

Modals are monsters: on indexical shift in English*

Paolo Santorio

Massachusetts Institute of Technology

Abstract I argue, *contra* Kaplan, that all English indexicals are shiftable. Starting from a new puzzle involving epistemic conditionals, I propose a semantics that accounts for embeddings of indexicals under non-metaphysical modals like *might* or *believe*. The central idea is that indexicals are variables: epistemic and doxastic modals are able to shift them by shifting the value of the assignment function. So all epistemic and doxastic modals turn out to be Kaplanian monsters. I close by arguing that the monstrous account has empirical advantages over the classical semantics for attitudes in the Kaplan tradition, and in particular over the recent version of it proposed by Percus and Sauerland.

Keywords: indexicality, monsters, modality, attitude reports, Kaplan, *de re*

1 Introduction

On Kaplan's (1989) classical theory of indexicality, the semantic values of indexicals like *I*, *you* and *now* are identified with their referents.¹ This theory leads to well-known difficulties when indexicals occur in attitude and speech reports. Spotting a man whose hair is enveloped in flames, I can warn him by saying

(1) I believe that you're on fire.

Moreover, unless I also believe that my own hair is enveloped in flames, I can add

(2) I believe that I am not on fire.

(1) and (2) both have a true reading even if, unbeknownst to me, I'm looking into a mirror and I am actually the man whose hair is on fire. This is puzzling: if *you* and *I* in (1) and (2) denote the same individual, on Kaplan's theory they have the

* For comments, questions, mentorship, and encouragement thanks to Fabrizio Cariani, Jennifer Carr, Irene Heim, Dilip Ninan, Orin Percus, Alejandro Pérez Carballo, Agustín Rayo, Philippe Schlenker, Yael Sharvit, Brad Skow, Bob Stalnaker, Andreas Stokke and to audiences at MIT, SPE 2009, the AAP 2009 conference in Melbourne, SALT 20 and the Arché Semantics Group in St. Andrews.

¹ Here and throughout the paper I use 'semantic value' to denote the kind of object that a lexical item contributes to the truth-conditional computation.

same semantic value. Then (1) and (2) should be heard as contradictory, or at least as ascribing to me contradictory beliefs. Yet they are perfectly good reports of my (consistent) attitudes.

Problems of this sort are familiar since Frege and have given rise to an extensive literature in philosophy and semantics.² One common assumption in this literature is that the interpretation of indexicals can't be affected by embeddings under *believe*. This is a straightforward consequence of Kaplan's theory. For Kaplan only context can fix the semantic values of indexicals: operators that shift indexicals, which Kaplan dubs 'monsters', are no part of the semantics of natural language (at least, of natural languages like English). Interestingly, recent literature has shown that the ban on monsters fails cross-linguistically.³ But this doesn't help with the puzzle in (1) and (2): in fact it remains virtually unquestioned that common English indexicals like *I*, *you* or *now* are unshiftable, just as Kaplan had it. So *you* and *I* in (1) and (2) pick out, as is customary, the addressee and the speaker, and our semantics for attitude reports must work around this basic fact.

Or so it is thought. In this paper I argue that in fact there is a monstrous treatment of indexicals which has empirical and conceptual advantages over traditional stories. On this account all English indexicals work as variables; modal quantifiers with epistemic flavor, such as *must*, *might* and *believe*, shift the assignment function and thereby shift indexicals. So for example in a sentence like

(3) I might be lost in Main Library, Stanford.

I is a variable and its semantic value is fixed by an assignment determined by *might*. As a result, *I* is associated with a range of *epistemic counterparts of the speaker*, i.e. possible individuals who, for all the speaker knows, the speaker herself might be. This theory should be preferred over its competitors, which involve heavier stipulations or lead to overgeneration in a number of cases. If the foregoing is correct, it is a partial argument that the semantics for attitude and speech reports is monstrous, and that—*pace* Kaplan—English is replete with monsters.

I proceed as follows. In section 2, I introduce new data showing that the Frege-style phenomenon generated by (1) and (2) can be reproduced with epistemic modals. The new data is interesting because, much more than its counterpart involving *believe*, it naturally points to a monstrous account. In section 3, I proceed to give a

² In both cases the starting point for a formal treatment of the problem is Kaplan 1968; cf. also section 20 of Kaplan 1989.

³ More precisely, it has been argued that speech report verbs in some languages (Amharic, Zazaki, Slave and perhaps, in a restricted fashion, even English) are able to produce shifted readings of indexicals. For monstrous semantic accounts of all these languages, cf. Schlenker 1999, 2003; for the Zazaki, Anand & Nevins 2004 and Anand 2006. An alternative account of the data is attempted by von Stechow 2002 and von Stechow 2003. For useful summaries of the debate, cf. Schlenker 2004, 2008.

monstrous theory which accommodates the data. Section 4 contains the argument that this theory is superior to the traditional Kaplan-style account of attitude reports (and in particular to one of the best and latest versions of it, that put forward by Percus & Sauerland (2003)). In the conclusion, I point to directions for future research.

2 The puzzle

2.1 Data

My starting observation is this: indexicals embedded under epistemic modalities can have *prima facie* ‘shifty’ readings. Consider the following scenario:

Rudolf Lingens and Gustav Lauben are kidnapped. Lingens and Lauben are amnesiacs: each of them knows that he is one of the two kidnapped amnesiacs, but not which one. They will be subjected to the following experiment. First, they will be anesthetized. Then a coin will be tossed. If the outcome is tails, Lingens will be released in Main Library, Stanford, and Lauben will be killed. If the outcome is heads, Lauben will be released in Widener Library, Harvard, and Lingens will be killed. Lingens and Lauben are informed of the plan and the experiment is executed. Later, one of them wakes up in a library. He says:

- (4) If the coin landed tails, I am in Main Library, Stanford.
- (5) If the coin landed heads, I am in Widener Library, Harvard.

Here are the data: (4) and (5) are both true (or at least appropriate) as uttered in the scenario. The problem is that standard accounts of conditionals miss this prediction.

For illustration, take a basic version of Kratzer’s (1981) semantics for conditionals. Conditionals are, as usual, modal statements where the *if*-clause plays the role of a restrictor. The domain of quantification (modal base) is specified by a covert variable the value of which is provided by context: for simplicity, here I assume that the modal base defaults to the epistemic state of the speaker, i.e. the set of worlds compatible with what the speaker knows.⁴ So the truth-conditions predicted for *if* p , q are, schematically:

$$\llbracket \text{must [if } p \text{]} [q] \rrbracket^{c,g} = \text{true iff for all } w' \text{ compatible with what the speaker knows in } w_c \text{ and such that } w' \in \llbracket p \rrbracket^{c,g}, w' \in \llbracket q \rrbracket^{c,g}.$$

⁴ I should flag that this assumption is rather simplistic, and in fact is rejected by virtually everyone in recent debates about epistemic modality. This is not important for my purposes, as the puzzle can be reproduced on more sophisticated assumptions.

Now, consider again the amnesiac scenario and suppose that Lingens is the speaker. On the assumption that *I* invariably denotes the actual speaker, (5) gets truth-conditions:

[[5]]^{c,g} = true iff for all w' in Lingens's epistemic state in w_c such that the coin landed heads in w' , Lingens is in Widener Library in w' .

But in all heads-worlds in Lingens's epistemic state Lingens is *not* in Widener Library—rather, he is dead. Hence (5) is wrongly predicted to be false. Conversely, if Lauben is the speaker we get a wrong prediction for (4).

In short, then, the difficulty is this: whatever the context of utterance is (i.e. whether we take Lingens or Lauben as the speaker), one of (4) and (5) is predicted to be false. But both conditionals are true according to the data.

2.2 Generalizing the problem

The puzzle generalizes along two dimensions: it arises both for other indexicals and for other modals. Let me look at other indexicals first. Here are examples involving *now* and *that*:

You fall asleep at around noon, and you wake up hearing the clock in your room chiming. You know it chimed once or twice, but you're not sure which. You say:

- (6) If the clock chimed once, it is now one.
- (7) If the clock chimed twice, it is now two.

A group of climbers has gotten lost in the mountains between France and Italy. They are finally descending on one of the two sides, though they don't know which. One points to a city in the distance and says:

- (8) If we are on the French side, that [pointing] is Lyon.
- (9) If we are on the Italian side, that [pointing] is Turin.

The problem is analogous. All of (6)–(9) are true, but on a standard Kratzer semantics one conditional in each pair turns out to be false, no matter what the actual context is. So far as I can see, analogous puzzles can be constructed for all English indexicals.

What about other modals? Interestingly, results differ depending on the kind of modality involved. Conditionals invoking metaphysical modality,⁵ such as counterfactuals, behave exactly as predicted by a standard Kratzer semantics. Suppose that

⁵ By 'metaphysical' here I mean to pick out simply whatever modality will turn out to do work in typical counterfactual conditionals.

the amnesiac in the original scenario is recapitulating to himself his predicament. Trying to describe the causal connections between possible outcomes of the coin toss and amnesiacs waking up in libraries, he says

(10) If the coin had landed tails, I would have been in Main Library, Stanford.

(11) If the coin had landed heads, I would have been in Widener Library, Harvard.

A joint utterance of (10) and (11) is infelicitous, and in any case *not* on a par with a joint utterance of (4) and (5). (One could reply “No, that’s wrong—it wouldn’t have been you in both cases.”)

On the other hand, the data can be easily replicated with other non-metaphysical modals. Following Hintikka (1962), I assume that attitude verbs also work as quantifiers over possibilities. For example, *believe* is a universal quantifier over worlds which model an agent’s *doxastic alternatives*—worlds that, for all the agent believes, might be the actual world. Schematically:

$$\llbracket x \text{ believes [that } p \rrbracket \rrbracket^{c,g} = \text{true iff for all } w' \text{ compatible with what } x \text{ believes, } w' \in \llbracket p \rrbracket^{c,g}$$

Now, suppose that, rather than (4) and (5), the amnesiac in the library scenario says:

(12) I believe that either the coin landed tails and I am in Main Library, Stanford, or the coin landed heads and I am in Widener Library.

(12) presents us with basically the same problem as the (4)–(5) pair. There is no individual that is going to be in Main in case of a tails outcome and in Widener in case of a heads outcome. So, on a simple Hintikka-style semantics for *believe*, (12) is predicted to be false no matter what the context is. For example, if the speaker is Lingers, its truth-conditions are

$$\llbracket (12) \rrbracket^{c,g} = \text{true iff for all } w' \text{ compatible with what Lingers believes in } w_c, \text{ either the coin landed tails and } \underline{\text{Lingers}} \text{ is in Main, or the coin landed heads and } \underline{\text{Lingers}} \text{ is in Widener in } w'.$$

In the remainder of the paper I use *might*, *must*, and *believe* as my running examples. But the phenomenon can be reproduced at least for a large subclass of attitude verbs: some examples are *know*, *suppose*, *imagine*, *conjecture*, *dream*. So the right generalization seems to be that the phenomenon concerns the behavior of indexicals in the scope of epistemic and doxastic modals in general.

Let me also mention two directions of generalization that I won’t explore for reasons of space: on the one hand tense, on the other referential expressions other than indexicals. I will come back to the latter briefly in the conclusion.

2.3 Preview

The phenomenon spotted in this section clearly belongs to a larger family of Frege-style issues concerning the interaction between indexicals and non-metaphysical modals. But it deserves specific attention because, much more than other puzzles in the vicinity, it suggests a natural solution. Consider again (4) and (5), repeated below:

(13) If the coin landed tails, I am in Main Library, Stanford.

(14) If the coin landed heads, I am in Widener Library, Harvard.

It seems intuitive that the two conditionals work in this way: the *if*-clause selects a subset of the worlds among those in the modal base, and the *I* occurring in the consequent refers to *the individual who is speaking in those worlds*.

In the next section I propose an account that endorses and implements this intuition: *I* in (4) and (5) works as a variable that is bound by epistemic *must*. Of course, the main challenge is to allow for this kind of monstrous behavior while avoiding overgeneration. In particular, any plausible account should recapture the standard data about unembedded indexicals, counterfactuals, etc., that motivated Kaplan's original account. This can be done, as I argue below.

3 A monstrous account

3.1 Theoretical background: counterparts and information states

Start from a general question: how should we evaluate indexicals embedded under non-metaphysical modals like *might* and *believe*?

This question is the linguistic counterpart of a traditional question in philosophy of mind. Epistemic and doxastic modals, recall, quantify over worlds modeling *information states* of a subject: namely, what a subject knows, believes, etc. We standardly talk of these information states as representing specific individuals: for example, we can say that Lingens believes that *I* am lost in a university library. But how am I individuated in Lingens's information state? I.e., what makes Lingens's information state an information state about *me*, a real individual in the actual world? Call this the *problem of individuation*.

The standard solution to the problem (coming from Hintikka (1969) and Lewis (1983)) is this: I have a representative in each of Lingens's belief worlds; Lingens counts as knowing that *I* am lost in a university library in virtue of the fact that each of my representatives in his knowledge worlds is lost in a university library.⁶ What individuals are my representatives? In informal terms, individuals that resemble me

⁶ Why can't *I myself* be my own representative in each of Lingens's belief states? The problem is

in certain relevant respects. In more formal terms: individuals that are *counterparts* of me, where a relation of counterparthood is just a relation of similarity (cf. Lewis 1968, 1983): *A* is a counterpart of *B* under a certain respect just in case *A* and *B* are sufficiently similar in that respect.

I suggest that the problem of evaluating indexicals under *might* or *believe* is essentially a linguistic version of the problem of individuation. So it can be solved with the same tools. Objects are individuated in an information state by individuating a set of their representatives, one for each world in the information state. Similarly, and in a slogan: indexicals embedded under *might*, *believe* and the like denote their referents by denoting a set of their representatives. Out of metaphor, this means that they work as bound variables and range over counterparts of the referents they pick out in ordinary linguistic environments. So e.g. in

(15) Lingens believes that I am lost in a university library.

I is bound and ranges over counterparts of the actual speaker. Similarly, bound occurrences of *you* will range over counterparts of the actual addressee; bound occurrences of *that* (like those occurring in (8) and (9)) will range over counterparts of the (*n*-th) object demonstrated; and so on.

I have talked about similarity and counterparts, but so far I haven't specified what notion of similarity we should use. For present purposes, it will be enough to import an idea from Kaplan's semantics for *de re* reports. I assume that a subject bears a relation of *acquaintance* to the individuals he has attitudes about. Acquaintance is supposed to be a kind of direct epistemic relation to an object, of which perception is the paradigm example. Kaplan's central idea is that acquaintance relations figure in the content of a subject's beliefs about objects. For example Lingens, having seen me wander around the library, thinks of me as *the man I saw wandering around the library*. So in each of his belief worlds there is a man he has seen wonder around the library: just these men, who are my counterparts by acquaintance, are my representatives in Lingens's belief state. More generally, bound occurrences of *I*, *you*, etc. range over *counterparts by acquaintance* of the actual speaker, addressee, etc.⁷

that real-world identity is ill-suited to model the ways agents are represented throughout information states. As Quine (1956) pointed out, I might enter Lingens's beliefs under two guises: for example as *that man in front of him*, next to the stack of semantics books, and *that man reflected in the mirror over there*, who (for all Lingens can see) might or might not be next to the stack of semantics books. If we invariably used myself as my own representative throughout Lingens's doxastic state we should say that Lingens has inconsistent beliefs, since in some alternatives I would both be and not be next to the stack of semantics books. But Lingens's beliefs are not contradictory—he's just ignorant about certain matters of fact. (For extended discussion of this problem, see Lewis 1983 and Ninan 2008)

⁷ For more details about acquaintance see, among others, Lewis 1979. I should flag that Kaplan's

3.2 Semantics: overview

The semantics is based on three main innovations on classical theories. First, indexicals are variables and their anchoring to the context is determined by extra-semantic constraints on the choice of the assignment. Second, modals like *might* and *believe* quantify not simply over worlds, but rather over pairs of an assignment and a world. Third, these modals shift the value of the assignment and as a result shift indexicals. Let me present each of these points informally before turning to the formal system.

First, indexicals are turned into simple variables. By ‘simple’ I mean that their lexical entries contain no reference to the context, not even in a presupposition. In a slogan, *I* is semantically analogous to *he* (at least with respect to its person features). The only difference is that indexicals are given referents not via the ordinary assignment, but by a second dedicated assignment, which I denote with ‘*a*’. The connection between indexicals and context is not part of the semantics proper, but takes place at the level of metasemantics, i.e. at the stage at which the dedicated assignment is selected. So nothing in the semantics of *I* tells us that it must refer to the speaker, but it is a constraint on the choice of *a* that *I* be invariably sent to the speaker.

Second, modals are turned into more powerful quantifiers. *might*, *believe* and the like quantify not simply over worlds, but rather over pairs of assignments (i.e. functions from indices to sequences of individuals), and centered worlds. This move is not new. It is a natural generalization of our standard semantics for *de se* reports and variants of it appear frequently in the literature on *de re* reports.⁸

So far, then, my account of modals looks similar to other proposals in the literature. The innovation consists in allowing the modal to shift indexicals by shifting the assignment *a*. In the new system *might*, *believe* and the like shift the value of the assignment *a*, forcing the evaluation of the embedded clause at the alternative assignments they quantify over. Schematically:

$$\llbracket x \text{ believes that } [\phi] \rrbracket^{c,g,a} = \text{for all } \langle a', w' \rangle \text{ compatible with } x\text{'s beliefs} \\ \text{at } w_c, \llbracket \phi \rrbracket^{c,g,a'}(w')$$

idea of using acquaintance for the semantics of *de re* attitudes is subject to a number of objections: see e.g. Ninan 2008, especially chapters 2 and 3. I am aware of this, and suspect that ultimately we should switch to a better notion. But the issue of what counterpart relation is in place, although important, seems orthogonal to my main point here. So it seems easier to go with Kaplan on this point.

⁸ Typical examples are accounts derived from Kaplan 1968: cf. among others Cresswell & von Stechow 1982 and Ninan 2008. Interestingly, Cumming 2008 has attitude verbs quantify over pairs of assignments and worlds and bind names. There are obvious similarities between Cumming’s semantics and mine, as well as important differences. (For example the system I suggest, unlike his, enforces obligatory shift of the assignment.)

Modals are monsters

This requires that the clausal arguments of modals be not propositions, but rather functions of higher type, i.e. functions from (again) an assignment and a centered world to a truth-value. We get these arguments by introducing a binder on the assignment. For example, in (15) (repeated below)

(16) Lingens believes that I am lost in a university library.

the clausal argument of *believe* is

$$\lambda a'. \llbracket \text{I am lost in a university library} \rrbracket^{c,g,a'}$$

It should be clear how, on the semantics just sketched, *might* and *believe* are able to affect the semantic values of indexicals. The idea is simple: they shift indexicals by shifting the assignment on which their semantic values depend.

Notice an interesting formal analogy between assignment-shifting modals and intensional operators in intensional systems. Both manipulate a metalanguage parameter: an assignment in the former case, and the world or time parameter in the latter. To visualize the point, consider a schematic intensional semantics for *must*:

$$\llbracket \text{must } [\phi] \rrbracket^{c,w,g} = \text{for all accessible } \underline{w}', \llbracket \phi \rrbracket^{c,w',g}$$

One feature of the shift produced in this way is that it is obligatory. The embedded clause must be evaluated at the shifted parameter (at least if no other operators are present). This point will be important for the comparison between the monstrous theory and more classical theories.

So much for an overview. Now let me introduce the system.

3.3 Semantics: formal details

In line with recent literature (see, for example, [Schlenker 2003](#)), I assume a fully extensional system, which involves object language quantification over worlds. The basic denotations of clauses are *centered propositions*, namely functions from worlds and individuals to truth-values. For example:

$$\llbracket \text{Ralph loves Ortcutt} \rrbracket^{c,g,a} = \lambda w. \lambda x. \text{Ralph loves Ortcutt in } w$$

For shorthand, I'm going to use '*i*' for a variable over centered worlds as well as for the semantic type of centered worlds.

3.3.1 Indexicals

Indexicals are variables with a special diacritic $[+c^*]$.⁹ Their semantics is analogous to the semantics of ordinary variables of the corresponding types, with the only difference that they are handled by the new assignment a . For example:

$$\begin{aligned} \llbracket \mathbf{I} \rrbracket^{c,g,a} &= \llbracket x_{1[+c^*]} \rrbracket^{c,g,a} = a(1[+c^*]) \\ \llbracket \text{you} \rrbracket^{c,g,a} &= \llbracket x_{5[+c^*]} \rrbracket^{c,g,a} = a(5[+c^*]) \end{aligned}$$

This makes indexicals bindable, but poses the problem of recapturing the connection between their free occurrences and the context of utterance. How to do this?

I suggest that this connection is independent of the compositional part of the semantics. Rather, it is captured at the level that Kaplan calls ‘metasemantics’, i.e. at the level at which the context of utterance determines an assignment. For present purposes, I simply assume the existence of a constraint on the selection of the special assignment a . This constraint rules out as inadmissible assignments which send indexicals into the wrong referents. I.e.,

An assignment a is admissible in c just in case $a(1[+c^*]) =$ the speaker of c , $a(5[+c^*]) =$ the addressee of c , etc.

Admittedly this account is somewhat stipulative: it would be preferable to provide some independent evidence for the metasemantic story. I am not in the position to do this here. On the upside, this is the only stipulative element I need.

3.3.2 Composition rules

Ordinary clauses denote centered propositions; at the same time, modals require a clausal argument of higher type. So there is systematic type mismatch between modals and their clausal arguments. This mismatch is fixed via a new composition rule which tells us to perform lambda-abstraction on the assignment a :

Monstrous Functional Application (MFA)

If α is a branching node and $\{\beta, \gamma\}$ the set of its daughters, then for any context c and assignment g , if $\llbracket \beta \rrbracket^{c,g,a}$ is a function whose domain contains $\lambda a'$. $\llbracket \gamma \rrbracket^{c,g,a'}$, then $\llbracket \alpha \rrbracket^{c,g,a} = \llbracket \beta \rrbracket^{c,g,a}(\lambda a'. \llbracket \gamma \rrbracket^{c,g,a'})$.

Notice that this is a second parallel between this system and intensional systems. Intensional systems normally contain an extra composition rule, Intensional Functional

⁹ The semantics for indexicals is indebted to an idea found in [Schlenker 2008](#), esp. pages 25–6. Schlenker attributes the crucial suggestion to Ede Zimmermann.

Modals are monsters

Application (see Heim & Kratzer 1998), which is also triggered by type mismatch between operators and their clausal argument.

3.3.3 Modals

My starting point is the semantics for epistemic and doxastic modals that is common currency in the literature on *de se* reports (see, among others, Lewis 1979, Chierchia 1989, Anand 2006). Modals are universal quantifiers over a contextually supplied set of centered worlds:

$$\llbracket \text{must} \rrbracket^{c,g} = \lambda p_{\langle i,t \rangle}. \lambda M_{\langle i, \langle i,t \rangle \rangle}. \lambda w. \forall i': i' \in ((M)(i)), p(i') = 1$$

Monstrous modals are different in three respects. I already mentioned two:

- they quantify over pairs of assignments and centered worlds;
- their clausal argument is, correspondingly, a function from assignment-centered world pairs to truth-values.

To these I add a third:

- epistemic and doxastic modals have an extra argument, which is filled by a variable ranging over sequences of counterpart functions.

This extra argument serves the purpose of determining the assignment-centered world pairs that the modal quantifies over. These counterpart functions (type $\langle es, e \rangle$) specify a way in which the subject of the epistemic state is acquainted with elements of the context. So f_1 specifies the way in which she's acquainted with the speaker, f_2 the way in which she's acquainted with the addressee, f_i the way in which she's acquainted with the i -th *demonstratum*, etc. We use them to determine sequences of individuals; in turn, these sequences are used to determine an assignment.

Here are the details. We start from a sequence $\langle f_1, f_2, \dots \rangle$. We plug each centered world in the modal base into all the functions in the sequence. This gives us sequences of possible individuals:

$$\begin{aligned} &\langle f_1(i_k), f_2(i_k), \dots, f_n(i_k) \rangle \\ &\langle f_1(i_l), f_2(i_l), \dots, f_n(i_l) \rangle \\ &\dots \end{aligned}$$

Notice that the individuals in each sequence are the counterparts of speaker, addressee, etc. in a determinate centered world. It is easy to recover an assignment function, once we have each of these sequences. We just take a function which maps the n -th index into the individual determined by the n -th counterpart function:

$$\begin{aligned}
1[+c^*] &\longrightarrow f_1(i_k) \\
2[+c^*] &\longrightarrow f_2(i_k) \\
\dots & \\
n[+c^*] &\longrightarrow f_n(i_k)
\end{aligned}$$

Now, let a_k be the assignment determined in this way by using centered world i_k . The assignment-centered world pairs that a modal quantifies over will be all the pairs $\langle a_k, i_k \rangle$, where i_k is a world in the modal base. For shorthand, I will say that an assignment-centered world pair is *compatible with a modal base M and a sequence of counterpart functions* $\langle f_1, f_2, \dots \rangle$ iff it can be generated from that modal base and that sequence in the way described.

All this is encoded in the lexical entry of modals. In essence, modals do two things:

- First, they ‘check’ that the counterpart functions provided by context are the right ones. I.e., they check that f_1 , if defined,¹⁰ does in fact specify the way that the subject is acquainted with the speaker, f_2 , if defined, does specify the way the subject is acquainted with the hearer, etc.
- Second, they quantify over the pairs that are generated in the way described.

Here is the entry for epistemic *must* (using A as a type for assignments):

$$\begin{aligned}
[[\text{must}]]^{c,g,a} &= \lambda \chi_{\langle A, \langle i, t \rangle \rangle}. \lambda \langle f_1, f_2, \dots \rangle. \lambda M_{\langle i, \langle i, t \rangle \rangle}. \lambda i. \\
& (f_1)(i) = (\text{if def.}) a(x_1[+c^*]), (f_2)(i) = (\text{if def.}) a(x_2[+c^*]) \dots \\
& \text{and } \forall \langle a', i' \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle, \chi(a')(i') = 1
\end{aligned}$$

And here is the entry for *believe*:

$$\begin{aligned}
[[\text{believe}]]^{c,g,a} &= \lambda \chi_{\langle A, \langle i, t \rangle \rangle}. \lambda \langle f_1, f_2, \dots \rangle. \lambda M_{\langle i, \langle i, t \rangle \rangle}. \lambda x. \lambda i. \\
& (f_1)(\langle w, x \rangle) = (\text{if def.}) a(x_1[+c^*]), (f_2)(\langle w, x \rangle) = (\text{if def.}) a(x_2[+c^*]) \\
& \dots \\
& \text{and } \forall \langle a', i' \rangle \text{ compatible with what } x \text{ believes and } \langle f_1, f_2, \dots \rangle, \\
& \chi(a')(i') = 1
\end{aligned}$$

¹⁰ Why the definedness condition? In most cases, the subject of the relevant information state won’t have a way of identifying all elements of the context of utterance. In this case I assume that the context will provide an undefined function for that object. Having undefined functions in the sequence doesn’t make a difference unless the function does actual work in fixing the range of an indexical appearing in the clause.

Modals are monsters

Here is a sample derivation. Take

(17) I must be in Main Library, Stanford.

which I assume has LF

(18) $\lambda i_0. \mathbf{must} [\lambda i_1. x_{1[+c^*]} \mathbf{be in Main Library, Stanford } i_1] (\langle f_1, f_2, \dots \rangle)$
 $(M_{\langle i, \langle i, t \rangle \rangle}) i_0$

Here is the computation (using ‘F’ as a variable over sequences of counterpart functions):

$\llbracket \lambda i_0. \mathbf{must} [\lambda i_1. x_{1[+c^*]} \mathbf{be in Main Library, Stanford } i_1] (\mathbb{F}) (M) i_0 \rrbracket^{c, g, a} =$
 (by FA and abstraction rules)

$\lambda i'. \llbracket \llbracket \mathbf{must} [\lambda i_1. x_{1[+c^*]} \mathbf{be in Main Library, Stanford } i_1] \rrbracket^{c, g[i_0 \rightarrow i'], a} (g(\mathbb{F})) (g(M))$
 $i' \rrbracket =$ (by MFA)

$\lambda i'. \llbracket \llbracket \mathbf{must} \rrbracket^{c, g[i_0 \rightarrow i'], a} (\lambda a'. \llbracket [\lambda i_1. x_{1[+c^*]} \mathbf{be in Main Library, Stanford } i_1] \rrbracket^{c, g[i_0 \rightarrow i'], a'})$
 $(g(\mathbb{F})) (g(M)) i' \rrbracket =$
 (computing the denotation of the clause)

$\lambda i'. \llbracket \llbracket \mathbf{must} \rrbracket^{c, g[i_0 \rightarrow i'], a} (\lambda a'. [\lambda i''. [a'(1[+c^*]) \text{ is in Main Library, Stanford in } i'']]$
 $(g(\mathbb{F})) (g(M)) i' \rrbracket =$
 (by the entry for **must**)

$\lambda i'. \llbracket [\lambda \chi. \lambda \langle f_1, f_2, \dots \rangle. \lambda M. \lambda i. (f_1)(i) = (\text{if defined}) a(x_{1[+c^*]}), (f_2)(i) = (\text{if}$
 $\text{defined}) a(x_{2[+c^*]}) \dots$
 $\text{and } \forall \langle a^*, i^* \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle, \chi(a^*)(i^*) = 1]$
 $(\lambda a'. [\lambda i''. [a'(1[+c^*]) \text{ is in Main Library, Stanford in } i'']] (g(\mathbb{F})) (g(M)) i' \rrbracket =$
 (by the definition of λ -notation)

$\lambda i'. \llbracket [\lambda \langle f_1, f_2, \dots \rangle. \lambda M. \lambda i. (f_1)(i) = (\text{if defined}) a(x_{1[+c^*]}), (f_2)(i) = (\text{if defined})$
 $a(x_{2[+c^*]}) \dots \text{ and } \forall \langle a^*, i^* \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle, a^*(1[+c^*]) \text{ is}$
 $\text{in Main Library, Stanford in } i^*] g(\mathbb{F}) (g(M)) i' \rrbracket =$ (by the definition of λ -notation
 again)

$\lambda i'. \llbracket (g(f_1))(i') = (\text{if defined}) a(x_{1[+c^*]}), (g(f_2))(i) = (\text{if defined}) a(x_{2[+c^*]}) \dots \text{ and}$
 $\forall \langle a^*, i^* \rangle \text{ compatible with } (g(M))(i') \text{ and } g(\mathbb{F}), a^*(1[+c^*]) \text{ is in Main Library, Stan-}$
 $\text{ford in } i^* \rrbracket =$ (fixing the value of g and a , and simplifying)

$\lambda i'. \llbracket (g(f_1))(i'), \text{ if defined, is the speaker and for all } \langle a^*, i^* \rangle \text{ compatible with what}$
 $\text{the speaker knows at } i' \text{ and the sequence of counterpart functions } g(\mathbb{F}), a^*(1[+c^*]) \rrbracket$
 $\text{is in Main Library, Stanford in } i^*$

3.4 The puzzle, solved

Here are the (slightly simplified) truth conditions assigned to (4) and (5):

$\llbracket(4)\rrbracket^{c,g,a} = \text{true}$ iff $(g(f_1))(i_c) = \text{the speaker of } c$ and for all i' compatible with what the speaker knows at i_c s.t. the coin landed tails in i' , the $g(f_1)$ -counterpart of the speaker in i' is in Main Library, Stanford in i' .

$\llbracket(5)\rrbracket^{c,g,a} = \text{true}$ iff $(g(f_1))(i_c) = \text{the speaker of } c$ and for all i' compatible with what the speaker knows at i_c s.t. the coin landed heads in i' , the $g(f_1)$ -counterpart of the speaker in i' is in Widener Library, Harvard in i' .

Provided that the context supplies an appropriate individual concept, this predicts that both (4) and (5) are true, exactly as we want. Although I won't go through that, it should be clear that the semantics is equipped also to handle more traditional puzzles, like the one generated by (1) and (2).

4 The comparison with a classical account

4.1 Kaplan-style semantics for the *de re*

If the foregoing is right, there is a monstrous semantics for English indexicals that accounts for their interaction with *might*, *believe* and the like. But is it better than alternatives? In this section I set up a comparison between the monstrous theory and what I take to be its main competitor, i.e. the semantics for attitude reports deriving from Kaplan 1968. I focus on a recent version of Kaplan's semantics proposed by Percus & Sauerland (2003) (see also Percus 2010). I argue that the monstrous theory has an empirical advantage: the Kaplan-style theory, but not the monstrous theory, overgenerates systematically when attitude reports are embedded in attitude reports.

The monstrous theory and the Kaplan-style theory share two main features. First, they both employ relations of counterparthood by acquaintance. Second, the Kaplan-style theory, somewhat similarly to the monstrous theory, treats *believe* as a quantifier over pairs of sequences of individuals and worlds. As a result, the truth-conditions predicted by the two are, at least for a wide range of cases, very close. So the difference lies mostly in the compositional implementation.

As is well-known, most versions of the Kaplan-style theory rely on important stipulations just at the compositional level. (For example, the one presented by Cresswell & von Stechow (1982) is non-compositional, and the one used by Heim (1994) and Anand (2006) must stipulate a kind of otherwise unattested syntactic movement.) Percus and Sauerland's version is smoother and avoids stipulations of this kind, hence it seems a good candidate for setting up a comparison.

4.2 Percus and Sauerland's semantics for attitude reports

The central notion of Percus and Sauerland's (henceforth, P&S) semantics is that of a *concept generator*. Concept generators are functions matching objects with counterpart functions (type $\langle e, ie \rangle$). The basic idea is that indexicals feed their semantic values into concept generators; the latter produce counterpart functions, which in turn generate descriptive truth-conditions for the sentence.

Here is how this is implemented. The syntax contains variables G_i ranging over concept generators, as well as abstractors over these variables. The clausal arguments of attitude verbs are just functions from concept generators to centered worlds propositions. For example, the denotation of the that-clause in (1) (repeated below) is given in (20)

(19) I believe that you are on fire.

(20) $\lambda G_0. \lambda i_1. [((G_0(\text{the addressee of } c))(i_1)) (\text{is on fire})](i_1)$

The denotation of *believe* involves reference to *acquaintance-based* concept generators: intuitively, these are simply concept generators that match objects with concepts that capture relations of acquaintance.¹¹ P&S's idea is that *believe* existentially quantifies over concept generators. Here is its denotation:

$$\llbracket \text{believe} \rrbracket^{c,s} = \lambda \phi_{\langle \langle e, ie \rangle, \langle i, t \rangle \rangle}. \lambda x_e. \lambda i: \text{There is some acquaintance-based concept generator } G \text{ for } x \text{ in } i \text{ s.t. } \forall i' \in \text{DOX}_x \text{ in } i, (\phi(G))(i') = 1$$

This generates plausible truth-conditions for (1):

$$\llbracket (1) \rrbracket^{c,s} = \text{true iff there is some acquaintance-based concept generator } G \text{ for the speaker of } c \text{ in } i_c \text{ s.t. } \forall i' \text{ compatible with what the speaker of } c \text{ believes, } ((\text{the addressee of } c)(G))(i') \text{ is on fire in } i'.$$

Even though P&S don't investigate epistemic modals, it is straightforward how to extend the proposal. We can use the following lexical entry for *must*:

¹¹ Here is the definition:

A function G of type $\langle e, ie \rangle$ is an acquaintance-based concept generator for x at centered world i iff:

- (i) $\text{Dom}(G) = \{y: x \text{ is acquainted with } y \text{ at } i\}$
- (ii) For all y in $\text{Dom}(G)$, there is some acquaintance relation R s.t.
 - x bears R uniquely to y in i , and
 - for all centered worlds i' in x 's doxastic set at i , the center of i' bears R to $G(y)(i')$ in i' .

$\llbracket \text{must} \rrbracket^{c:s} = \lambda \phi_{\langle \langle e, ie \rangle, \langle i, it \rangle \rangle} \cdot \lambda M_{\langle i, it \rangle} \cdot \lambda i$: There is some acquaintance-based concept generator G for the individual at i s.t. $\forall i' \in M(i)$ in i , $(\phi(G))(i') = 1$

This solves our puzzle. As an example, here are the truth-conditions of (4):

$\llbracket (4) \rrbracket^{c:s} = \text{true}$ iff there is some acquaintance-based concept generator G for the speaker of c in i_c s.t. $\forall i'$ compatible with what the speaker knows at i_c and s.t. the coin landed tails in i' , $((\text{the speaker of } c)(G))(i')$ is in Main Library, Stanford in i' .

4.3 The argument from embedded reports

Despite their similarities, P&S's theory and the monstrous theory diverge in an important respect. P&S's theory introduces counterpart functions via variables that appear in the clausal complements of modals. By contrast, on the monstrous theory counterpart functions are arguments of the modal itself. This difference has an interesting consequence. Consider a sentence with two doxastic modals, like

(21) John believes that I believe that you are on fire.

The monstrous theory predicts that the range of a shifted indexical is invariably fixed by the closest modal. For example, the range of *you* in (21) must be fixed by the lower *believe*. By contrast P&S's theory is more flexible: the range of *you* will depend on the coindexing patterns of concept generator variables.

The argument I'm about to give is based just on this difference. P&S's theory predicts more readings than the monstrous theory. But these readings are unavailable, so P&S's theory overgenerates while the monstrous theory is empirically adequate. In essence, the kind of contrast at stake is the contrast between an extensional and an intensional treatment of a parameter—in this case, counterpart functions. P&S's theory handles them via variables in the clause, while the monstrous theory handles them via an operator that affects everything falling in its scope (i.e., the monstrous modal). The argument establishes that the extensional theory overgenerates and the operator theory is correct.

Start by considering (21). P&S's theory assigns to it this LF:

λi_0 . John believes that
 λG_0 . λi_1 . $(G_0(I))(i_1)$ believe that
 λG_1 . λi_2 . $((G_k(\text{you}))(i_2))$ (are on fire)(i_2)
 $(i_1)(i_0)$

Modals are monsters

The crucial issue is the index of the underlined variable. That variable may have index 0: i.e. it may be coindexed with the *higher* abstractor. If this is the case (21) gets (roughly) the following truth-conditions: it is true iff (a) there is some acquaintance-based way in which John thinks of the speaker (say, *the F*), (b) according to John's beliefs, the F believes that *the individual that the F thinks of in the way in which John thinks of the hearer* is on fire. And this is obviously wrong. To see this, consider the following scenario:

John has seen you, an amnesiac, wander around Main Library, Stanford, and has come to think of you as *the individual I saw wander around Main Library, Stanford*. He also believes that I have seen some other amnesiac, different from you, wander around the library, that I've come to think of him under the description that I would specify as *the individual I saw wander around Main Library, Stanford* and that I believe that he is on fire.

(21) is predicted to be true in the scenario described. But this is obviously wrong: my having beliefs that involve a counterpart function that is also involved in John's beliefs is not enough to give me a belief about you.

By contrast, the monstrous semantics handles (21) quite easily. Here is the LF of (21):

John believes that

[λi_1 . I believe that

[λi_2 . you are on fire (i_2)]

$((f'_1, f'_2, \dots))(M_1)(i_1)((f_1, f_2, \dots))(M_0)(i_0)$

What the semantics does is shift the assignment twice over. So *you* will be evaluated at the assignments quantified over by the lower *believe*. Schematically:

$\llbracket (21) \rrbracket^{c,g,a} = \text{for all } \langle a', i' \rangle \text{ compatible with John's beliefs at } i_c, \text{ for all } \langle a'', i'' \rangle \text{ compatible with the beliefs of } a'(1[+c^*]) \text{ at } i', \llbracket \text{you are on fire} \rrbracket^{c,g,a''}(i'')$

The truth-conditions will be:

$\llbracket (21) \rrbracket^{c,g,a} = \text{true iff John is acquainted in the way specified by } g(f_1) \text{ with the actual speaker in } w_c, \text{ and for all } i' \text{ compatible with John's beliefs at } i_c, \text{ the } g(f_1)\text{-counterpart of the speaker in } i' \text{ is acquainted in the way specified by } g(f'_2) \text{ with the } g(f_2)\text{-counterpart of the addressee in } i' \text{ and for all } i'' \text{ compatible with what the } g(f_1)\text{-counterpart}$

of the speaker in i' believes in i' , the $g(f'_2)$ -counterpart of the addressee in i'' is on fire in i'' .

which seems correct. So the monstrous theory avoids overgeneration.¹²

5 Conclusion

I have argued that there is a monstrous theory that accounts smoothly for the interaction between indexicals and non-metaphysical modalities. This theory is well-equipped to handle traditional puzzles in the semantics for belief, as well as a new

¹² Percus (p.c.) suggests a reply on behalf of the P&S theory. We can assume that a concept generator for an individual x at a world w only yields counterpart functions that are defined at the doxastic alternatives of x at w . If we do this, then (21) ends up having undefined truth-conditions when the lower variable is coindexed with the higher abstractor. This is because, assuming that John and the speaker have doxastic sets that are at least partially disjoint, $(G_k(\text{the addressee of } c))$ is undefined for some i compatible with John's beliefs. So the problematic reading is blocked.

The reply is that one can construct similar examples where the two believers have coinciding belief states. In these cases the problematic reading is not blocked, and the theory still yields a wrong prediction. Here is one example of this kind:

Lingens and Lauben are amnesiacs in two different though qualitatively identical libraries. Adjust things in a way that their doxastic states (i.e. the sets of centered worlds that are compatible with their respective beliefs) are completely identical: so e.g. assume that they have identical subjective experiences. Moreover, make it the case that they're aware that they are in qualitatively indistinguishable situations. (Say, the experimenters tell them that this is the case.) And now, suppose that I am in Lingens's library, that Lingens sees me, and comes to think of me as *the man I saw in the library*. And suppose that B.J. Ortcutt (a qualitative duplicate of me) is in Lauben's library, that Lauben sees him, and comes to think of him under the concept *the man I saw in the library*. And now consider the sentence

(22) Lingens believes that Lauben _{i} believes that I am in his _{i} same library.

The case is extremely contrived, but the judgment is still clear. According to Lingens's beliefs, Lauben has never even seen me, hence he certainly does not believe I am in his same library. So (22) is false. But P&S predict that it has a true reading. The relevant LF is, again, the one where the lower variable over concept generators is coindexed with the higher abstractor:

Lingens believes that

$$\lambda G_0. \lambda i_1. (G_0(\text{Lauben}_j))(i_1) \text{ believe that}$$

$$\lambda G_1. \lambda i_2. ((G_0(\mathbf{I}))(i_2)) (\text{am in his}_j \text{ same library})(i_2)$$

$$(i_1)(i_0)$$

The point is the same as for (21). Both Lingens and Lauben think of someone (respectively, me and Ortcutt) in the way specified by the function $\lambda w. \lambda x. \text{the } y \text{ that } x \text{ saw in the library}$. And, as Lingens knows, Lauben thinks, of the individual that he identifies in the same way in which Lingens identifies me, that he's in his same library. On the P&S theory, this is enough to give (22) a true reading.

puzzle that was my starting data. Arguably, this theory is also superior overall to its main competitor, Kaplan-style semantics for attitude reports.

Should we then reject Kaplan-style semantics in favor of the monstrous theory? Not yet. The argument I've given is very partial. At best, I've shown how to give a monstrous theory that handles the behavior of indexicals under non-metaphysical modals. But Kaplan-style semantics has a much broader coverage: it handles the behavior of all referential expressions in modal environments, as well as quantification into these environments. So a full comparison must wait until the present account is extended to handle names, *de re* occurrences of descriptions, and quantification in modal contexts. But what I've said should be enough to establish that the monstrous option deserves to be taken seriously.

Let me close by pointing to three topics that deserve further research: the bearing of the foregoing on the semantics of metaphysical modals; the interaction between epistemic modals and attitude verbs on the monstrous picture; the existence of independent evidence for the metasemantic account of free indexicals.

References

- Anand, Pranav. 2006. *De de se*: MIT dissertation.
- Anand, Pranav & Andrew Nevins. 2004. Shifty operators in changing contexts. In R. B. Young (ed.), *Proceedings from SALT XIV*, Cornell University: CLC Publications.
- Chierchia, Gennaro. 1989. Anaphora and attitudes *de se*. In J. van Benthem R. Bartsch & van Emde Boas (eds.), *Semantics and contextual expression*, Dordrecht, Netherlands: Foris.
- Cresswell, Max J. & Arnim von Stechow. 1982. *De re* belief generalized. *Linguistics and Philosophy* 5(4). 503–535. doi:10.1007/BF00355585.
- Cumming, Sam. 2008. Variabilism. *Philosophical Review* 117(4). 525–554. doi:10.1215/00318108-2008-015.
- Heim, Irene. 1994. Puzzling reflexive pronouns in *de se* reports. Handout for Bielefeld conference, March 1994.
- Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford, UK and Malden, MA: Blackwell Publishers.
- Hintikka, Jaakko. 1962. *Knowledge and belief*. Cornell University Press.
- Hintikka, Jaakko. 1969. On the logic of perception. In *Models for modalities*, Dordrecht-Holland: Reidel.
- Kaplan, David. 1968. Quantifying in. *Synthèse* 19(1-2). 178–214. doi:10.1007/BF00568057.
- Kaplan, David. 1989. Demonstratives. In J. Almog, J. Perry & H. Wettstein (eds.), *Themes from kaplan*, Oxford: OUP.

- Kratzer, Angelika. 1981. The notional category of modality. In H. J. Eikmeyer & H. Rieser (eds.), *Words, worlds, and contexts: New approaches to word semantics*, Berlin: de Gruyter.
- Lewis, David K. 1968. Counterpart theory and quantified modal logic. *Journal of Philosophy* 65(5). 113–126. doi:10.2307/2024555.
- Lewis, David K. 1979. Attitudes *de dicto* and *de se*. *Philosophical Review* 88(4). 513–543. doi:10.2307/2184843.
- Lewis, David K. 1983. Individuation by acquaintance and by stipulation. *Philosophical Review* 92(1). 3–32. doi:10.2307/2184519.
- Ninan, Dilip. 2008. *Imagination, content, and the self*: MIT dissertation.
- Percus, Orin. 2010. The syntax of *de re*. Handout for the Goettingen workshop “A Matter of Attitude”.
- Percus, Orin & Uli Sauerland. 2003. On the LFs of attitude reports. In M. Weisgerber (ed.), *Proceedings of Sinn und Bedeutung 7*, Konstanz: Universitaet Konstanz.
- Quine, Willard V. O. 1956. Quantifiers and propositional attitudes. *Journal of Philosophy* 53(5). 177–187. doi:10.2307/2022451.
- Schlenker, Philippe. 1999. *Propositional attitudes and indexicality (a cross-categorical approach)*: MIT dissertation.
- Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26(1). 29–120. doi:10.1023/A:1022225203544.
- Schlenker, Philippe. 2004. Person and binding (a partial survey). *Italian Journal of Linguistics/Rivista Italiana di Linguistica* 16(1). 155–218.
- Schlenker, Philippe. 2008. Indexicals and *de se* reports. Manuscript available at <http://homepages.nyu.edu/~pds4/>.
- von Stechow, Arnim. 2002. Binding by verbs: Tense, person and mood under attitudes. Unpublished manuscript.
- von Stechow, Arnim. 2003. Feature deletion under semantic binding: Tense, person and mood under verbal quantifiers. In M. Kadawaki & S. Kawahara (eds.), *Proceedings of the 33th Annual Meeting of the North East Linguistic Society (NELS 33)*, .

Paolo Santorio
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
32 Vassar St 32D808
Cambridge, MA 02139 USA
santorio@mit.edu