anand says, judgments about sentences with main-clause negations are actually very delicate (see Anand [2006, p. 21], discussing a proposal of Abusch). Even if we control for problems about neg-raising by using an expression like “is sure” the judgments in related sentences remain less clear.

\[41\]Related points apply to an account of \(14\), \(15\), \(24\) and \(25\) as what Bumblung and Lederman [2020] call “revisionist reports”. While such an account seems to offer some promise for these four sentences it is unclear to me how to generalize it to \(16\), \(19\).
one particular outline of a functionalist response to Anand’s cases. But the challenge behind
the problem is more general. Since functionalists predict that [16] and [17] are true in exactly the
same contexts, it is just extremely hard to see how any reasonable story could be told about
why one is heard as true, and the other as false. It is even harder to see how such a story
could be told which would also predict that [14] and [15] are both true, [24] is true and [25] is false
(after Anand’s story). By contrast the existentialist faces a very different, and less daunting
challenge: they need only to tell a story about why [25] is naturally heard as false after hearing
Anand’s story, and true after hearing mine. It is not obvious how an existentialist should
meet this challenge, but there is no principled reason to doubt that it can be met.

7.2 Problems with plural subjects

Here, I will work a simpler example than [20], though the morals for [20] should be clear from
my discussion. The running example will be:

26. Eve and Dawn know that Venus is not a star.

The assumptions described above predict the following syntax for this sentence:

![Syntax Tree]

---

42 Anand’s second argument, “the argument from only” is based on a case from [Percus and Sauerland 2003, p. 234]. The argument has the same structure as the one we’ve just considered. Anand gives a story where it is natural to hear one sentence as true, and another as false, but observes that the functionalist will predict that there are true readings of each of them. As with the argument above, I think the problem here is not so much that the functionalist predicts a true reading of each of the relevant sentences (I think everyone should want to predict this), but rather that the fact that one of the sentences is read as false shows that a flat-footed story about how hearers access true readings of both [14] and [15] will overgenerate.

In response to this second example (along with others), Anand suggests that existentialists should hold that a concept-generator delivering the self concept for every individual is included in the domain of concept-generators salient for every individual in every context ([Anand 2006, p. 39-40]). Interestingly, this constraint itself suggests a way that functionalists could reply to Anand’s argument: they could claim that in general contexts where individuals’ self-concepts are salient relative to those individuals are somehow preferred if these concepts are relevant at all. But I myself am uncertain about Anand’s generalization, though this is not the place to discuss the issue in detail.
Suppose that there are three relevant elements of $D_e$, $h$, $p$ and $v$, corresponding to “Hesperus”, “Phosphorus” and “Venus”. In the context produced by the background story for we may suppose that, relative to Eve and all worlds, $p$ and $v$ are equivalent and $h$ is only equivalent to itself, while relative to Dawn and all actual worlds, $h$ and $v$ are equivalent and $p$ is only equivalent to itself. Supposing that $f(Eve, w)$ and $f(Dawn, w)$ are the sets of permutations $\pi$ such that for all $x \pi(x)$ is equivalent in their respective equivalence relations, then will be true relative to Monday at noon. Importantly, under the same specification of $f$, the sentence would not be true at any earlier times on Monday morning: at those times, Eve did not know or believe that Phosphorus was not a star, and Dawn did not know or believe that Hesperus was not a star.

The entry thus allows us to avoid the problem with No element of $D_e$ satisfies both (i) and (ii) (from section 5.2), but can be true in spite of this fact, because it may be that $f(Eve, @) \neq f(Dawn, @)$.

Once again it will help to see how this example constrains the official theory by considering an alternative class of theories. A theory is insensitive if it takes the parameter $f$ to be simply a function from worlds to sets of permutations; it is sensitive if it takes the parameter to be a function from worlds and individuals to sets of permutations. To illustrate, here is an insensitive lexical entry for “believe”.

**Existentialist Insensitive Believe**

$$[\text{believe}]_{g.S.f} = \lambda w. \lambda p. \lambda x. \text{for some } \pi \in f(w), \forall w' \in \text{DOX}(x)(w), p(\pi) = 1.$$  

Again, I’ve also left out complications required to deal with cases where there is not exactly one permutation variable bound just the verb, since this extra complexity won’t matter in the rest of this section. The official theory is, of course, sensitive.

Insensitive theories cannot accommodate a true reading of the sentence On such theories $f$ is sensitive only to a world argument, so the same set of permutations will be used for each attitude holder. If this set of permutations includes one which maps $[\text{Venus}]_{g.S.f}(w)$ to $[\text{Hesperus}]_{g.S.f}(w)$, or one which maps $[\text{Venus}]_{g.S.f}(w)$ to $[\text{Phosphorus}]_{g.S.f}(w)$, then the proposition expressed by the complement clause of “learned” will fail condition (i): either Eve or Dawn would have known it before. On the other hand, if the set of permutations contains no permutations which either map $[\text{Venus}]_{g.S.f}(w)$ to $[\text{Hesperus}]_{g.S.f}(w)$, or map $[\text{Venus}]_{g.S.f}(w)$ to $[\text{Phosphorus}]_{g.S.f}(w)$, then the proposition expressed by the complement clause of “learned” will fail (ii), since neither person will know the relevant proposition on Monday at noon.

---

43I’ve spoken of a sentence being true relative to Monday at noon so as not to take a stand on whether a single proposition may be true or false at different times. But nothing important here turns on how one settles that issue.

44Schiffer’s famous “Madonna problem” [Schiffer, 1992, p. 507-8] could be handled by either a sensitive theory, or by an existentialist one. Dorr’s example goes beyond standard arguments based on Schiffer’s example, by forcing a sensitive one. Moss [2012, p. 516] presents an example which similarly suggests a sensitive theory.
These basic points apply not just to fine-grained theories but to Millian ones as well. (For the Millian theory, simply take $f$ above to be a function from worlds to concept-generators.) In either setting, a sensitive functionalist theory could account for the true reading of 20 and 21 but not for the contrast between 16 and 17. In either setting, an insensitive existentialist theory could account for the contrast between 16 and 17 but not for the true reading of 20 and 21. In this sense, the two sets of examples impose independent constraints on fine-grained theories like mine as well as on the CG-theory itself.

7.3 The bound de re

I now rehearse observations from Charlow and Sharvit [2014] to show how my proposal handles 22. We assume the following syntax (abstracting from irrelevant world-pronouns and abstraction over worlds, and grouping some abstractions for the sake of space):

$\exists \lambda_3 \lambda_\pi_5 \lambda_\pi_7 \lambda_{s_1}$

$\lambda_3$

John

believes

$\lambda_\pi_5 \lambda_\pi_7 \lambda_{s_1}$

$t_{\pi_5} \quad t_3 \quad$ is Jupiter $s_1$ and $t_{\pi_7} \quad t_3 \quad$ is Mars $s_1$

The key point is that although the two occurrences of $t_3$ are forced by the usual clause for predicate abstraction to take the same value, different permutation pronouns take these two occurrences of the variable as arguments. Recall that in the setup for this example, John believes that Hesperus is Jupiter and Phosphorus is Mars. The clause below “$\lambda_{s_1}$” can express the proposition that Hesperus is Jupiter and Phosphorus is Mars relative to an assignment, if the value of $t_3$ relative to the assignment is the denotation of “Hesperus”, the value of the first permutation pronoun on this assignment maps the denotation of “Hesperus” to itself, and the value of the second permutation pronoun on this assignment maps the denotation of “Hesperus” to the denotation of “Phosphorus”. Thus our account can deliver an intuitive true reading of the sentence, since it allows different permutations to map the same element of $D_e$ to different elements of $D_e$. 
More formally, relative to a $g, S, f$, such that $[\text{Jupiter}]^{g, S, f}$ is $\lambda w.j$, while $[\text{Mars}]^{g, S, f}$ is $\lambda w.m$, and simplifying away the quantification over $n$, the clause below “John” will evaluate to:

- $\lambda x$. there are $\pi_1, \pi_2 \in f(x, w)$ such that for all $w' \in \text{DOX}(x)(w)$, $\pi_1([t_3]^{g, S, f})_{E_{w'}j}$ and $\pi_2([t_3]^{g, S, f})_{E_{w'}m}$.\(^{45}\)

And this condition can be non-trivially be satisfied, since $\pi_1$ and $\pi_2$ can vary independently.

Once again considering an alternative class of theories will help to show how the example constrains the official theory. A theory is type-simple if according to it, there is only a single pronoun for permutations; it is type-variable otherwise. To illustrate, here is one type-simple lexical entry for “believe”:

**Existentialist, Sensitive, Type-Simple Believe**

$[\text{believe}]^{g, S, f} = \lambda w. \lambda p. \lambda x. \text{for some } \pi \in f(x, w), \forall w' \in \text{DOX}(x)(w), p(\pi) = 1.$

Here the quantification over $n$ that appears in the official entry is no longer required: a single abstraction over permutations is guaranteed to bind any number of occurrences of the single pronoun for permutations. Type-simple theories allow the verb “believe” always to take an argument of the same type. The example lexical entries I gave in the previous two subsections were both type-simple. But the official theory is type-variable.

Type-simple theories cannot accommodate a true reading of 22. Since they assume that there is only one pronoun for permutations, they predict that in the appropriate version of the syntax displayed in section 7.3, the same permutation pronoun occurs as sister to both occurrences of the bound trace $t_3$. Thus the clause below “John” in the syntax displayed above would evaluate to:

- $\lambda x$. there is an $\pi \in f(x, w)$ such that for all $w' \in \text{DOX}(x)(w)$, $\pi([t_3]^{g, S, f})_{E_{w'}j}$ and $\pi([t_3]^{g, S, f})_{E_{w'}m}$.\(^{46}\)

Since John was assumed to know that Jupiter and Mars are distinct, he does not satisfy this condition: there is no single element of $D_e$ which stands in $E_{w'}$ to Jupiter and to Mars at any of his belief-worlds $w'$, never mind at all of them.

Once again, the constraints imposed on our theory by this example are in an important sense independent of the constraints imposed by the previous two sets of examples. For example, a functionalist, insensitive type-variable theory could deliver a true reading of 22, but it would predict neither the contrast between 16 and 17, nor a true reading of 20.

\(^{45}\)This assumes also that the world-argument of “believe” has been saturated by a world-pronoun which is not made explicit above.

\(^{46}\)Here I am assuming assuming the world-argument of “believe” has been saturated.
The following table summarizes the ways in which the examples constrain the final theory (as well as the CG-theory), and exhibits how the constraints they impose are independent from one another. “F” stands for “functionalist” and “E” for “existentialist”; “I” stands for “insensitive” and “S” for sensitive; “TS” stands for “Type-simple” and “TV” for “Type-variable”. “E, S, TS” is thus the official theory (and, in a Millian setting, the CG-theory itself).

<table>
<thead>
<tr>
<th></th>
<th>16 vs. 17 (Teachers)</th>
<th>20 (Dorr’s datum)</th>
<th>22 (Bound de re)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F,I,TS</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>E,I,TS</td>
<td>√</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>F,S,TS</td>
<td>x</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>E,S,TS</td>
<td>√</td>
<td>√</td>
<td>x</td>
</tr>
<tr>
<td>F,I,TV</td>
<td>x</td>
<td>x</td>
<td>√</td>
</tr>
<tr>
<td>E,I,TV</td>
<td>√</td>
<td>x</td>
<td>√</td>
</tr>
<tr>
<td>F,S,TV</td>
<td>x</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>E,S,TV</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

8 The Indexed-Domain CG-theory

At the end of section 2, I noted that there are Millian variants on the CG-theory which avoid the problems the CG-theory faced with 3-5. In this section, I consider such a variant on the CG-theory and argue that the fine-grained theory should be preferred to it.

An acceptable variant on the CG-theory must not only allow an intuitive reading of 5, but also (given the arguments of the previous section) be existentialist, sensitive, and type-variable.\[47\] Letting $f$ be a function from individuals to worlds to natural numbers to sets of concept-generators, the following is a minimal alteration of CG-Believes which satisfies these desiderata:

**Indexed-Domain CG-Believes** $[\text{believes}]^{g,f} = \lambda p. \lambda x. \lambda w. \text{either for all } w' \in \text{DOX}(x)(w), p(w') = 1 \text{ or for some } n \geq 1, \text{there are } G_1 \in f(x)(w)(1), ..., G_n \in f(x)(w)(n) \text{ such that for all } w' \in \text{DOX}(x)(w), p(G_1) ... (G_n)(w') = 1$\[48\]

\[47\]Perhaps the most obvious variant of the CG-theory which would avoid its predictions about 3-5 would be a functionalist theory. Such a theory can allow for an intuitive true reading of 5 provided the concept-generator variables “wrapping” the occurrences of “Hesperus” and “Phosphorus” are assigned different values, so that (in effect) these names are associated with different individual concepts. But as we saw in detail in section 7.1, there are a number of independent reasons to reject such a functionalist theory.

\[48\]The theories of Ninan [2012] and Rieppel [2017] produce essentially the same results as this lexical entry; it can be simply thought as an implementation of their theories using the machinery of the CG-theory. (Ninan sometimes uses set-notation and speaks of context as supplying “sets” of acquaintance relations, but he uses numerical indices on the acquaintance relations and in correspondence he confirmed that his intention was to have context supply a sequence of such relations.)
In the CG-theory, context supplied a set of concept-generators as salient relative to each individual; for this reason we could call it (and my theory) a single-domain theory. But in the Indexed-Domain CG-Believes, the extra numerical argument of \( f \) means that context now supplies a sequence of such sets; this is why I have called it an indexed-domain theory. The extra structure of these indexed domains allows different domains of concept-generators to be associated with different concept-generator variables, which allows the theory to avoid the problem with \([5]\). To see this point, consider an \( f \) such that \( f(\text{Plato})(1) \) contains a single concept-generator which when applied to Venus produces the individual concept corresponding to “the planet Plato sees in the evening”, while \( f(\text{Plato})(2) \) contains a single concept-generator which when applied to Venus produces the individual concept corresponding to “the planet Plato sees in the morning”. Relative to such an \( f \), Indexed-Domain CG-Believes will allow for an intuitive true reading of \([5]\) this sentence will be effectively interpreted as equivalent to “Plato did not believe that the planet Plato saw in the evening was the planet Plato saw in the morning”. The availability of this reading depends on the fact that \( f \) takes different values at different numerical arguments.

Should we prefer this theory, or mine? At first sight, the indexed-domain CG-theory might seem to have a clear advantage in terms of simplicity over mine. Since it is a Millian theory, it does not require that we have multiple elements of \( D_e \) corresponding to a single individual, or that we employ the machinery of surrogate domain-restrictions.\([49]\) These benefits in simplicity come at what might seem the small cost of adding an additional numerical argument to the function which determines which concept-generators are salient relative to an individual.

But I think this cost is actually very large, and that it provides a reason to prefer the fine-grained theory over this Millian one. In section \([6]\) I described how the permutations made salient relative to each person and world can be thought of as induced by a contextually supplied relation of equivalence among ways of thinking about individuals (i.e. elements of \( D_e \)) relative to each person and world. That sketch is just the beginning of a full story about how background features of conversational participants’ psychology and surroundings contribute to determining what permutations are salient relative to an individual at a world, but it is at least a beginning. By contrast it is unclear how the indexed-domain CG-theory can give even the beginning of such a story. This theory places special weight on the order in which names occur in the complement clause of an attitude report. For instance, suppose that Plato thought the planet he saw in the evening was brighter than the one he saw in the morning, and consider again the \( f \) described above as delivering an intuitive reading of \([5]\). Relative to this \( f \), the sentence “Plato believed Hesperus was brighter than Phosphorus” would have an intuitive true reading, roughly paraphrasable as “Plato thought the planet he saw in the evening was brighter than the one he saw in the morning”. This is a good result. But trouble sets in when we observe that relative to this \( f \) the sentence “Plato believed Phosphorus

\[49\] Though see n. \([19]\) for a way of making surrogate domain restrictions less flexible.
was brighter than Hesperus" would have the very same reading, and, more generally, that on
the indexed-domain CG-theory the first of these sentences will be true in exactly the contexts
where the second is. At this point, one might think that the indexed-domain CG-theory could
simply appeal to differences in the words used in the complement clauses of these reports to
explain why they are typically interpreted in one way rather than another, by analogy to the
strategy described in the introduction for explaining the contrast between 1 and 2. But there
are important differences between those examples and these ones. In explaining the contrast
between 1 and 2 the CG-theorist could appeal to the natural idea that there might be “seen
in the evening” contexts and “seen in the morning” contexts. But this idea does not yield
sufficiently fine-grained f to deliver an intuitive true reading of 5 to do that, we would need
the idea of a “first name is seen in the morning, and second name is seen in the evening”
context. But it is hard to understand what kind of context that would be. Moreover, there
is general concern: that the only natural ways of saying why a particular fine-grained f is
used for one sentence as opposed to another would be in effect to say that “Hesperus” has a
different compositional semantic value than “Phosphorus”, i.e. to endorse not a fine-grained
theory of the f supplied by context, but a fine-grained theory of the semantics of names.50

The above line of thought is my main reason for concluding that the fine-grained theory
should be preferred to the indexed-domain CG-theory as an account of 5. But I would feel
more confident in this conclusion if I had some data which clearly supported it. At present, I
don’t have robust, crystal clear examples which do this. But to illustrate what such evidence
might look like, I will present some very delicate examples which at least have the right
structure to discriminate between the two theories:

**Context** Amalia selects ten subjects who are known all to have genes which differ from
one another’s in a particular part of the genome. She runs two identical samples from
the relevant part of each subject’s genome through a sequencing machine. Amalia’s
technician has two pictures of each of the subjects. To make the data easier to analyze,

---

50 In Goodman and Lederman [forthcoming] §9.1 we develop related objections to a different Millian theory
which delivers a true reading of 5.

A different reason for preferring the present theory might be that it handles certain instances of Mates’
puzzle straightforwardly (Mates [1952]). To recall, Mates’ puzzle concerns the fact that data like 1 and 2 (and
even to some extent 5) can be produced with predicates which intuitively denote the same property:

27. Mary believes Woody is a woodchuck.
28. Mary believes Woody is a groundhog.

These data can be handled straightforwardly within my system, as follows. We may take “woodchuck” and
“groundhog” to denote the same property in the sense of denoting properties which have the same extension at
all possible worlds. But they can have different extensions at impossible worlds. By fine-graining the semantics
of names, we also fine-grain the denotations of predicates in such a way as to allow this way of accounting
for data related to Mates’s puzzle. More work would have to be done to see how general this account of
Mates’s puzzle is. But even handling the above basic data within the indexed-domain CG-theory is much more
complex; it requires an elaborate extension of the mechanics of concept-generators to properties. See Baron
2015.
he is supposed to staple one photo of each subject to each of the machine’s two printouts of the sequence for that subject. The technician attaches one of the photos of Issa (one of the subjects) to the correct printout, but he attaches the second photo of Issa to the wrong printout. Amalia works through the data using the photos as mnemonics for the people. When she comes across the pair of Issa’s photos, she points at the photos in order and says to herself “This person shares no relevant genes with that person, so even though they look similar in the photo, this person is not that one.” Later, the lab manager is explaining what happened to a friend, and says: “Because I switched the photos...”

27. Amalia thought Issa wasn’t Issa.

28. Amalia didn’t think Issa was Issa.

29. Amalia didn’t know that Issa was Issa.

30. Amalia thought Issa didn’t have any relevant genes in common with Issa.

31. Amalia didn’t think Issa had any relevant genes in common with Issa.

32. Amalia didn’t know that Issa had any relevant genes in common with Issa.

Judgments about these sentences are subtle. But below, I will report my own judgments about them, and document how those judgments would bear on our two theories. Nothing I say is meant to be conclusive.

To my ear, the most acceptable of these six sentences are 27 and 30, the two in which the negation takes narrow scope over only the complement clause of the attitude verb. The next most acceptable are 28 and 29, where the negation takes wide scope over the attitude verb, and the complement clause features the copula. The worst (and indeed flatly unacceptable) for me are 31 and 32, where the negation is wide-scope over the attitude verb and the complement clause of the report features an expression which denotes an uncontroversially reflexive relation.

If these are the facts about these sentences, then they provide some evidence in favor of my theory, and against the indexed-domain CG-theory. Both theories predict true readings of 27 and 30, and both theories can explain the contrast between 28 and 29 on the one hand and 31 and 32 on the other, given the hypothesis that the copula’s default use is not to express the reflexive relation of identity. But only my fine-grained theory gives a properly semantic explanation of the unacceptability of 31 and 32. My theory predicts that these sentences have no intuitive true readings, essentially for the same reason that the original CG-theory

51Same-name cases like these are typically associated with Kripke [1979]; this case is more similar to the “Thelma” case of Schiffer [1979], cf. Dorr [2014], Goodman and Lederman [forthcoming, §3].
predicted that 3-5 have no intuitive true readings. By contrast, the Indexed-Domain CG theory predicts that 31 and 32 are on a par with 3-5, and so in principle both sentences have true readings. Of course, the proponent of the indexed-domain CG-theory can supplement their theory with a pragmatic principle that explains why it is hard to access the true readings of 31 and 32 by comparison to 3 and 5. But if my judgments about these sentences are correct, the fact that the indexed-domain CG-theory requires this kind of supplementation would be some evidence against it.

Note finally that, although my theory predicts that 31 and 32 are false, it does allow true readings of more traditional “Paderewski”-style sentences (Kripke [1979]). For instance suppose that the printout to which Issa’s photo was incorrectly attached showed him as having Gene G, while the correct printout showed him as having Gene H (and not Gene G), and consider:

33. Amalia thought Issa had Gene G.
34. Amalia did not think Issa had Gene G.

Both the CG-theory and my theory will (correctly) predict that there is no context where both of these sentences are true together, but there are (different) contexts in which each of these receive intuitive true readings.

9 Conclusion

I have presented and argued for a theory which combines a fine-grained theory of the semantics of names with some key ideas of the CG-theory. Unlike the CG-theory, this theory straightforwardly allows intuitive true readings of 3-5—without postulating a structural ambiguity in those sentences. And unlike simpler fine-grained theories, it accommodates a range of complex examples, which the CG-theory handles smoothly. The model theory I’ve used to develop this theory includes impossible worlds, but only a highly constrained version of them, so that the models are comparable to ordinary possible-worlds models in terms of their predictive strength.

A main goal of the paper has been to show how the central examples (3-5; 16-19; 20-21; 22) impose independent structural constraints on any theory of attitude reports. For concreteness I have described these constraints within a general framework inspired by the CG-theory. This framework allowed me to describe the constraints as arising from the settings of specific parameters within the theory (Existentialist vs. Functionalist; Sensitive vs. Insensitive; Type-Variable vs. Type-Simple; Single-domain vs. Indexed-domain). But the examples impose constraints on a wide array of theories of attitude reports, and I hope the discussion here will spur further exploration of how they might be accommodated in such other frameworks as well.
References


A Concrete Versions of the Model Theory

In this appendix I present different ways of developing my formal models into a semantics in the philosopher's sense. I show how, starting from ingredients that should be acceptable to proponents of various theories of the meaning of names, one can construct a subclass of the models I have used in the main text, which is nevertheless sufficiently rich to do everything I’ve done in the paper. These models are not meant in any sense to be the final word on how proponents of these theories might adapt the work I’ve done in this paper; they are simply intended to witness my claim that the abstract models can be developed into a semantics.

A.1 Descriptivism / Predicativism

We assume that referential uses of names denote individual concepts, functions from worlds to individuals. (Since I am interested here only in referential uses of names, “the”-predicativism can be thought of, for my purposes, as a form of descriptivism.) Following Kripke [1972], we
assume that such uses of names are modally rigid: in our setting this means that if two worlds are possible relative to one another, the individual concept denoted by a name has the same value at each of them. In addition to a set of individuals \( X \), we take as given a set of worlds \( W \), which includes impossible worlds, and is equipped with a relation \( R \) of relative possibility. The impossible worlds in this set could be thought of as reflecting the epistemic possibility that the (rigid) description is satisfied by an individual different from the individual who actually satisfies it. We then let \( D_e \) be the set \( \{ f \in X^W | \text{ if } wRw' \text{ then } f(w) = f(w') \} \), and the relation \( E_w \) be \( \{ (f, f') | f(w) = f'(w) \} \).

We could now treat quantifiers exactly as I did in section 4. Alternatively, we could follow Aloni [2005] and replace my surrogate domain restrictions with conceptual covers, where a set of individual concepts is a conceptual cover of \( X \) in \( W \) if and only if for every \( x \in X \), and every \( w \in W \), there is exactly one \( f \in C \) such that \( f(w) = x \). Such conceptual covers correspond to a proper subset of the surrogate domain restrictions available in the models described in the previous paragraph, namely those which are (i) constant on their world argument, and (ii) such that, for every world \( w \), and every element \( A \in I_w \), there is exactly one \( a \in S(w) \) such that \( a \in A \). Since this Aloni-inspired version of my theory is more restrictive than Basic Surrogatism, and the arguments of section 5 concerned ways in which Basic Surrogatism itself was too restrictive, those arguments also show that the more restrictive Aloni-inspired version of the theory too, should be supplemented in some way.

Interestingly, using permutations (which are naturally taken in this setting to be functions from conceptual covers to conceptual covers) in the way described in section 6 would also solve these problems for the restrictive,

\[ n \text{In fact, the more restrictive features of Aloni’s proposal lead it to have some additional problematic predictions, beyond those of Basic Surrogatism. For (i), consider a scenario where John is looking at three pictures, } a, b, \text{ and } c, \text{ knows that they are pictures of exactly two people, but doesn’t know which pictures represent distinct people. One might think there could be a true reading of “John knows there is an } x \text{ and a } y \text{ such that } x \neq y \text{ and } (x = a \text{ or } y = a) \text{ and } (x = b \text{ or } y = b) \text{ and } (x = c \text{ or } y = c)”. \text{ Basic Surrogatism delivers this verdict but Aloni’s system doesn’t. For (ii), note that relative to a single conceptual cover, Aloni’s theory cannot deliver a true reading of: “there are two people that John thinks are identical” (i.e. there are } x, y \text{ such that } x \neq y \text{ and John thinks } x = y); \text{ Basic Surrogatism can handle this sentence straightforwardly. This second issue is recognized by Aloni herself and discussed also in detail by Holliday and Perry [2014, p. 620f.].}

Interestingly the main issues I am aware of that might lead one to reject the restrictive features of Aloni’s conceptual covers (e.g. the issue discussed under heading (ii) above, and those mentioned by Holliday and Perry [2014]) are eliminated once we move to a version of the theory which includes permutations. Such a theory and the final theory of this paper would still be different in the mechanics of how they handle various examples, but in the end I don’t know of a clear way in which they differ on English data.

Holliday and Perry [2014] develop a theory using individual concepts which eliminates (for instance) Aloni’s “uniqueness assumption” on conceptual covers, in a way somewhat reminiscent of Basic Surrogatism. Holliday and Perry seek to provide a formal system in which one can re-describe the English data, and do not provide a systematic translation manual between English and their logical language (see §2.3 of their paper for some discussion). They allow that some English sentences should be regimented using quantifiers (in effect) over individual concepts, but hold that others (and presumably “at least two” in [11]) should be regimented using quantifiers which range only over objects. Holliday and Perry [2014, p. 619f.] Basic Surrogatism provides a new way of having some of the benefits of the ways in which they relax Aloni’s assumptions, while giving uniform lexical entries for English quantificational expressions.

43
Aloni-inspired version of the theory.\footnote{Aside from the differences between surrogate domain restrictions and conceptual covers, my models are more flexible than Aloni’s in other ways too: it is not always possible to view $I_{\emptyset}$, the set of equivalence classes of $D_e$ at the actual world, as a set $X$ of individuals in the Aloni-inspired version of the models. A straightforward way of seeing this is to consider a model where there is one equivalence class under $E_{\emptyset}$ at the actual world, but two equivalence classes under $E_{\emptyset}$ at some impossible world. This simple version of the difference would be erased if we consider variable-domain versions of Aloni’s models, but there are further differences between the variable-domain versions of each theory.}

These descriptivist ways of viewing the model theory “explain away” the different elements of $D_e$, but they do not do the same for the impossible worlds in $W$. This may not be a problem. Descriptivists and predicativists might be happy to take epistemically possible, but metaphysically impossible worlds as primitive: since it is plausibly not $a$ priori that a given individual satisfies a given description, it is natural to assume that there is a notion of epistemic possibility according to which it is possible that the planet actually called Hesperus was not Venus but something else. But it is worth noting that we can go further and explain away the impossible worlds as well, as I will now describe.

We begin with a primitive set of individuals $X$ and a nonempty set $S$ of possible worlds, and let $C$ be the set of conceptual covers of $X$ in $S$. We then define the ingredients of my models as follows:

- $W = S \times C$;
- $R = \{ \langle \langle s, c \rangle, \langle s', c' \rangle \rangle \in W \times W \mid c = c' \}$;
- $D_e = \{ f \in (X^S)^C \mid f(C) \in C \text{ for all } C \in C \}$;
- for all $\langle s, C \rangle \in W$, $E_{\langle s, C \rangle} = \{ \langle a, b \rangle \in D_e \times D_e \mid a(C) = b(C) \}$.

Here the conceptual covers can be thought of as ways of conceiving what individuals there are. If a person thinks that Hesperus is not Phosphorus, their doxastic possibilities include only those where these names correspond to different elements of the relevant covers.

### A.2 Variabilism, Interpretations, Mental Representations

I now turn to give concrete versions of the model theory based on notions that would be acceptable to proponents of variabilist theories of names. We again take as given a set $X$ of individuals and a set $S$ of possible worlds. We assume that names are treated like variables: they are associated with numerical indices and mapped to individuals in $X$ by an assignment function, which I’ll call $i$ here (for reasons that will become clear in a moment). Using $N$ for the set of indices, the set of such assignment functions is $X^N$. We then define the ingredients of my models as follows:

- $W$ is $S \times X^N$, the set of pairs of possible worlds and assignment functions;
• $R$ is $\{(s, i), (s', i') \mid i = i'\};$

• $D_e$ is $N$

• for all $(s, i) \in W$, $E_{(s, i)}$ is $\{(n, m) \in D_e \times D_e \mid i(n) = i(m)\}$. 

This variabilist model-theory, when combined with my overall theory, does not make the common variabilist assumption that attitude verbs introduce binding over assignment functions (Cumming [2008], Pickel [2015], Schoubye [forthcoming]). As is well-known, on a flat-footed development of that approach, the assignment functions for ordinary pronouns would also be bound by attitude verbs, so that sentences like “there’s someone John thinks is wonderful” do not receive the correct truth-conditions (the sentence would be equivalent to “there’s someone$_1$ such that John thinks he$_2$ is wonderful”). Cumming [2008] and Pickel [2015] offer different proposals to solve this problem. A quite different kind of variabilist proposal (which does not need to deal with this issue) is found in Ninan [2012].

In constructing these models I did not appeal to variabilists’ distinctive syntactic-semantic thesis about the way names interact with assignments function in context. If we take the set $N$ not to be the set of indices, but instead as the set of names, and take $i$ not as an assignment function but as an interpretation function which maps names to individuals, we would obtain a different theory, where the semantic values of names are taken to be the names themselves. (Perhaps this is most naturally taken to be part of a more general picture where the $DOX$ function takes individuals and worlds to sets of pairs of worlds and whole interpretations of the language.)

A limitation of this proposal is that it is not always clear that elements of $D_e$ need to correspond to any public language name. In the set up for [16][19] for instance, it is not required that each of the teachers have multiple public language names associated with them, but the model theory employs multiple elements of $D_e$ corresponding to each individual. (This problem doesn’t clearly apply to the variabilist theory, since it is less obviously problematic to require that the actual assignment function assign multiple indices to each teacher, though it may be difficult to provide independent motivation for this requirement.)

A third way of viewing these models can be seen as motivated by this problem with the theory which takes the semantic values of names to be themselves. We take the set $I = D_e$ to be the set, not of public language names, but of mental representations of individuals. These might be understood as “names” in a “language of thought” (Fodor [1975]), or “mental files” associated with individuals, or in any number of other ways. We then take the $i$ above to be a map from mental representations of individuals to individuals.

A great deal more would have to be said to make this model plausible. Here is one obvious issue. If public-language names are to be assigned elements of $D_e$ then the mental representations should be “had” by every individual. On many theories of mental representations, however, different individuals never possess the same mental representations. To use
this class of models, proponents of such theories might take as given a primitive (perhaps interest-dependent) notion of synonymy among different agents’ mental representations, and think of the elements of $D_e$ as equivalence classes of mental representations which are treated as synonymous relative to the speakers’ interests.

### B In Scope Existentialism

This appendix is dedicated to an intriguing piece of data from [Charlow and Sharvit 2014] that the official theory does not account for. Charlow and Sharvit present the following examples, to be assessed in the context of 16–19:

35. John believes every teacher is Italian.
36. ? John believes no teacher is French
37. John believes both teachers are Italian.
38. ? John believes neither teacher is French.

They claim that 35 and 37 are naturally interpreted as true, while 36 and 38 are naturally interpreted as false. My consultants have almost universally disagreed with these judgments. Some report something of a contrast between 37 and 38, but almost none reported that contrast for 35 and 36. Further work needs to be done to see whether the contrast is genuine. But in case further work does bear it out, I will show in this appendix how the official theory could be extended to predict the contrast.

To predict the contrast, I propose altering the lexical entries for non-upward monotone quantifiers, for instance:

**Surrogatist No**

\[
\text{Surrogatist No} = \lambda w.\lambda F.\lambda G. \text{no } x \in S_w \text{ is such that}
\]

- for some $n \geq 0$ there are $\pi_1 \ldots \pi_n \in \bigcup_{x \in D_e} f(x, w)$ such that $F(\pi_1) \ldots (\pi_n)(x) = 1$, and
- for some $m \geq 0$ there are $\pi_1 \ldots \pi_m \in \bigcup_{x \in D_e} f(x, w)$ such that $G(\pi_1) \ldots (\pi_m)(x) = 1$.

54 Cable [2018] develops an account designed to predict the contrast between these in-scope data. His account of this contrast is very elegant. Unfortunately, the proposal has several undesirable features. Here are two. First, it predicts that in any context in which “Plato thought that Hesperus rose in the evening and Phosphorus rose in the morning” is true, “Plato knew that Hesperus rose in the evening and Plato did not know that Hesperus rose in the evening.” will also be true. (Note the main clause negation in the second conjunct here.) Second, Cable’s theory cannot accommodate the data discussed in section 5.2.

55 In the Millian setting of the CG-theory, if the individual concepts which are in the range of concept generators can produce different individuals at different metaphysically possible worlds, an analogous clause for negative quantifiers would produce undesirable results for sentences like “no planets are necessarily identical with Phosphorus”, where “no” governs a binder which binds into an alethic modals like “necessarily”. I see no obstacle to using an analogous entry with the CG-theory (or variants of it) provided the theory uses metaphysically impossible worlds and the relevant individual concepts are assumed only to vary between worlds which are impossible relative to one another.
Here “$F(\pi_1) \ldots (\pi_0)$” abbreviates “$F$”, and similarly for $G(\pi_1) \ldots (\pi_0)$; these abbreviations allow us to accommodate in a single clause both the degenerate case where there are no permutations, and the interesting case where there are one or more. The key idea is that the determiner now introduces a sequence of existential quantifiers over salient permutations. Whereas the arguments of attitude verbs were functions from permutations...to permutations to propositions, here the arguments are functions from permutations....to permutations to extensions. When working with attitude verbs we assumed that different permutations were salient for different individuals. Since determiners do not have individuals as arguments, we must quantify over a set of permutations which are salient simpliciter, and not (as before) salient relative to some individual. There are various ways one might operationalize this idea, but in the entry above the permutations which are salient simpliciter are defined as those such that there’s some individual relative to whom they are salient at that world.

This entry is designed to work with a syntax like the following:

![Syntax Diagram]

Here the binder $\lambda_{\pi_5}$ is produced beneath “no teacher” and the binder $\lambda_{\pi_4}$ is produced below “no”. We assume that even with simple predicates like “is a teacher” there is a $\lambda$-abstract produced by movement.

The key result is that provided the syntax above, where the quantifier over permutations has scope below “no”, we naturally predict that 36 is interpreted as false. The transparent reading of that sentence relative to an $S$ and $f$ (and assuming natural indexing of world-pronouns not displayed above) will now express:

- $\lambda w. \forall w' \in DOX(John)(w), \text{ no } x \in S_w \text{ is such that for some } \pi \in \bigcup_{x \in D_e} f(x, w) \pi(x) \text{ is a teacher at } w, \text{ and for some } w'-\text{admissible } \pi \pi(x) \text{ is French at } w'$.

Given this entry, assuming that all $w$-permutations are salient, 35 would be true, and 36 false.\(^{56}\)

\(^{56}\)Charlow and Sharvit suggest that either attitude verbs are ambiguous between a “universal” and an
I've assumed that the only altered entries are for non-upward monotone determiners. This is all that is needed to handle Charlow and Sharvit’s judgments. But it is an interesting question whether we should also alter the entries for other determiners. One way to motivate such an extension comes from sentences like the following:

39. John thinks at least one planet which shares matter with Jupiter also shares matter with Mars.

I will consider only the reading of this sentence on which “at least one” receives an opaque interpretation, i.e. the traditional de dicto interpretation. The sentence should intuitively be false on this reading.

But both my theory and the CG-theory say that, on this reading, the sentence can be regimented syntactically in at least two different ways, on one of which the sentence can be true. On a first regimentation, all permutation-variables in the sentence are coindexed, and we get the correct prediction that the sentence is false regardless of what context it is interpreted in. But on a second, where the permutation variables associated with the traces in the relative clause (“which shares matter with Jupiter”) differ from those associated with traces in main clause of the prejacent (“also shares matter with Mars”), we predict that the sentence has true readings (in contexts where 22 is true). This second true reading would be eliminated if the permutation-variables were bound beneath “at least one”. For then, on the opaque reading of this determiner, the permutations would be required to be admissible not at world argument of “thinks” (i.e. the worlds of the speakers’ context), but instead at the world-argument of “at least one”. Since on the opaque reading, the world argument of “at least one” ranges over John’s belief-worlds, even if the permutation variables are not coindexed, they still must respect the identity relation as it is at John’s belief-worlds, so the sentence will be guaranteed to be false. Probably the best version of this theory would require that binders over permutations must occur as low in the tree as is consistent with the syntax being interpretable, to guarantee that the permutations are governed by the determiner.

"existential" entry (where the universal has a universal quantifier over permutations), or that attitude verbs always have the universal entry, but that the only non-singleton restrictions arise when non-upward-monotone quantifiers are embedded in the attitude report. The second of these options cannot predict the contrast between 16 and 17 since no non-upward monotone quantifiers occur embedded in attitude reports in these sentences. The first of these theories is supplemented by the idea that the “universal” entry tends to be preferred when the attitude report features a non-upward monotone quantifier. If this idea is understood in such a way that 19 is supposed to suggest a universal interpretation of the attitude verb, the theory would be unable to handle that sentence, since under the universal interpretation, it would be true that there is no one that John thinks is French, and hence true that no one John thinks is French, is French. Charlow and Sharvit might therefore hold that the universal interpretation is suggested only when the non-upward-monotone quantifier occurs inside the scope of the attitude verb. But the proposal was already somewhat ad hoc and the fact that this restriction is needed makes it even more so.

The present theory also explains some further data which puzzle Charlow and Sharvit and which pose problems for their own theory (see Charlow and Sharvit [2014, pg. 37 example 72]).

Richard [1983] and Soames [1985] take related data to be so clear that they use them as the basis for objections to particular theories of direct reference, and propositions respectively. See Salmon [1986, p. 401-405] for helpful discussion.
when it occurs within the scope of the attitude verb.\textsuperscript{58}

I’m inclined to say that anyone who accepts Charlow and Sharvit’s judgments should accept this argument. But I should note that there is a possible counter-argument, which might motivate one to reject the idea that permutations can scope under quantifiers like “some”:

**Context** John thinks Hesperus and Phosphorus are distinct planets, and says to himself “exactly one of Hesperus and Phosphorus rises in the evening, and no other planets do”.

40. John thinks some interior planet rises in the evening,

Some of my consultants say that this sentence does not have a true reading. (I think I can access one.) If the quantifier over permutations were only available just below the verb (but not under the quantifier) we would make the prediction that there is no true reading of the sentence. But if it is produced below “some” in the sentence, then the existential quantification over permutations scopes under the universal quantification over worlds, so we can choose different permutations for different worlds, allowing a true reading, when the quantifier receives a transparent interpretation. This example would not be enough to move me to reject the proposal in this appendix, if I were convinced it was well-motivated in the first place. But I’m also not moved by Charlow and Sharvit’s original examples, and I don’t think the judgments about 39 are strong enough on their own to motivate the altered syntax and lexical entries for determiners, so I’m not convinced the proposal is well-motivated.\textsuperscript{59}

\textsuperscript{58}A different way of ruling out the problematic ruling would be to stipulate that every permutation variable under a determiner which receives an opaque interpretation must be coindexed. But such a stipulation would have to be motivated in a more general way.

\textsuperscript{59}There is no formal reason why, in the lexical entries for attitude verbs in the main text, we could not have stipulated that the quantifier over permutations scopes under the universal quantification over worlds. The problem is that this theory (as Charlow and Sharvit note) makes questionable predictions. Consider:

**Context** Mary has two pictures of Martin. She thinks they are different people. She says: “exactly one of these people is French”, pointing to the two photos.

- Mary thinks Martin is French.

Charlow and Sharvit think this is intuitively false. My consultants go different ways on it. But I’ve preferred the theory which rejects it, since such a theory is stronger, and even if accepted the data are middling.