

Fine-grained semantics for attitude reports

Harvey Lederman*

March 3, 2021

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 themselves are of interest independently of my 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, concept-generators

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

*Thanks to audiences at Philosophical Linguistics and Linguistical Philosophy (PhLiP) 5, the Princeton Talks in Epistemology and Metaphysics, and the NYU Semantics Group 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 great deal about these issues from my joint work with Jeremy Goodman, and the paper is heavily indebted to the many conversations we have had about them.

This is a preprint version of a paper published by *Semantics & Pragmatics*, which includes two appendices and a number of footnotes that had to be cut from the published paper. Aside from the appendices, the principal differences between this version and the published one are: the inclusion of the final paragraph of section 1, a different version of the final paragraph of section 3, and the inclusion of notes 9, 19, 25, and 29. Please cite the published version, available at <https://doi.org/10.3765/sp.14.1>, where possible.

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

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.¹

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, this idea is more of a wish-list than a theory. How should we think about these 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 $\lceil \text{Plato believed } \delta \text{ is visible in the evening} \rceil$ 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, 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.

¹Some take a passage in *Laws* 821c, where the character Kleinias describes the paths of “Hesperus and Phosphorus and other stars”, to be evidence that the historical Plato did not know that the planet called “Hesperus” was the planet called “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] predicts that none of them has an intuitive true reading.

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 drives 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.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 this basic theory cannot handle, but which the theory of Percus and Sauerland [2003] handles smoothly. These examples motivate a new 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 new theory accounts for the examples. In that section, I discuss systematically how my examples each impose independent constraints on the shape of my theory, and how they go beyond examples in the literature (in particular those discussed in Anand [2006]) designed to motivate particular features of the concept-generator theory (most notably, its use of existential quantification over concept-generators).

With the new theory before us, we face an important question: should we prefer this new 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

²Given my definitions, a theory of the semantics of names is fine-grained if and only if it is not Millian. Many theories of names are naturally seen as fine-grained according to these definitions. “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. Non-descriptivist variants on the ideas of Frege [1892] may also be fine-grained theories, 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). And it is natural to see “variabilist” theories (Dever [1998, §2.3], Cumming [2008], Pickel [2015], Schoubye [forthcoming]) as falling in this class as well, 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.

be preferred over this alternative.³

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*), which was first published in Percus and Sauerland [2003] (building on notes of Irene Heim), and argue that it fails to predict relevant true readings of 3-5.⁴

The CG-theory aims to predict that, for instance, “Plato believes Hesperus is bright” is true in a context (roughly) if and only if there is a contextually salient definite description δ such that ‘Plato believes δ is bright’ is true in that context. A main goal of the theory is to predict these truth-conditions without requiring that the name “Hesperus” (implausibly) undergoes syntactic movement out of the clausal complement of “believe”. The formal background for the theory is that of standard possible-worlds semantics for attitude verbs (Hintikka [1962]): we take as given a non-empty set of worlds W , a set of individuals X , and a function $DOX : X \rightarrow (W \nrightarrow \mathcal{P}(W))$ (\nrightarrow indicating that the function may be partial) 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 goes beyond this standard framework in its use of *concept-generators*, functions from individuals to individual concepts, where an *individual concept* is in turn a function from worlds to individuals. In particular, we assume that, when a name or pronoun occurs within the scope of an attitude verb, a covert pronoun which denotes a concept generator takes the name or pronoun as its argument; the result of composing the concept-generator with its argument (which denotes an individual) will be an individual concept (the type of the denotation of definite descriptions). To produce existential quantification over such definite descriptions (as in the target truth-conditions), we assume that concept-generator variables themselves are bound below some relevant attitude verb, and that the attitude verb introduces existential quantification over the concept-generators. By varying which concept-generator operates on a given name or pronoun, we indirectly vary which individual concept (denotation of a definite description) is associated with the indi-

³I will assume throughout the paper that any satisfactory theory must accommodate intuitive true readings of 3-5, and related sentences. But I am in fact open to the idea that this assumption is false, and that the best overall theory may predict that these sentences do not have true readings at all (see Goodman and Lederman [forthcoming, §11]). The paper can thus be read as exploring what follows from this assumption, while leaving it open that broader theoretical considerations could lead us ultimately to reject it.

⁴While 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].

vidual denoted by the name or pronoun. We can thus produce the desired truth-conditions without requiring undesirable syntactic movement.

The assumption that covert concept-generator variables can be bound beneath attitude verbs forces a modification to the usual lexical entries for attitude verbs themselves. A verb like “believe” will no longer always have a function from worlds to truth-values as its argument. If there is one concept-generator pronouns bound beneath the verb, its argument will denote instead a function from concept-generators to functions from worlds to truth-values. If there are more concept-generator pronouns bound there, then the argument will be more complicated still. To handle this variation in the type of the argument of “believe”, we use the following lexical entry instead of a more standard one:

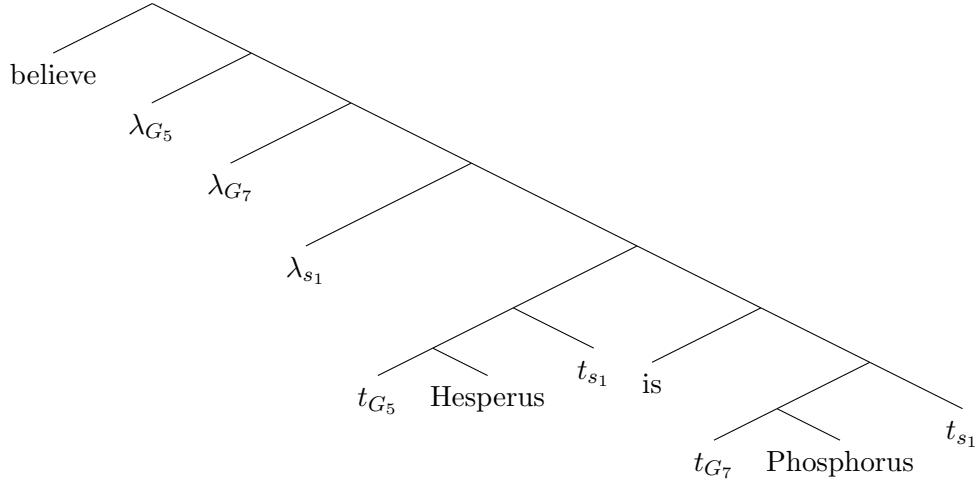
CG-Believes $\llbracket \text{believes} \rrbracket^{g,f} = \lambda p. \lambda x. \lambda w. \text{ either for all } w' \in DOX(x)(w), p(w') = 1, \text{ or for some } n \geq 1, \text{ there are } G_1, \dots, G_n \in f(x) \text{ which are salient for } x \text{ and such that for all } w' \in DOX(x)(w),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

We assume that context determines a function f , which maps each individual x to a set of concept-generators, intuitively, those which are contextually salient relative to the individual x . In this entry, 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 of the lexical entry (“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 mentioned above, where p may be a function from concept-generators to functions from concept-generators...to functions from worlds to truth-values. The entry in effect introduces a sequence of existential quantifiers over concept-generators, of the exact length needed to saturate the first argument of “believes”, so that it yields a function from worlds to truth-values.

So far, so good. But, as I will now show in detail, the theory cannot produce a reasonable interpretation of 5. Given the syntactic assumptions sketched above, the theory predicts that the following is the natural syntax for the VP of 5, at an appropriate level of abstraction:⁶

⁵Strictly 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.

⁶Throughout 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 propositions (functions from worlds to truth-values) when required (see e.g. Percus [2000]). I will use the simplest, highly unconstrained, version of this theory, and in working examples, will simply cherry-pick my preferred syntax from the huge array of available ones. I note, though, that everything I do below is compatible with the more constrained (and to my mind preferable) system of Schwarz [2012], where the only constituents which take world-pronouns are determiners. My basic theory could also be developed using quite different approaches to the “*de re*”/“*de dicto*” or “transparent”/“opaque” ambiguity, for instance, a “split intensionality” theory ([Keshet, 2008, 2010, 2011]).



Given the lexical entry for “believe”, the denotation of the VP of 5 with the above syntax will be:

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 $\text{Hesperus} = \text{Phosphorus}$, 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 thus for the sentence as a whole: 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 allow for the intuitive true readings that these sentences seem to have: readings on which they describe Plato’s specific ignorance or lack of opinion about a particular astronomical fact.⁷

⁷Allowing 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 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 it 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 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.⁸ 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”. More importantly (since such semantic phenomenology is not probative), 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

against the right background, for instance, “Plato did not believe that Hesperus was Phosphorus, Phosphorus was Venus, or Hesperus was Venus”. Thanks to Josh Dever here.

Note that for the kind of sentences I’ve considered in this section, Santorio [2014] is simply a different implementation of the same truth-conditions as the CG-theory, and so is subject to the same problem. Like the CG-theory, Santorio’s theory can be modified along the lines I describe in section 8 to avoid the problem.

⁸Indeed, in correspondence Percus and Sauerland have said that their theory should be supplemented in this way; see also Sauerland [2015, p. 77].

evening did not rise in the evening”). But there is no similar felt change of perspective between readings of 5, and, crucially, referential uses of names 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 *de re* and *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. The theory I develop below will not require 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.⁹ In section 8 I will consider the prospects for a Millian theory which escapes this argument.¹⁰ But first – and for most of the paper – I will explore a different response, which sees the argument as casting doubt on the underlying Millianism of the CG-theory, and thus takes it to motivate developing a fine-grained, non-Millian alternative.

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 my 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;

⁹The 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.

¹⁰I 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. 46. Others are Crimmins and Perry [1989], Crimmins [1992] and the theories described in Goodman and Lederman [forthcoming, §9.1 and §9.2].

- D_e , a set;
- $DOX : D_e \rightarrow (W \nrightarrow \mathcal{P}(W))$, 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”;
- $E : W \rightarrow \mathcal{P}(D_e \times D_e)$ 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 $\mathbf{2}$ for the set of truth values $\{0, 1\}$ and D_p for the set of functions from worlds to truth values, i.e. $\mathbf{2}^W$. I sometimes call these “propositions”.¹¹

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 also 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 $E_{@}$ (where “@” here and throughout stands for the actual world). 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 $D_e = \{h, p, pl\}$, $W = \{@, i\}$, R is the identity relation on W , $E_{@}$ 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):

- $\llbracket \text{Hesperus} \rrbracket^g = \lambda w.h,$
- $\llbracket \text{Phosphorus} \rrbracket^g = \lambda w.p,$

¹¹For 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, though neither of these presents any real challenge, as far as I can see.

- $\llbracket \text{Plato} \rrbracket^g = \lambda w. pl$
- $\llbracket \text{is} \rrbracket^g = \lambda w. \lambda x. \lambda y. x E_w y;$
- $\llbracket \text{it's not the case that} \rrbracket^g = \lambda w. \lambda x. \mathbf{2}. 1 - x..$

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

We can now give a straightforward treatment of 5. The set $DOX(pl)(@) = \{i\}$, and it is not the case that hE_ip . 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 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. I will sometimes speak of it as a “semantics”. By this I mean that it is a formal model used to make predictions both about the truth and falsity of sentences in context and about entailment relations among sentences (Yalcin [2018]). I do not mean that the model gives a “semantics” in some heavier-weight philosophers’ sense of that term. It is just a model, to be judged by its simplicity, tractability and predictive strength. In all three of these dimension, my models are comparable to possible-worlds models. Most importantly, just as in standard possible-worlds models, at every world in every model I consider, Boolean connectives such as “it’s not the case that” will behave standardly. As a result propositions themselves will form a Boolean algebra under the usual set-theoretic operations.¹³ The only non-standard feature of the models will be that identity is interpreted by a non-trivial equivalence relation on D_e , 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.

¹²Again, 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.

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 $x \in D_e$ is 1 if and only if for every y such that $x E_w y$, 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 .

Although I do not aim to give a semantics in the philosophers' heavier-weight sense here (Burgess [2008]), one might wonder whether the models I develop could be used as the basis for one. In appendix A, I give a first stab at answering this question, by showing how a wide array of fine-grained theories of names can develop what they might take to be a semantics in this sense using models that are isomorphic to subclasses of mine. As I show, 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.¹⁴ Although I will be officially neutral on how to think about elements of D_e for most of 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 at some points it may bear a little 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 one might reconstrue my impossible worlds in other terms.

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 of D_e , the semantic value of "Hesperus", and the semantic value of "Phosphorus", which satisfy the predicate "is visible in the evening".¹⁵

¹⁴A 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.

¹⁵This problem with 11 arises for fine-grained theories like mine on which the semantic values of intuitively

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.¹⁶

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

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

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:

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

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, 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

extensional predicates like “is an interior planet” are functions from worlds to functions from *the domain of the semantic values of names* to truth-values. (It would also arise for theories on which such predicates denoted functions from the semantic values of names to propositions; the key point is that, no matter where they occur, intuitively extensional predicates denote functions which in some natural sense operate on the domain of the semantic values of names.) An alternative style of fine-grained theory takes occurrences of intuitively extensional predicates which are not in the scope of attitude verbs to denote functions not on the semantic values of names, but on equivalence-classes of them. The most natural versions of such theories do not have any trouble with 11, but they face a related problem with sentences which involve binding into the scope of attitude verbs, like “There are at least two interior planets which Plato thinks are visible in the evening”. To handle such examples, these theories typically employ a non-standard rule for predicate abstraction (Bigelow [1978], Yalcin [2015], Lederman [2020]). I am inclined to see such a change to the rule for abstraction as more disruptive than the domain restrictions I will impose below to handle 11. But I will not give a systematic comparison between the two approaches: I have simply wanted to observe that while 11 poses a problem for the style of fine-grained theory I will be developing here, it does not pose a problem for all fine-grained theories.

¹⁶To my knowledge, Kaplan [1986, p. 258–9] first gave the name “surrogatism” to a related proposal (see Section XVI for development of the view). Aloni [2005] cf. Ninan [2018]) Dorr [2014] and Bacon and Russell [2017] can also be thought of as “surrogatists” in Kaplan’s sense, though the parallel is not exact in each case.

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.¹⁷

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. 45 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.¹⁸

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*. This proposal gives an account of a broad array of data without using anything resembling concept-generators. In the next section I'll present three problems for this theory

¹⁷Here 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.

¹⁸Throughout 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 the restriction used in every context. For instance, descriptivists who see elements of D_e as individual concepts might take the surrogate domain restriction in every context to be the set of constant functions which return the same individual at every world where the individual exists. For details on a descriptivist way of understanding the proposal see appendix A.1.

and go on to propose a refinement of it.¹⁹

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

¹⁹There are two salient alternatives to Surrogatism; in this note I'll say briefly why I disprefer them. Alternative “Existentialist” and “Universalist” proposals are as follows:

Existentialist Two

$\llbracket \text{at least two} \rrbracket^g = \lambda w. \lambda F. \lambda G. \lambda w' \text{ for at least two } Z \in I_w, \exists x \in Z \text{ such that } F(x)(w) = 1, \text{ and } \exists y \in Z, \text{ such that } G(y)(w') = 1.$

Universalist Two $\llbracket \text{at least two} \rrbracket^g = \lambda w. \lambda F. \lambda G. \lambda w' \text{ at least two } Z \in I_w \text{ are such that } \forall x \in Z F(w)(x) = 1 \text{ and } \forall x \in Z G(w')(x) = 1.$

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.

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 Caie et al. [2020] for treating “every” and “some” as unrestricted even in a fine-grained setting.

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 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.²⁰

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

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

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

²⁰Barker [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.

assumption one and only one of these classes has elements which are bright at w' for all $w' \in DOX(\text{Plato})(@)$ (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.²¹

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 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.²²

²¹It 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.

²²The 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 3-5), 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 5 and 10, 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 are used referentially in the complement clause of an attitude report (e.g. “Mary didn’t know he was him”). The system also handles cases with coreferential occurrences of demonstratives (“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 $E_@$ (for the example, see Perry [1977, p. 12-13]).

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.

5 Three Problems for Basic Surrogatism

In this section I present three problems for Basic Surrogatism, which the CG-theory avoids. In the next section I 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 first problem for Basic Surrogatism comes from the following example:

Context John has four pictures in front of him, two pictures each of two teachers. The teachers are Anna and Beau; we think think of the photos of Anna as A_1 and A_2 , and the photos of Beau 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 Anna is Italian and Beau is French.²³

14. Someone John thinks is French is French.
15. ?Everyone John thinks is Italian is Italian.
16. Someone John thinks is Italian is French.
17. ?No one John thinks is French is French.

The sentences 14 and 16 are naturally heard as true, whereas the sentences 15 and 17 are naturally heard as false. (They all have true, and false, readings in this scenario; my claim is just that there is a contrast in immediate acceptability between these pairs.) But on the natural assumption that the only relevant ways of thinking about individuals (i.e. elements of D_e) correspond to the four pictures of the teachers, Basic Surrogatism predicts that 14 is true in a context if and only if 15 is true in that context, and that 16 is true in a context if and only if 17 is true in that context. Moreover, it predicts that 14 is true in a context if and only if 17 is false in that context.

So 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 3-5.

²³This style of “pictures” case was introduced by Charlow and Sharvit [2014]. The example sentences used here are new.

When I return to discuss this example in more detail, in section 7.1, I will argue that it motivates the use of existential quantification over concept-generators within the CG-theory (and an analogous feature of my own theory). There, I will discuss in detail how the argument based on this example complements and goes beyond some previous arguments for this feature of the theory (in particular, one based on the “double vision” scenario of Quine [1956] and those developed by Anand [2006]).

5.2 Problems with plural subjects

The following example, due to Cian Dorr, presents a different kind of problem for Basic Surrogatism:²⁴

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

18. On Monday at noon, Eve and Dawn learned that Venus is not a star.
19. 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 elements 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 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 $\lambda w.p$,

²⁴Jeremy Goodman and I discussed a related example in earlier versions of Goodman and Lederman [forthcoming].

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 14-17.²⁵

5.3 The bound *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]:

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.

20. There’s something John thinks is Jupiter and is Mars.
21. 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 21.

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.

²⁵A 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.

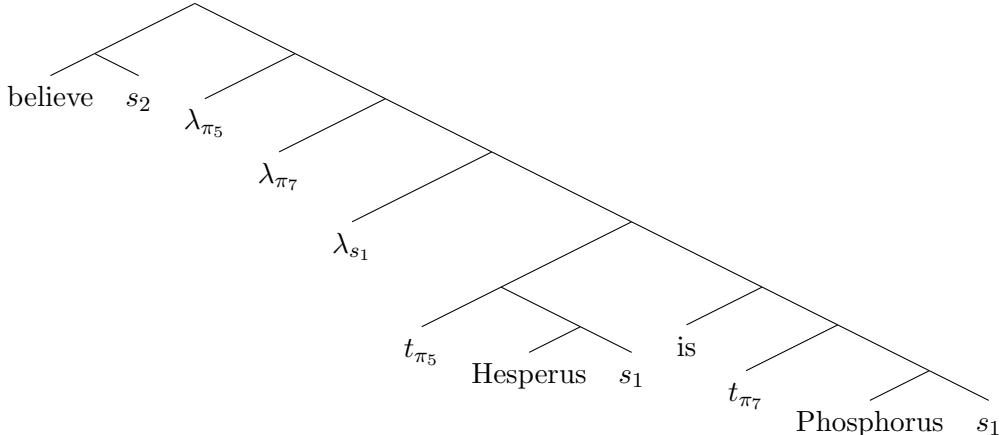
As far as I can see this proposal would handle the present example, but it does not give us a way of handling examples like 14-17: those examples seem to require some kind of existential quantification over ways of thinking. My way of handling these examples makes it natural to handle Dorr’s data by allowing the domain of permutations to vary from attitude-holder to attitude-holder, rendering this kind of type-shifting unnecessary.

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 with ideas from the CG-theory to produce a theory which handles both sets of data.

A bijection $\pi : D_e \rightarrow D_e$ is a *permutation*. A permutation π is *w-admissible* if and only if for all $x \pi(x)E_wx$; 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 @-permutations which map the semantic value of "Hesperus" to the semantic value of "Phosphorus". But there are no @-permutations that map the semantic value of "Hesperus" to the semantic value of "Mars".²⁶

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_{π_i} , I assume that the assignment function g is extended to be defined on new indices π_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:



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

²⁶There 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).

see this f as induced by contextually supplied equivalence relations among ways of thinking about individuals, which are defined relative to each person and 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 do care about whether Plato thinks about Venus in one way as opposed to another, and they do care 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 such an equivalence relation among ways of thinking about things for each person and 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 person and world. If we take this set of permutations as the value of f relative to that person and world, then the first context above, where differences between “Hesperus” and “Phosphorus” are unimportant relative to Plato and the actual world, will be associated with an f such that $f(\text{Plato}, @)$ contains a permutation which maps the denotation of one to the other (and one which maps the denotation of the other to the one). By contrast, the second contexts, 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.²⁷

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

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

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

²⁷The informal discussion using “contextual equivalence” makes it natural to impose further constraints on the values of $f(x, w)$ for every 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 π 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.

case, where p is a function from permutations to functions from permutations...to functions from worlds to truth-values.²⁸

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

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

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 $w \in DOX(x)(\llbracket s_2 \rrbracket^{g,S,f}), hE_w p$,

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.

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. 27). If we make this assumption, then the choice of semantic values for names *does* matter.²⁹

²⁸The 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.

²⁹One 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 π be the type of permutations. Then as usual we have base types e, p for names and sentences respectively, and in addition π for permutations. The simple variable type is χ . The types are then the members of the smallest set containing e, p, π, χ , and such that if σ and τ are members of the set, so is $\sigma \rightarrow \tau$. A simple abnormal type is anything of the form $\pi_1 \rightarrow (\dots(\pi_n \rightarrow t))$. Expressions of type $\chi \rightarrow \tau$ combine with an expression of a simple abnormal type to produce an expression of type τ . For instance, the type of attitude verbs (again, ignoring the world-pronouns we had above) is $\chi \rightarrow (e \rightarrow p)$. 29 can then be regimented as $\forall^\chi(\lambda P^\chi. \text{if Mary believes } P^\chi \text{ then John believes } P^\chi)$. We can assume a set W of worlds and let D_p be $\mathcal{P}(W)$, D_e be an arbitrary set, and Π 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.

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.³⁰

³⁰One 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 for all w' such that wRw' , $p(w') = p'(w')$, and consider:

Propositional Believe $\llbracket \text{believe} \rrbracket^{g,S,f} = \lambda w. \lambda p. \lambda x. \text{for some } p' \in f(x)(w)(p), \forall w' \in DOX(x)(w), p'(w') = 1$, where $f(x)(w)(p)$ is assumed to contain only *w*-variants of p ; intuitively, *w*-variants of p which are salient relative to x . This entry is essentially the lexical entry of Richard [1990], transposed to the present unstructured setting. It can accommodate all of the data we have considered to this point, and has the significant advantage of not requiring a complex syntax with permutation variables, allowing the arguments of attitude verbs to be propositions.

The idea behind this simpler theory is attractive. But as it stands it is too unconstrained. 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.

There is however a further, natural constraint which would eliminate this prediction, building on the notion of "intensional isomorphisms" from Carnap [1947]. We require not just that the elements of $f(x)(w)(p)$ be *w*-variants of p , but also that they be expressed by a (salient) sentence s which is *w-intensionally epimorphic* in context to the complement clause s' . (Since this will be a pragmatic constraint on which propositions are salient in context, imposing it does not require that any expression have the quotation-name of itself or another expression as part of its semantic value.) Say that w' and w'' are *w-intensionally equivalent* if and only if wRw' and $w' = w''$, that $x, y \in D_e$ are *w-intensionally equivalent* if and only if $xE_w y$, that $n, m \in \mathbf{2}$ are *w-intensionally equivalent* if and only if $n = m$, and, finally, that $f, f' \in D_r^{D_\sigma}$ are *w-intensionally equivalent* if and only if for every *w-intensionally equivalent* a, b , $f(a)$ is intensionally equivalent to $f(b)$. Then a sentence (or more properly: a syntactic parse of a sentence) s is *w-intensionally epimorphic* to a sentence s' in a context if and only if there is a surjection j from (not necessarily terminal) nodes of s to the terminal nodes of s' such that (i) if $j(\alpha)$ dominates $j(\beta)$ in s' then α dominates β in s and such that (ii) for all α in the domain of j , the interpretation of α in this context is *w-intensionally equivalent* to the interpretation of $j(\alpha)$ in this context. This proposal strikes me as in some ways more attractive than the one in the main text, but I have focused on that one because it allows an easier comparison to the CG-theory.

A quite different 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.

7 Solving the problems

In this section, I'll describe how the new fine-grained theory solves the three problems I described for Basic Surrogatism (sections 7.1-7.3). In each subsection I will also discuss in more detail how my examples constrain the theory I've developed, as well as the CG-theory itself. Readers primarily interested in the positive proposal of the paper may wish to skim or even skip this section; a good deal of it is taken up with discussion of how alternative theories fail to handle the three examples.

7.1 Beyond double vision

I suggested that on their most salient readings, 14 and 16 are true, while 15 and 17 are false. I'll now show how my theory accounts for this contrast.

In spelling out these predictions, I'll call the equivalence classes corresponding to each teacher **A** and **B**, and 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 one 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 14 expresses:

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

But under the same assumptions, we predict that 15 will be false. Relative to any S and f , our lexical entry for “believe” predicts that on the most natural syntax 15 expresses:

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

The reader may readily verify that 16 will similarly be predicted to be true, and 17 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

$$\llbracket \text{believe} \rrbracket^{g,S,f} = \lambda p. \lambda x. \lambda w. \forall w' \in DOX(x)(w'), p(f(x,w))(w') = 1.^{31}$$

In this entry I’ve assumed that the values of f are just permutations (not singleton sets of permutations) and I’ve left out complications required to deal with cases where different numbers of permutation variables are bound by the verb, since they won’t matter here.

Under the natural assumptions I made at the start of this section, a functionalist theory will predict that 14 is true in a context if and only if 15 is, and that 16 is true in a context if and only if 17 is. Given appropriate analogues of those natural assumptions, related theories which use concept-generators in a Millian setting instead of permutations in a fine-grained one (let f in the entry above supply a single concept-generator for each individual and world) make exactly this same bad prediction. (Anand [2006, p. 25] calls this the “Skolemized” proposal, but I will call “functionalist CG-theory”.) So functionalist theories of all stripes fail to predict the observed contrast in immediate acceptability between these pairs of sentences.³²

This new example complements and goes beyond some earlier arguments against functionalist theories. Perhaps the most famous such argument, often attributed to Quine [1956], starts from the following example:

³¹A 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.

³²One might think that the different words “French” and “Italian” in the complement clauses of the reports above on their own 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 14 and 17, and substitute “is not French” for the relevant occurrences of “is Italian” in 15 and 16, the modified examples lead to the same 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. Extreme versions of contextualism could escape these arguments by holding that context changes are cued by the minute differences between these examples (e.g. by the use of the word “not” in the complement clause), but insisting on such magical context changes hardly leads to an attractive or plausible theory.

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.

22. Ralph believes that Ortcutt is a spy.
23. Ralph believes that Ortcutt is not a spy.

Both of these sentences are intuitively true in this scenario. But 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, a functionalist fine-grained theory will predict that there is no context where both are true.³³ Similarly, since there is plausibly no relevant description δ which refers to Ortcutt in the actual world and such that $\lceil \text{Ralph believes that } \delta \text{ is a spy} \rceil$ and $\lceil \text{Ralph believes that } \delta \text{ is not a spy} \rceil$ are both true, the functionalist CG-theory also predicts that there is no context where both of these sentences are true.

This example has played a central role in the development of semantic theories of attitude reports. But the argument based on it is not particularly strong. For it relies on the claim that the two sentences must be true *in the same context*. And this premise can be denied, without giving up the far more important claim that both sentences are typically true when uttered. As Anand [2006, p. 24-5] notes (citing Zimmerman [1991] and Heim [1998]), proponents of the functionalist CG-theory (and, we might add, a fine-grained functionalist theory) may claim that different concept-generators (respectively, permutations) are salient in the different contexts in which these different sentences are typically assessed, and thus accommodate the judgment that both are true when uttered, even though there is no single context in which both are true.³⁴

After describing precisely this limitation of arguments based on Quine's example, Anand [2006, p. 32-33] develops two new arguments against the functionalist CG-theory (which apply

³³It 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 $\lambda w. o \text{ is-a-spy-at-}w$, and if one comes to believe that Ortcutt is not a spy by seeing him on TV, one believes the proposition $\lambda w. o \text{ is-not-a-spy-at-}w$. 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 $\lambda w. x \text{ is-a-spy-at-}w$ while also believing the proposition $\lambda w. x \text{ is-not-a-spy-at-}w$ 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.

³⁴Basic Surrogatism itself allows a similar response to this example: one can hold that names are context-sensitive, and can denote different elements of D_e in different contexts. An alternative contextualist theory treats this case as an example of what Blumberg and Lederman [2020] call "revisionist reports" (for discussion, see Blumberg and Lederman [2020, §7]). But both of these alternative forms of contextualism also have trouble with the cases I use to argue against functionalist theories below.

straightforwardly to a fine-grained functionalist theory as well). My diagnoses of these two different arguments are essentially the same, so I will only discuss one of them here. The first 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 Orcutt 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, who have all come to suspect that Olson, the police chief, is a spy. One night, while watching over the docks, they all see someone – as it happens, Ortcutt the mayor – in shady circumstances, and conclude that the person is a spy. But they all think that the person they saw was Olson; none suspects it was Ortcutt. They are led to this conclusion in part because they believe that Ortcutt the mayor is in the clear: he is not a spy. Thus, although each man thinks Olson is a spy, (25) no man thinks Ortcutt is a spy.³⁵

This story is simply a more detailed version of Anand's: 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 to another they believe he is not. 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. Perhaps the most obvious way for the functionalist to account for the change in context

³⁵For some of my consultants the final sentence is improved by deleting “a spy”, adding “yet” before “thinks”, or changing “no man” to “no investigator”, but all agree that the sentence is true in this setting.

between 22 and 23 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 them. So Anand’s case shows that functionalists need a more nuanced story about how 22 and 23 are both heard as true, which does not also predict that 25 will be heard as true in his case.³⁶

We can now see at last 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 it has a sufficiently easy to access true reading. The idea would then be that in the original Ralph story it is sufficiently easy to access both a true reading of 22 and a true reading of 23, but after hearing Anand’s story it is sufficiently easy to access a true reading of 24, but not of 25. This blueprint of a story does not pretend to be explanatory or predictive, but we can set that point aside. The problem is that the theory *still* fails to account for my examples. Since 14 and 15 are true in the same relevant contexts, the functionalist should hold that a true reading of 15 will be *just as easy* to access as a true reading of 15 (and similarly for 16 and 17).

More generally: since functionalists predict that 14 and 15 are true in exactly the same contexts, it is hard to see how they can tell a reasonable story about why one is heard as true, and the other as false. It is even harder to see how they could tell such a story which would also predict that 22 and 23 are both heard as true, and that 24 is heard as true and 25 as false (after Anand’s story). By contrast, the existentialist faces a much 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 may not be obvious how an existentialist should meet this challenge, but there is no principled reason to think that it cannot be met.

7.2 Problems with plural subjects

The new theory handles 18 by allowing different sets of permutations to be salient for different individuals. To see how this works, suppose that there are three relevant elements of D_e , h , p and v , corresponding to the names “Hesperus”, “Phosphorus” and “Venus”. In the context produced by the background story for 18 we may suppose that, relative to Eve and all possible worlds, p and v are equivalent and h is only equivalent to itself, while relative to Dawn and all

³⁶It 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 23, 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.

possible worlds, h and v are equivalent and p is only equivalent to itself. In a context where $f(\text{Eve}, w)$ and $f(\text{Dawn}, w)$ are the sets of permutations π such that for all $x \pi(x)$ is equivalent in these different equivalence relations, 18 will be true. Before Monday at noon, Eve did not know or believe that Phosphorus was not a star, while Dawn did not know or believe that Hesperus was not a star. But on Monday at noon, Eve came to know that Phosphorus was a star, and Dawn came to know that Hesperus was.

Once again it will help to see how this example constrains the official theory by considering an alternative. 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 f 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

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

Again, I’ve 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 here.

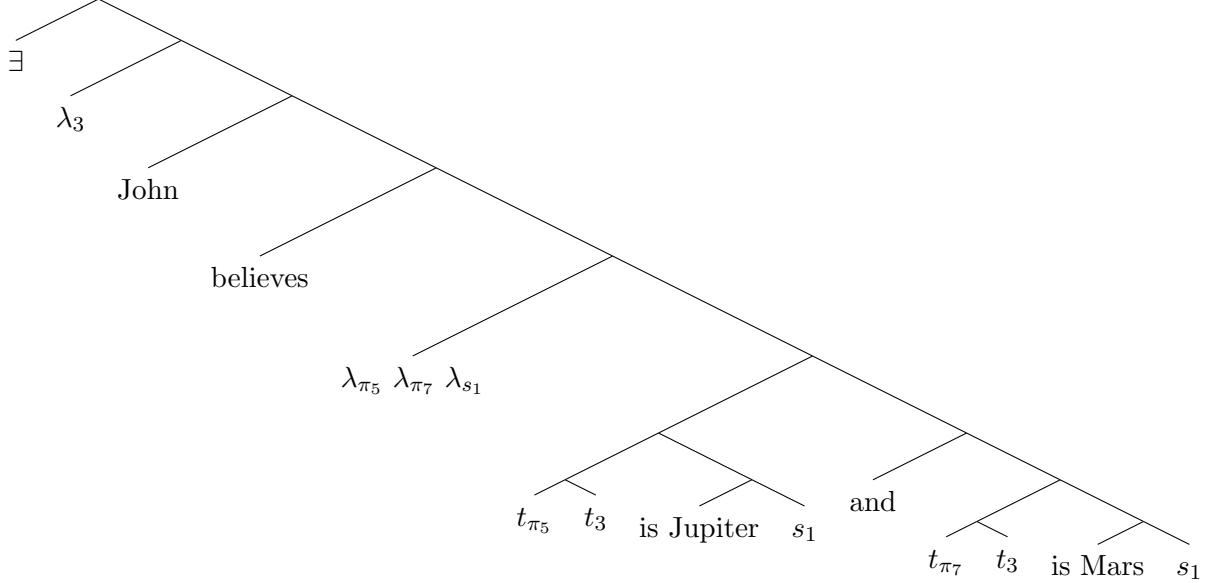
Insensitive theories cannot accommodate a true reading of the sentence 18. On such theories f only takes a world argument, so the same set of permutations will be used for each attitude holder. If this set includes one which maps $\llbracket \text{Venus} \rrbracket^{g,S,f}(w)$ to $\llbracket \text{Hesperus} \rrbracket^{g,S,f}(w)$, or one which maps $\llbracket \text{Venus} \rrbracket^{g,S,f}(w)$ to $\llbracket \text{Phosphorus} \rrbracket^{g,S,f}(w)$, then the proposition expressed by the complement clause of “learned” will fail condition (i) (from the conditions enumerated in section 5.2): either Eve or Dawn would have known it before. On the other hand, if the set of permutations contains no permutations which either map $\llbracket \text{Venus} \rrbracket^{g,S,f}(w)$ to $\llbracket \text{Hesperus} \rrbracket^{g,S,f}(w)$, or no permutations which map $\llbracket \text{Venus} \rrbracket^{g,S,f}(w)$ to $\llbracket \text{Phosphorus} \rrbracket^{g,S,f}(w)$, then the proposition expressed by the complement clause of “learned” will fail (ii), at least one person will not know the relevant proposition on Monday at noon.³⁷

These basic points apply not just to fine-grained theories but to Millian ones as well. (For the Millian version, take f in the entry above to be a function from worlds to concept-generators.) In either setting, a sensitive functionalist theory could account for the true reading of 18 and 19, but not for the contrast between 14 and 15. In either setting, an insensitive existentialist theory could account for the contrast between 14 and 15 but not for the true reading of 18 and 19. In this sense, the examples impose independent constraints on both the CG-theory and my fine-grained one.

³⁷Schiffer’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.

7.3 The bound de re

To see how the proposal handles 20, we assume the following syntax (abstracting from irrelevant world-pronouns and abstraction over worlds, and grouping some abstractions for the sake of space):



As observed by Charlow and Sharvit [2014], the key fact is that *different* permutation pronouns govern the different occurrences of the variable t_3 in this syntax. Recall that in the setup for this example, John believes that Hesperus is Jupiter and Phosphorus is Mars. The clause below “ λ_{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”.

More formally, relative to a g, S, f , such that $\llbracket \text{Jupiter} \rrbracket^{g, S, f}$ is $\lambda w.j$, while $\llbracket \text{Mars} \rrbracket^{g, S, f}$ 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 $w' \in DOX(x)(w)$, $\pi_1(\llbracket t_3 \rrbracket^{g, S, f})E_{w'}j$ and $\pi_2(\llbracket t_3 \rrbracket^{g, S, f})E_{w'}m$.³⁸

And this condition can be non-trivially be satisfied, since π_1 and π_2 can vary independently. By allowing different permutations to map the same element of D_e to different elements of D_e , the account can deliver an intuitive true reading of the sentence.

³⁸This assumes also that the world-argument of “believe” has been saturated by a world-pronoun which is not made explicit above.

Once again considering an alternative class of theories will help to show how the example constrains the 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

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

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.

Type-simple theories cannot accommodate a true reading of 20. Since they assume that there is only one pronoun for permutations, they predict that in the appropriate version of the syntax displayed above, 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 a $\pi \in f(x, w)$ such that for all $w' \in DOX(x)(w)$, $\pi(\llbracket t_3 \rrbracket^{g,S,f})E_{w'}j$ and $\pi(\llbracket t_3 \rrbracket^{g,S,f})E_{w'}m$.³⁹

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 20, but it would predict neither the contrast between 14 and 15, nor a true reading of 18.

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).

³⁹Here I am assuming assuming the world-argument of “believe” has been saturated.

	14 vs. 15 (Teachers)	18 (Dorr's datum)	20 (Bound de re)
F,I,TS	x	x	x
E,I,TS	✓	x	x
F,S,TS	x	✓	x
E,S,TS	✓	✓	x
F,I,TV	x	x	✓
E,I,TV	✓	x	✓
F,S,TV	x	✓	✓
E,S,TV	✓	✓	✓

8 The Indexed-Domain CG-theory

At the end of section 2, I noted that there are Millian variants on the CG-theory which can handle 3-5. In this section, I present such a theory but argue that my 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.⁴⁰ 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 $\llbracket \text{believes} \rrbracket^{g,f} = \lambda p. \lambda x. \lambda w. \text{either for all } w' \in DOX(x)(w), p(w') = 1 \text{ or for some } n \geq 1, \text{ there are } G_1 \in f(x)(w)(1), \dots, G_n \in f(x)(w)(n) \text{ such that for all } w' \in DOX(x)(w), p(G_1) \dots (G_n)(w') = 1.$ ⁴¹

In the CG-theory, f supplies a set of concept-generators as salient relative to each individual; we might call this theory a *single-domain* theory. Here, however, since f takes an extra numerical argument, it in effect supplies a sequence of sets of concept generators. The extra structure of these “indexed domains” allows the theory to escape the problem with 5. To see how, 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

⁴⁰Perhaps the most obvious variant of the CG-theory would be a functionalist one, which allows for an intuitive true reading of 5 by allowing the concept-generator variables “wrapping” the occurrences of “Hesperus” and “Phosphorus” to be 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.

⁴¹The theories of Ninan [2012] and Rieppel [2017] produce essentially the same results as this entry; it can be 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.)

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 predicts that 5 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”, a very good result.

At first sight, the indexed-domain CG-theory might seem to have clear advantages over my theory. Since it is a Millian theory, it does not require the use of multiple elements of D_e corresponding to a single individual, or the machinery of surrogate domain-restrictions.⁴² These benefits in simplicity come at what might seem the small cost of adding an additional numerical argument to f , the function which determines which concept-generators are salient relative to individuals.

But this is in fact a significant cost, which in my view 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 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 and 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, and this aspect of the theory leads ultimately to problems down the road. 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”. That is a good result. But relative to this f the sentence “Plato believed Phosphorus was brighter than Hesperus” would have the very same true reading, and, more generally, on the indexed-domain CG-theory the first of these sentences will be true in exactly the contexts where the second is, not a happy prediction.

In response to this problem, one might think that the indexed-domain CG-theory could 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 earlier for how the CG-theory could explain 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

⁴²Though see n. 18 for a way of making surrogate domain restrictions less flexible.

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. It is hard to understand what kind of context that would be. More generally, there is a concern that any natural way 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.⁴³

This line of thought gives my main reason for concluding that the fine-grained theory is preferable to the indexed-domain CG-theory. 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 of this kind. But I do have some subtle examples which at least have the right structure to discriminate between the theories, and I will present those examples to illustrate how such an argument might go:

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, producing two printouts for each individual. Amalia’s technician has two pictures of each of the subjects. To make the data easier to analyze, he is supposed to attach exactly one photo to every printout, matching the photos of the subjects with printouts of their genetic sequence. 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...”

26. Amalia thought Issa wasn’t Issa.
27. Amalia didn’t think Issa was Issa.
28. Amalia didn’t know that Issa was Issa.
29. Amalia thought Issa didn’t have any relevant genes in common with Issa.
30. Amalia didn’t think Issa had any relevant genes in common with Issa.
31. Amalia didn’t know that Issa had any relevant genes in common with Issa.⁴⁴

⁴³In Goodman and Lederman [forthcoming, §9.1] we develop related objections to a different Millian theory which delivers a true reading of 5.

⁴⁴Same-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].

Judgments about these sentences are very delicate. But 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 26 and 29, the two in which the negation takes narrow scope over only the complement clause of the attitude verb. The next most acceptable are 27 and 28, 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 30 and 31, 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 evidence for my theory, and against the indexed-domain CG-theory. Both theories predict true readings of 26 and 29, and both theories can explain the contrast between 27 and 28 on the one hand and 30 and 31 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 30 and 31. It predicts that these sentences have no intuitive true readings, essentially for the same reason that the original CG-theory predicted that 3-5 have no intuitive true readings. By contrast, the indexed-domain CG theory predicts that 30 and 31 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 30 and 31, 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 30 and 31 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:

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

Both the CG-theory and my theory will (correctly) predict that there is no single context where both of these sentences are true, but there are (different) contexts in which each of them receives a natural true reading.

9 Conclusion

In this paper, 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 as illustrated in section 5. In developing my account I have used 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 simplicity and predictive strength.

A main goal of the paper has been to show how the central examples (3-5; 14-17; 18-19; 20) impose distinct structural constraints on any theory of attitude reports. For concreteness I have developed this point within a general framework inspired by the CG-theory, where the constraints can be seen described in terms of 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 constrain a wide array of theories of attitude reports, and I hope my discussion here will spur further exploration of how they might be accommodated in other frameworks as well.

References

- Maria Aloni. Individual concepts in modal predicate logic. *Journal of Philosophical Logic*, 34(1):1–64, 2005. doi: 10.1007/s10992-004-4065-8.
- Pranav Anand. *De De Se*. PhD thesis, MIT, 2006.
- Andrew Bacon and Jeffrey Sanford Russell. The logic of opacity. *Philosophy and Phenomenological Research*, 2017. doi: 10.1111/phpr.12454.
- Chris Barker. The linguistics of coordinated thought. Unpublished Manuscript, February 2016.
- John C. Bigelow. Believing in semantics. *Linguistics and Philosophy*, 2(1):101–144, 1978. doi: 10.1007/BF00365131.
- Jens Christian Bjerring. Impossible worlds and logical omniscience: An impossibility result. *Synthese*, 190(13): 2505–2524, 2013. doi: 10.1007/s11229-011-0038-y.
- Jens Christian Bjerring and Wolfgang Schwarz. Granularity problems. *Philosophical Quarterly*, 67(266):22–37, 2017. doi: 10.1093/pq/pqw028.
- Kyle Blumberg and Harvey Lederman. Revisionist reporting. *Philosophical Studies*, 2020. doi: 10.1007/s11098-020-01457-4.
- Tyler Burge. Reference and proper names. *Journal of Philosophy*, 70(14):425–439, 1973. doi: 10.2307/2025107.
- John P. Burgess. Tarski’s tort. In *Mathematics, Models, and Modality: Selected Philosophical Essays*, page 149–168. Cambridge University Press, 2008. doi: 10.1017/CBO9780511487347.010.
- Seth Cable. A choice-functional semantics for *De Re* attitude reports. Unpublished Manuscript, July 2018.
- Michael Caie, Jeremy Goodman, and Harvey Lederman. Classical opacity. *Philosophy and Phenomenological Research*, 101(3):524–566, 2020. doi: 10.1111/phpr.12587.
- Rudolf Carnap. *Meaning and necessity: a study in semantics and modal logic*. University of Chicago Press, 1947.
- Simon Charlow and Yael Sharvit. Bound ‘de re’ pronouns and the LFs of attitude reports. *Semantics and Pragmatics*, 7(3):1–43, 2014. doi: 10.3765/sp.7.3.
- M. J. Cresswell. Hyperintensional logic. *Studia Logica*, 34(1):25–38, 1975. doi: 10.1007/BF02314421.

- Maxwell J Cresswell and Arnim Von Stechow. De re belief generalized. *Linguistics and Philosophy*, 5(4): 503–535, 1982. doi: 10.1007/BF00355585.
- Mark Crimmins. *Talk about beliefs*. Mit Press, 1992.
- Mark Crimmins and John Perry. The prince and the phone booth: Reporting puzzling beliefs. *The Journal of Philosophy*, 86(12):685–711, 1989. doi: 10.2307/2027014.
- Sam Cumming. On an alleged ambiguity in attitude reports. Unpublished Manuscript, 2016.
- Samuel Cumming. Variabilism. *Philosophical Review*, 117(4):525–554, 2008. doi: 00318108-2008-015.
- Joshua Bruce Dever. *Variables*. PhD thesis, University of California, Berkeley, 1998.
- Cian Dorr. Transparency and the context-sensitivity of attitude reports. In Manuel Garcia-Carpintero and Gómez Martí, editors, *Empty Representations: Reference and Non-existence*, pages 25–66. Oxford University Press, 2014. doi: 10.1093/acprof:oso/9780199647057.003.0002.
- Paul Elbourne. *Situations and Individuals*. MIT Press, 2005.
- Delia Graff Fara. Names are predicates. *Philosophical Review*, 124(1):59–117, 2015. doi: 10.1215/00318108-2812660.
- Jerry A. Fodor. *The Language of Thought*. Harvard University Press, 1975.
- Gottlob Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50, 1892. doi: 10.2307/2181485. Translated as “On Sense and Reference” by Max Black in *Philosophical Review*, 57.3: 209–230, 1948.
- Jeremy Goodman and Harvey Lederman. Perspectivism. *Noûs*, forthcoming. doi: 10.1111/noûs.12322.
- Irene Heim. Anaphora and semantic interpretation: A reinterpretation of reinhart’s approach. In Uli Sauerland and Orin Percus, editors, *The interpretive tract: Working papers in syntax and semantics*, pages 205–246. Department of Linguistics, Massachusetts Institute of Technology, Cambridge, Mass., 1998.
- James Higginbotham. Belief and logical form. *Mind and Language*, 6(4):344–369, 1991. doi: 10.1111/j.1468-0017.1991.tb00261.x.
- Jaakko Hintikka. *Knowledge and Belief*. Ithaca: Cornell University Press, 1962.
- Wesley H Holliday and John Perry. Roles, rigidity, and quantification in epistemic logic. In *Johan van Benthem on Logic and Information Dynamics*, pages 591–629. Springer, 2014. doi: 10.1007/978-3-319-06025-5_22.
- David Kaplan. Quantifying in. *Synthese*, 19:178–214, 1968. doi: 10.1007/BF00568057.
- David Kaplan. Opacity. In Lewis Edwin Hahn and Paul Arthur Schilpp, editors, *The Philosophy of W. V. Quine*, pages 229–289. Open Court, 1986.
- Ezra Keshet. *Good Intensions: Paving Two Roads to a Theory of the De re/De dicto Distinction: Massachusetts Institute of Technology Ph. D.* PhD thesis, MIT, 2008.
- Ezra Keshet. Possible worlds and wide scope indefinites: A reply to Bäuerle 1983. *Linguistic Inquiry*, 41(4): 692–701, 2010. doi: 10.1162/LING_a_00019.
- Ezra Keshet. Split intensionality: a new scope theory of *de re* and *de dicto*. *Linguistics and Philosophy*, 33: 251–283, 2011. doi: 10.1007/s10988-011-9081-x.
- Saul A. Kripke. Naming and necessity. In Donald Davidson and Gilbert Harman, editors, *Semantics of Natural Language*, pages 253–355, 763–769. Dordrecht: D. Reidel, 1972. doi: 10.1007/978-94-010-2557-7_9.
- Saul A Kripke. A puzzle about belief. In Avishai Margalit, editor, *Meaning and use*, pages 239–283. D. Reidel, Dordrecht, 1979. doi: 10.1007/978-1-4020-4104-4_20.
- Richard Larson and Gabriel Segal. *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT Press, 1995. doi: 10.7551/mitpress/4076.001.0001.
- Harvey Lederman. Fregeanism, sententialism, and scope. Unpublished Manuscript, January 2020.
- David Lewis. Attitudes *de dicto* and *de se*. *Philosophical Review*, 88:513–543, 1979. doi: 10.2307/2184843.
- David K. Lewis. General semantics. *Synthese*, 22(1-2):18–67, 1970. doi: 10.1007/BF00413598.
- Ora Matushansky. On the linguistic complexity of proper names. *Linguistics and Philosophy*, 31(5):573–627, 2008. doi: 10.1007/s10988-008-9050-1.
- Sarah Moss. The role of linguistics in the philosophy of language. In Delia Graff Fara and Gillian Russell, editors, *Routledge Companion to the Philosophy of Language*. Routledge, 2012.
- Dilip Ninan. Counterfactual attitudes and multi-centered worlds. *Semantics and Pragmatics*, 5(5):1–57, 2012.

doi: 10.3765/sp.5.5.

Dilip Ninan. Quantification and epistemic modality. *Philosophical Review*, 127(4):433–485, 2018. doi: 10.1215/00318108-6973010.

Orin Percus. Constraints on some other variables in syntax. *Natural language semantics*, 8(3):173–229, 2000. doi: 10.1023/A:1011298526791.

Orin Percus and Uli Sauerland. On the LFs of attitude reports. In Matthias Weisgerber, editor, *Proceedings of Sinn und Bedeutung*, 2003.

John Perry. Frege on demonstratives. *Philosophical Review*, 86(4):474–497, 1977. doi: 10.2307/2184564.

Bryan Pickel. Variables and attitudes. *Noûs*, 49(2):333–356, 2015. doi: 10.1111/nous.12044.

Willard V Quine. Quantifiers and propositional attitudes. *the Journal of Philosophy*, 53(5):177–187, 1956. doi: 10.2307/2022451.

Mark Richard. Direct reference and ascriptions of belief. *Journal of Philosophical Logic*, 12(4):425–52, 1983. doi: 10.1007/BF00249259.

Mark Richard. *Propositional attitudes: An essay on thoughts and how we ascribe them*. Cambridge University Press, 1990. doi: 10.1017/CBO9780511625282.

Michael Rieppel. Names, masks, and double vision. *Ergo*, 2017. doi: 10.3998/ergo.12405314.0004.008.

Bertrand Russell. On denoting. *Mind*, 14:479–493, 1905. doi: 10.1093/mind/XIV.4.479.

Nathan Salmon. Reflexivity. *Notre Dame Journal of Formal Logic*, 27(3):401–429, 1986. doi: 10.1305/ndjfl/1093636683.

Paolo Santorio. On the plurality of indices. Unpublished Manuscript, 2014.

Uli Sauerland. Counterparts block some ‘de re’ readings. In Uli Sauerland and Luka Crnič, editors, *The art and craft of semantics: A festschrift for Irene Heim*, volume 2, pages 65–85. Department of Linguistics, Massachusetts Institute of Technology, Cambridge, Mass, 2015.

Stephen Schiffer. Naming and knowing. In A. French Peter, E. Uehling Theodore, Howard Jr, and K. Wettstein, editors, *Midwest Studies in Philosophy II: Contemporary Perspectives in the Philosophy of Language*, pages 28–41. University of Minnesota Press, 1979.

Stephen Schiffer. Belief ascription. *The Journal of Philosophy*, 89(10):499–521, 1992. doi: 10.2307/2941005.

Anders Schoubye. Names are variables. *The Philosophical Review*, forthcoming.

Florian Schwarz. Situation pronouns in determiner phrases. *Natural Language Semantics*, 20(4):431–475, 2012. doi: 10.1007/s11050-012-9086-1.

Yael Sharvit. Covaluation and unexpected bt effects. *Journal of Semantics*, 28(1):55–106, 2010. doi: 10.1093/jos/ffq012.

Scott Soames. Lost innocence. *Linguistics and Philosophy*, 8(1):59–71, 1985. doi: 10.1007/BF00653375.

Scott Soames. Pronouns and propositional attitudes. *Proceedings of the Aristotelian Society, New Series*, 90: 191–212, 1989–90. doi: 10.1093/aristotelian/90.1.191.

Scott Soames. Attitudes and anaphora. *Philosophical Perspectives*, 8:251–272, 1994. doi: 10.2307/2214173.

Seth Yalcin. Quantifying in from a Fregean perspective. *Philosophical Review*, 124(2):207–253, 2015. doi: 10.1215/00318108-2842186.

Seth Yalcin. Semantics as model-based science. In Derek Ball and Brian Rabern, editors, *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, pages 334–360. Oxford University Press, 2018. doi: 10.1093/oso/9780198739548.003.0012.

Thomas Ede Zimmerman. Kontextabhängigkeit. In A. von Stechow and D. Wunderlich, editors, *Semantik: Ein Internationales Handbuch Der Zeitgenössischen Forschung*, pages 151–229. De Gruyter, Berlin, 1991.

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 heavier-weight 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 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 here; 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 as a form of descriptivism.) Following Kripke [1972], we assume that such uses of names are modally rigid: 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 can 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 $\{f \in X^W \mid \text{if } wRw' \text{ then } f(w) = f(w')\}$, and the relation E_w be $\{\langle f, f' \rangle \mid f(w) = f'(w)\}$.

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 version of the theory should be modified in some way.⁴⁵ Interestingly, using permutations (which are naturally

⁴⁵In 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 , and c , 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 and a y such that $x \neq y$ and $(x = a \text{ or } y = a)$ and $(x = b \text{ or } y = b)$ and $(x = c \text{ or } y = c)$ ”. 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 such that $x \neq y$ and John thinks $x = y$); 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

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, Aloni-inspired version of the theory.

These descriptivist ways of viewing the model theory “explain away” elements of D_e in other terms, but they do not do the same for the impossible worlds in W . Some descriptivists and predicativists may not see this as a problem, if they are happy to take epistemically possible, but metaphysically impossible worlds as primitive. But it is worth noting that we could also 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 \mathcal{C} be the set of conceptual covers of X in S . We then define the ingredients of my models as follows:

- $W = S \times \mathcal{C}$;
- $R = \{\langle\langle s, c \rangle, \langle s', c' \rangle\rangle \in W \times W \mid c = c'\}$;
- $D_e = \{f \in (X^S)^{\mathcal{C}} \mid f(C) \in C \text{ for all } C \in \mathcal{C}\}$;
- 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 \mathcal{N} for the set of indices, the set of such assignment functions is $X^{\mathcal{N}}$. We then define the ingredients of my models as follows:

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.

- W is $S \times X^{\mathcal{N}}$, the set of pairs of possible worlds and assignment functions;
- R is $\{\langle\langle s, i \rangle, \langle s', i' \rangle \rangle \mid i = i'\}$;
- D_e is \mathcal{N}
- for all $\langle s, i \rangle \in W$, $E_{\langle s, i \rangle}$ is $\{\langle n, m \rangle \in D_e \times D_e \mid i(n) = i(m)\}$.

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 someone₁ such that John thinks he₂ is wonderful”). Cumming [2008] and Pickel [2015] offer different solutions to 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 assignment functions in context. If we take the set \mathcal{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 14-17, 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, since there are few if any theory-independent constraints on when indices are assigned to objects.)

A third way of viewing these models can be seen as motivated by this last problem with taking the semantic values of names to be themselves. We take the set $\mathcal{I} = D_e$ 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. 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 14-17:

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

They claim that 34 and 36 are naturally interpreted as true, while 35 and 37 are naturally interpreted as false. My consultants have almost universally disagreed with these judgments. Some report something of a contrast between 36 and 37, but almost none report that contrast for 34 and 35. 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.⁴⁶

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

Surrogatist No

$$[\![\text{no}]\!]^{g,S,f} = \lambda w. \lambda F. \lambda G. \text{ no } x \in S_w \text{ is such that}$$

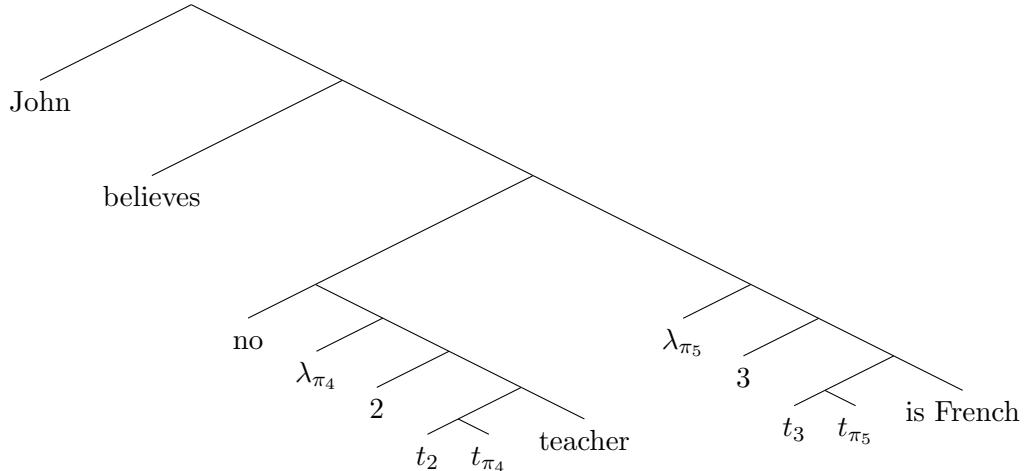
for some $n \geq 0$ there are $\pi_1 \dots \pi_n \in \bigcup_{x \in D_e} f(x, w)$ such that $F(\pi_1) \dots (\pi_n)(x) = 1$, and
for some $m \geq 0$ there are $\pi_1 \dots \pi_m \in \bigcup_{x \in D_e} f(x, w)$ such that $G(\pi_1) \dots (\pi_m)(x) = 1$.⁴⁷

⁴⁶Cable [2018] develops an account designed to predict the contrast between these in-scope data. His account of this contrast is very elegant. But his 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 of the second sentence here.) Second, Cable’s theory cannot accommodate the data discussed in section 5.2.

⁴⁷In 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

Here “ $F(\pi_1) \dots (\pi_0)$ ” abbreviates “ F ”, and similarly for $G(\pi_1) \dots (\pi_0)$; 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:



Here the binder λ_{π_5} is produced beneath “no teacher” and the binder λ_{π_4} is produced below “no”. We assume that even with simple predicates like “is a teacher” there is a λ -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 35 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:

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

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.

Given this entry, assuming that all w -permutations are salient, 34 would be true, and 35 false.⁴⁸

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:

38. 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.⁴⁹ 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 20 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 the 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,

⁴⁸Charlow and Sharvit suggest that either attitude verbs are ambiguous between a "universal" and an "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 14 and 15, 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 17 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]).

⁴⁹Richard [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.

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 when it occurs within the scope of the attitude verb.⁵⁰

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".

39. 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 that proposal was well-motivated in the first place. But I'm not moved by Charlow and Sharvit's original examples, and I don't think the judgments about 38 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.⁵¹

⁵⁰A 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.

⁵¹There 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.