

Reference and Monstrosity

Paolo Santorio

AUSTRALIAN NATIONAL UNIVERSITY

*My mommy always said there were no monsters—no real ones—but there are.
—from 'Aliens'*

I want to introduce a puzzle about indexicals. The puzzle shows that *I*, *you*, *now*, and the like can fail to pick out elements of the actual circumstances of speech: for example, *I* can fail to refer to the speaker. Rather, indexicals sometimes work as bound variables, despite an entrenched dogma that they can never be shifted or bound. This has some interesting philosophical consequences—among them, that it vindicates a broadly Fregean perspective on referential expressions, refuting the idea that indexicals are rigid designators. Or so I say.

1 Overview

It is a truism that some expressions in language depend on other expressions for their interpretation. Definite descriptions are paradigm examples of this dependence. For example, in

- (1) Juventus was the best soccer team in Europe.

the past tense shifts backward the time at which the rest of the clause is evaluated. Thus the description *the best soccer team in Europe* picks out not the current best soccer team in Europe, but the best soccer team in Europe at some time before the time of utterance. It is less of a truism that other expressions don't conform to this model. In fact, the discovery that some expressions behave differently from *the best soccer team in Europe* was a turning point which shaped contemporary theories of content and reference. Proper names like *Juventus* or *Titus Flavius Vespasianus*, as Kripke (1980) famously pointed out, are expressions of this kind. Indexicals like *I*, *you*, *now*, and *that*, which are my focus in this essay, are another example. Indexicals seem to be insensitive to the effects

of those operators that are able to manipulate descriptions. Here is a telling example from Kaplan (1989):

- (2) It is possible that in Pakistan, in five years, only those who are actually here now are envied.

Actually, here, and now in (2) obviously pick out the world, location, and time of utterance, despite the presence of expressions (call them ‘operators’) that normally shift the world, location, and time at which a clause is evaluated.

The phenomenon displayed by (2) lends support to a simple and elegant theory. Indexicals are directly referential: their semantic contents are simply their referents. This distinguishes them from descriptions like *the best soccer team in Europe*, whose contents are functions from shiftable parameters (world, time, etc.) to truth-values. This semantic difference explains the differences in linguistic behavior. Tense and modal operators can affect the contents of descriptions because the latter are sensitive to the time or world of evaluation. But the same operators are inert with indexicals because their contents make no reference to shiftable parameters.

The claim that indexicals are directly referential is at the heart of contemporary orthodoxy about reference. The orthodox picture is wrong. In this paper I argue that all indexicals can depend for their interpretation on other expressions. More precisely, indexicals are akin to variables of first-order logic. They can be free: in this case they invariably pick out elements of the context, like the speaker, the addressee, or the time of utterance. But they can also be bound: this happens when indexicals occur in the scope of epistemic modals or attitude verbs. One typical example is *might*. I claim that *I* in

- (3) I might be in Chicago.

is a variable bound by *might*. It ranges over *epistemic counterparts* of the speaker, namely individuals who, for all the speaker knows, she might be. So (3) says that there is a possibility in which the speaker’s counterpart in that possibility is in Chicago. Kaplan dubs operators which shift or bind indexicals ‘monsters’ and bans them from the semantics of natural language (or at least, from the semantics of natural languages like English). So one pithy way to put my claim is that epistemic modals and attitude verbs are all Kaplanian monsters.

The ban on monsters is a cornerstone of the received view of context, reference, and semantic theory. Lifting this ban requires major changes to this picture. In this paper I discuss two of them. First, indexicals are not rigid designators: they are semantically sensitive to the possible world at which they are evaluated. This allows indexicals to contribute to truth-conditions something more informative than just their

referents (roughly, something like a Fregean sense (1892/1997)). Interestingly, the denial of rigidity is compatible with recognizing that indexicals differ significantly from descriptions in their linguistic behavior. My proposed account will indeed vindicate many (though not all) of the ideas behind direct reference theory, showing that they can coexist with a semantic notion of cognitive significance. Second, while standardly compositional semantic theories make use of a context parameter, the new framework relegates context outside the scope of compositional semantics proper. In short, compositional semantics doesn't need to look at the context of utterance. Aside from its technical interest, this point has potential consequences for views about linguistic competence and the semantics/pragmatics interface.

Here is a synopsis of the paper. I first introduce a phenomenon, epistemic shift, which challenges the orthodox view of indexicality (section 2). Then I present an account of epistemic shift: the key idea is that indexicals are variables and that they are systematically bound in all sorts of epistemic contexts, e.g. in the scope of *might* and *believe* and in epistemic conditionals (section 3). Then I examine the impact of this view on the general architecture of semantic theories. The new view suggests that the connection between indexicals and context is not part of the compositional semantics, but is rather part of a different level of a theory of meaning, which is sometimes called 'postsemantics' (section 4). I close by surveying some extensions of the theory (section 5) and mentioning some outstanding issues (section 6).

2 Epistemic shift

2.1 Double indexing

As a start, let me survey the view of context dependence that is nowadays standard. Modern work on context dependence begins with the observation that theories of meaning need to track contextual parameters, such as the world or the time of utterance, in two distinct ways. This observation dates back to Kamp (1971) and is implemented formally by using *double indexing*, i.e. by relativizing the interpretation of object language expressions to two kinds of parameter. On the one hand, indexicals directly invoke context in their lexical meaning. For example, in

(4) Juventus is the best soccer team in Europe now.

now directly latches on to the time of utterance. This is captured by relativizing the interpretation function (i.e. the function which maps expressions of a language into their meanings, usually denoted by the double brackets '[[·]]'') to a context parameter. The meaning of indexicals directly mentions that parameter; for example:

$\llbracket \text{now} \rrbracket^c = \text{the time of } c$

On the other hand, the semantics keeps track of contextual coordinates separately, via an n -tuple called *index of evaluation*. The index is needed because certain expressions are sensitive to the time and world at which they are evaluated. For example, the description *the best soccer team in Europe* picks out a different referent at different times and worlds. So the interpretation function is relativized to an index of evaluation besides a context. Index coordinates can be shifted by words, like tenses and modals, that are often referred to as ‘operators.’ For example, in

(1) Juventus was the best soccer team in Europe.

the past tense shifts backwards the time at which the rest of the clause is evaluated:¹

$\llbracket \text{Juventus was the best soccer team in Europe} \rrbracket^t = \text{true}$ iff for some time t' before t , $\llbracket \text{Juventus be the best soccer team in Europe} \rrbracket^{t'} = \text{true}$

The crucial claim is that these two forms of interaction between meaning and context are independent and can’t affect each other. In particular, operators that shift index parameters can never shift indexicals, as example (2) suggests. This offers support to the directly referential view. If, as direct reference theory claims, the contents of indexicals incorporate no reference to times and worlds, then it is immediately predicted that no operators can have semantic influence on them.

So much for the standard view. Now I start building my case against it.

2.2 Shifty conditionals

Consider the following scenario:²

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 doesn’t know which. They will be subjected to the following

¹Modern accounts of tense in semantics actually lean towards a different hypothesis: tenses are explicit quantifiers which bind object-language variables ranging over times. (See, among many others, King (2003) and Kusumoto (2005).) The point is irrelevant for my purposes; I’m using tense just as a convenient illustration of the functioning of operators.

²The examples that follow are original, but the characters are borrowed from classical literature on indexicality and reference. The amnesiac Rudolf Lingens is Perry’s creation (1977) and Gustav Lauben (who also appears in Perry’s paper) can be traced back to Frege (1918/1967). It should be flagged that I’m not the first to notice the ‘shifty’ properties of indicative conditionals: cf. Jackson (1987), Weatherston (2001), Nolan (2003), and Williamson (2006). These philosophers mostly focus on the capacity of indicative conditionals to shift *actually* or natural kind terms (though, interestingly, Nolan has a clear case of epistemic shift involving *that* and briefly considers a view that is very close to my account). The crucial observation is that, as I’m going to argue, *all* indexical reference works in a non-standard way in epistemic contexts.

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:

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

Here is my starting datum: utterances of (5) and (6) are perfectly felicitous. Indeed, they seem to capture exactly what the lost amnesiac knows about his own predicament. If he were asked “Where are you?”, a joint utterance of (5) and (6) would be a good and informative answer. To better appreciate the point, it’s useful to contrast (5) and (6) with the corresponding counterfactuals. Consider:

- (7) If the coin had landed tails, I would have been in Main Library, Stanford.
- (8) If the coin had landed heads, I would have been in Widener Library, Harvard.

For concreteness, suppose that the amnesiac utters (7) and (8) in the context of recapitulating to himself the causal connections between possible outcomes of the coin toss and awakenings in libraries. (For this sort of use of counterfactuals, see Stalnaker (1975).) Now, there is a stark contrast in the acceptability of the two pairs. (7) and (8) are not good things to say in the scenario. They sound false, or at least inappropriate: in any case, they are clearly worse than (5) and (6).

My next observation is that a Kaplanian account of indexicals predicts that at least one of (5) and (6) should be infelicitous, contrary to the data. So (5) and (6) give rise to a puzzle for Kaplan’s theory.

This is the problem, intuitively: Kaplan’s account tells us that all tokens of *I* invariably refer to the speaker. So the two tokens of *I* in (5) and (6) should pick out the same individual. But two different individuals are in the relevant libraries in the two relevant scenarios: if *I* invariably refers to the speaker, it cannot pick out both of them. So, no matter who the speaker is, one of the two sentences should come out false.

Here is a more formal analysis of the difficulty. For illustration, take a standard version of possible worlds semantics for conditionals. All conditionals involve universal quantification over sets of possible worlds. The domain of quantification is provided by the context and the *if*-clause is used to specify a restriction on that domain.³ Condition-

³Syntactically, these truth-conditions are achieved by assuming the existence of an unpronounced modal quantifier taking scope over the whole conditional. The *locus classicus* for this theory is Kratzer (1981) and (1991). Kratzer builds her account on the classical work of Stalnaker (1968) and Lewis (1973);

als like (5) and (6) have epistemic flavor: they quantify over a set of worlds representing an epistemic state. There is controversy about how this set is determined: here I use a simplified account, where the relevant epistemic state is just that of the speaker.⁴ So I take the truth-conditions of a conditional with epistemic flavor to be, schematically:

$$\llbracket \text{if } p, q \rrbracket^{c,i} = \text{true iff for all } w' \text{ compatible with what the speaker of } c \text{ knows at } i \text{ and such that } p \text{ is true in } w', q \text{ is true in } w'$$

Now, fix a context: suppose that Lingens is the speaker. Assuming that *I* invariably denotes the speaker, the truth-conditions of (6) are

$$\llbracket (6) \rrbracket^{c,i} = \text{true iff for all } w' \text{ compatible with Lingens' knowledge that are such that the coin landed heads in } w', \text{ Lingens is in Widener Library in } w'.$$

But now, in all the heads-worlds compatible with Lingens' knowledge, the amnesiac lost in Widener is Lauben and Lingens is dead. So (6) is predicted to be false. Conversely, in contexts where Lauben is the speaker there is a wrong prediction about (5). Notice that the problem is not just that certain utterances get the wrong truth-values. No matter what the context is, one of (5) and (6) is predicted to be false. Then a joint utterance of (5) and (6) should sound contradictory. But clearly it doesn't.

There is an intuitive diagnosis of what happens in (5) and (6): *I* picks out not the actual speaker, but whatever individual is speaking in the circumstances singled out by the antecedent. In short, the referent of *I* seems to shift on the basis of the antecedent of the conditional. It's convenient to have a name for the phenomenon displayed by the (5)–(6) pair. Drawing on this intuitive diagnosis—and without, for the moment, making any commitments to its correctness—let me call it 'epistemic shift'. My main concern in this paper is to explain epistemic shift and draw some general consequences for semantic theory.

2.3 Three quick replies

Let me forestall three quick attempts at accounting for epistemic shift.

Non-possible worlds theories. The first is that the puzzle can be solved simply by switching to a non-possible worlds theory of conditionals. One obvious candidate is the material conditional analysis championed by Lewis (1976) and Jackson (1987).

for a state-of-the-art version of this semantics, see von Stechow & Heim (2010).

⁴Here I'm skirting an ongoing debate in philosophy of language: different schools of thought (contextualism, relativism, expressivism) disagree about the exact mechanisms through which this parameter is determined. (For representative positions, see DeRose (1991), MacFarlane (2008), and Yalcin (2007).) But for the purposes of this puzzle the issue is irrelevant; the puzzle can be reproduced on any of these positions.

This analysis identifies the truth-conditions of indicative conditionals with the truth-conditions of corresponding material conditionals of first-order logic. Interestingly, on this account both (5) and (6) come out true.⁵ So it would seem that endorsing it is sufficient to defuse the puzzle.

This conclusion is too fast. As is well-known, the material conditional analysis declares true a large number of indicative conditionals that are infelicitous. So the semantics needs to be supplemented with an account of the assertability of these conditionals. The puzzle resurfaces at this level. Jackson (1987), for example, claims that an indicative conditional is assertable just in case the speaker assigns a high credence to the consequent, conditional on the antecedent. So, in our scenario, (6) is assertable just in case the speaker has a high conditional credence that Lingens is in Widener, given a heads outcome. But the speaker assigns little or no credence to the proposition that Lingens is in Widener: hence (6) is declared to be unassertable, contrary to fact.

An analogous point holds for no-truth-value theories *à la* Edgington (1995), who analyzes assertions of indicative conditionals as speech acts of conditional assertions. These theories also use conditional probabilities to make predictions about assertability of conditionals. But again, (6) is assertable, even though on standard assumptions about indexicals it is predicted to have low or zero probability. More generally, the puzzle can be recreated on virtually any theory of epistemic conditionals, on three minimal assumptions: the semantics of indexicals is Kaplanian; both conditionals are true, or at least appropriate; the speaker in the scenario has no inconsistent beliefs.

Haecceities. The second attempted solution is condensed in the following line of thought:

Assume that there are non-qualitative individual essences or haecceities.⁶ Suppose also that *I* invariably refers to the individual with the haecceity of the actual speaker. Finally, assume that the individual with the speaker's haecceity has Lingens's qualitative properties in tails-worlds and Lauben's qualitative properties in heads-worlds. Then both (5) and (6) come out true.

The basic idea is that *I* in (5) and (6) invariably refers to one individual, namely the actual speaker. However, that individual (singled out on the basis of his haecceity) has very different qualitative properties in the two relevant scenarios: in tails-scenarios he has Lingens's qualitative properties, in heads-scenarios Lauben's.

⁵Suppose that Lingens is the speaker. Then the coin landed tails and the antecedent of (6) is false at the actual world. On the material conditional account, this is sufficient to make the whole conditional true. Similarly, *mutatis mutandis*, if the speaker is Lauben.

⁶Non-qualitative individual essences, or haecceities, are first introduced in connection with direct reference theory by Kaplan (1975). See also Lewis (1988), pages 220-248 for discussion of the topic.

I agree that this strategy accommodates my original example (though at the cost of buying into a controversial ontology involving non-qualitative essential properties). But the solution doesn't generalize: it's easy to produce examples that resist it. One way to do this is to use sentences involving two indexicals:

Things are as in the original amnesiac scenario, but for one tweak. In case of a tails outcome, Lingens will wake up in front of a mirror. In case of a heads outcome, Lauben will wake up in front of a glass window, with a perfect replica of himself (perhaps a twin, perhaps a clone) on the other side. And now suppose that the waking amnesiac points to the individual in front of him and says:

- (9) If the coin landed tails, I am you.
- (10) If the coin landed heads, I am not you.

Both conditionals are felicitous and, as before, a standard semantics misses the prediction. But here haecceities don't help. If the amnesiac is in fact looking at a mirror, then *I* and *you* both pick out the very same individual, hence only one haecceity is in play. But then (10) is false. If the amnesiac is in fact looking at his replica, then *I* and *you* pick out two distinct individuals, hence different haecceities are called into question. Hence (9) is false.

The haecceitistic route can provide at most a local fix to some of my examples, but not a general one. So I discard it. An account of epistemic shift calls for adjustments in the semantics of indexicals, rather than in our metaphysics.⁷

Vacuity. The third attempt flatly denies that one of the two conditionals is false on the standard semantics. Rather, both are true, even though one is only vacuously true.⁸ Here is a way to reach this conclusion. We claim that, contrary to what I have assumed so far, the surviving amnesiac does know what library he's in. In fact he knows the proposition expressed by *I am here*: this, the move goes, is sufficient to have non-trivial knowledge about where he is. For example, if the amnesiac is Lingens, his knowing the proposition expressed by *I am here* is sufficient to know that he, Lingens, is in Main Library. It follows that Lingens also knows that the coin landed tails, since only in that case he would make it alive to the library. But then (6) is vacuously true, since its antecedent is false in all of Lingens's knowledge worlds.

This is a radical route to take, since claiming that Lingens knows the outcome of the coin toss is obviously a large bullet to bite.⁹ In any case, the suggestion can be dismissed

⁷Thanks to Agustín Rayo for pushing me to consider this objection.

⁸Thanks to an anonymous referee for both mentioning this objection and suggesting the response.

⁹Moreover, the fact that the antecedent of (6) is known to be false raises questions about its felicity.

on independent grounds. The key point is that occurrences of epistemic shift is not limited to conditionals. Rather than (5) and (6), the amnesiac could say simply:

(11) I might be in Main Library, Stanford.

(12) I might be in Widener Library, Harvard.

(11) and (12) are equally problematic for Kaplan's theory. By assumption, the speaker knows that only Lingens could be in Main Library, and only Lauben could be in Widener Library. But then a joint utterance of (11) and (12) is problematic on the assumption that *I* invariably refers to the speaker. The speaker should know that, no matter who he is, one between (11) and (12) is false.¹⁰

2.4 Generalizations: other indexicals, other modals

It's time to chart some empirical ground. So far I have only talked about *I*; now I turn to investigating how far and along what dimensions the puzzle generalizes. As I'm going to point out, epistemic shift is a pervasive phenomenon that concerns all indexicals and all operators with epistemic flavor.

To start with, two quick observations. The first is that epistemic shift generalizes to all indexicals, including demonstratives. The amnesiac scenario is obviously contrived. But this is only because scenarios where speakers are unaware of their own identity are hard to come by. We can construct less far-fetched variants of the puzzle by switching to more prosaic varieties of ignorance. Here is one involving temporal indexicals. Suppose that you fell asleep at noon, and wake up without knowing whether you slept one or two hours. You say:

(13) If I slept one hour, it is now one.

(14) If I slept two hours, it is now two.

The problem, again: orthodox semantics wrongly predicts one of (13) and (14) to be false, no matter what time it is.

The second observation is that shifty readings are absent in conditionals with non-epistemic flavor. I already contrasted (5) and (6) with the corresponding counterfactual-

Stalnaker (1975) points out that indicative conditionals whose antecedents are incompatible with what is commonly known by the speakers are infelicitous. But perhaps the supporter of vacuity can claim that in this case knowledge doesn't iterate: Lingens knows that the coin landed tails, but he doesn't know that he knows.

¹⁰Even though the data involving possibility modals is relevant, I will keep using (5) and (6) as my running examples. I have two reasons. First, conditionals provide clear support for my analysis, suggesting indexical shift as an intuitive diagnosis. Second, conditionals are not amenable to a pragmatic treatment of epistemic shift (see footnote 12). So they make a better case for a semantic analysis.

als. Now let me point out that the contrast extends beyond counterfactuals. Philosophers tend to group all indicative conditionals together. But it is standard in formal semantics to assume that at least some indicative conditionals involving *will* express the same kind of modality that is expressed by counterfactuals.¹¹ Following this assumption, it is unsurprising that, in the amnesiac scenario, we can formulate *will*-conditionals that pattern with counterfactuals. Suppose that, after having been informed about the experiment but before undergoing it, one of the amnesiacs says:

(15) If the coin lands tails, I will be in Main Library, Stanford.

(16) If the coin lands heads, I will be in Widener Library, Harvard.

Like (7) and (8), (15) and (16) are not good utterances in the scenario. Intuitively, (15) sounds true only if the speaker is Lingens, (16) only if the speaker is Lauben. Since the speaker is uncertain of his identity, neither of them is felicitous (given the Gricean requirement that speakers should not assert what they don't have evidence for).

This establishes that epistemic shift concerns specifically epistemic conditionals. Can it generalize beyond conditionals? I already showed that it can be replicated under existential epistemic modals. It's natural to suspect that it can take place under other verbs with epistemic flavor: for example, attitude verbs like *believe*. The issue is not straightforward.

On the one hand, for the moment I cannot produce data showing that epistemic shift obtains under *believe*. On the other, we are familiar since Frege (1892/1997) with the idea that referential expressions, including indexicals, are problematic in belief reports. Here is an illustration: suppose that Lingens mistakes a mirror image of himself for a different individual—perhaps Fidel Castro, who wears a beard similar to his. Then he can truly utter, addressing the individual he's seeing:

(17) I believe that I am lost in a university library.

(18) I believe that you are not lost in a university library.

despite the fact that in the context *I* and *you* pick out the same individual. Using a standard label, indexicals display failure of substitutivity in attitude reports. Even though the specific form of the two puzzles is different, it is natural to suspect that both epistemic shift and failures of substitutivity are symptoms of a unique underlying difficulty concerning verbs and operators that have epistemic flavor.

I endorse the kinship between the two puzzles and look for a unitary solution. My main motivation is that there is a conceptually and empirically uniform account that

¹¹This analysis is substantiated by the morphological claim that *will* and *would* both realize a unique modal auxiliary, usually referred to as *woll*. (For discussion, see, among others, Abusch (1988), Condoravdi (2002), Werner (2006).) Notice that the view still leaves room for some *will*-conditionals to be epistemic.

manages to explain both phenomena. Just the existence of this account provides evidence that the two are closely related. Moreover, there is an empirical connection between the two: it's possible to recreate a restricted kind of epistemic shift under attitude verbs. This connection emerges not with indexicals, but rather with other pronouns, like *she* and *he*: I present the relevant data in section 5.

Epistemic shift also extends beyond indexicals. In fact, it can be reproduced with virtually any kind of referential expression, including names and deictic pronouns. But taking up the task of rewriting the semantics for all these expressions would take me too far. For this reason, throughout the bulk of the paper I'm going to focus on indexicals. I will point out, in section 5, how my account can be used as a blueprint for building parallel semantics for other referential expressions.

2.5 Review

I have introduced epistemic shift by means of epistemic conditionals involving *I*. Then I have observed that the phenomenon generalizes to all indexicals. I have shown that epistemic shift is tied specifically to epistemic conditionals and disappears when we switch to other conditionals. Finally, I have suggested that epistemic shift is part of a broader phenomenon concerning verbs and operators that represent information states.

The fact that epistemic shift is systematically linked with a particular class of expressions is evidence that it is a semantic and not a pragmatic phenomenon. The natural hypothesis is that the semantics of *might*, *must*, *believe* and the like involves some undiscovered feature that is responsible for the shifty behavior of indexicals in their scope. This is the kind of account I pursue in the next sections.¹²

¹²Notice, in particular, that the problem can't be solved by using Stalnaker's (1978) pragmatic diagonalization strategy. This strategy consists in taking the proposition communicated by a sentence *S* in a context to be the so-called diagonal, that is, (roughly) the proposition true at those possibilities *p* where the content expressed by *S* as uttered in *p* is true.

To see that this can't help, consider the diagonal proposition communicated by (6) in the amnesiacs scenario. Since we're considering indexicals, the diagonal is a function from contexts to truth-values (following Lewis (1980), just take contexts to be centered worlds). Hence the diagonal of (6) is

$$\lambda c. \llbracket (6) \rrbracket^{c,i_c} = \lambda c. \text{for all } w' \text{ compatible with the speaker's knowledge in } w_c \text{ that are such that the coin landed heads in } w', \text{ the speaker of } c \text{ is in Widener Library in } w'.$$

Now, let the speaker of the context be Lingens (let the context be $\langle w_T, t, \text{Lingens} \rangle$, where w_T is a tails-world). The centered proposition above is false in this context, since in no world compatible with Lingens's knowledge does Lingens end up in Widener library.

So diagonalization fails. Intuitively, the problem is that (as it happens in the standard Kratzer semantics) the antecedent of the conditional has no semantic effect on *I*. But what we want is just to let the antecedent determine the context in which *I* is evaluated. Hence we would need a kind of diagonalization operation that functions at the embedded level. Taking this route leads essentially in the monstrous direction. Indeed, my account can be seen as a generalization of a semantic version of Stalnaker's diagonalization.

3 Informational modals are monsters

3.1 Preliminaries

Recall the intuitive diagnosis of the puzzle. In conditionals like (5) and (6), *I* picks out not the actual speaker, but rather whatever individual is speaking in the circumstances singled out by the antecedent. In short, the referent of *I* seems to shift on the basis of the antecedent of the conditional. The account I propose endorses and develops this intuition. To be sure, my claim will be that *I* in (5) and (6) is technically a bound variable, hence it's not used to refer to a particular individual. But the intuitive diagnosis captures perfectly the main insight: the antecedents of (5) and (6) have semantic effects on *I*. Similarly for other verbs with epistemic flavor, like *believe* and *might*, and for all indexicals.

My presentation of epistemic shift has been fully framework-neutral. In particular, I made no assumptions about the semantics of conditionals, modality, or belief. To state an account of the phenomenon, I must abandon this neutrality. I help myself to two main assumptions. First, some kind of possible worlds account of epistemic conditionals is correct. Second, attitude verbs like *believe* should be analyzed, in the fashion of Hintikka (1962), as modal operators ranging over possible worlds compatible with the subject's attitudes. I choose these assumptions simply because they capture the standard treatment of conditionals and attitude verbs in compositional semantics. But let me stress that, although I use them to state my account, they are not essential to it. My central claims—that indexicals work as variables, that they are systematically bound in epistemic environments, and that they range over epistemic counterparts of their referents—are still framework-neutral and can be combined with different accounts of conditionals and attitude verbs.

It's useful to have a unique phrase to denote all the expressions I'm concerned with. I use 'informational modals' as a blanket label for the modals which report, describe, or express subjects' attitudes. Typical cases of informational modals are *might*, *believe*, and epistemic *must*; but also verbs like *suppose*, *imagine*, *desire*, or *wish* will be in this class.

This section is devoted to giving my account of informational modals. I do this relatively informally and defer the heaviest technical details to the appendix.

3.2 Variables over counterparts

Epistemic shift and substitutivity puzzles arise in connection with modals that describe (or report, express, etc.) information states, that is, mental states with representational content. The question I've been investigating is how indexical reference works in the scope of these modals. Epistemic shift and substitutivity puzzles show that the answer

suggested by standard theories—that indexicals invariably contribute their referents to truth-conditions—is inadequate. What are the alternatives?

Here is a suggestion. When *I* occurs under an informational modal, it refers not to the actual speaker, but rather to representatives of the actual speaker in the relevant information state. Think of an information state as a set of possible worlds, namely, the worlds that are compatible with the relevant attitude. Now, suppose that the subject of the information state has some attitude about the speaker of a certain context: for example, he thinks that the speaker has a flowing beard. Then in all the worlds compatible with his beliefs there will be a flowingly bearded person representing the speaker. Roughly, these representatives are individuals that, for all the subject knows, the speaker might be. I will elucidate the notion of a representative shortly: before that, let me illustrate the idea via an example.

Take again the library scenario and suppose that Lingens, who is in fact the man lost in the library, meets a second amnesiac, Herman. Lingens tells Herman about the coinflip scenario and Herman expresses sympathy for Lingens's predicament. When they part, Lingens says:

(19) Herman believes that I am in a sad predicament.

Who does *I* in (19) refer to? Upon meeting Lingens, Herman has come to think of him in a number of ways: for example as *the other amnesiac lost in the library* or *the survivor of the gruesome coinflip experiment*. Of course, a number of metaphysically different individuals fit these properties. For example, it is compatible with Herman's knowledge that Lingens is the survivor of the coinflip experiment, and also that Lauben is. I claim that these individuals act as representatives of Lingens in Herman's information state. It is these individuals that the occurrence of *I* in (19) denotes.

Let me clarify two points. First, the talk of *I* denoting a multiplicity of individuals is obviously metaphorical. As I anticipated, the precise statement of the point is that *I* is semantically a bound variable. The individuals representing Lingens in the relevant epistemic state will constitute its range.

Second, the informal talk of representatives can be made precise by switching to talk of epistemic counterparts. In general, a relation of counterparthood is simply a relation of similarity (see Lewis (1968) and (1983)): *x* is a counterpart of *y* under a certain respect just in case *x* and *y* are sufficiently similar in that respect. More specifically, epistemic counterparthood is a three-place relation of similarity (*x* is an epistemic counterpart of *y* for a subject *S*) which captures a way a subject thinks of a certain object. *y*, *z*, . . . are epistemic counterparts of *x* for *S* just in case (a) *S* has beliefs about *x* and (b) *y*, *z*, . . . possess all the properties that *S* attributes to *x*.¹³ Epistemic counterparts are

¹³Notice that Lewis (1983) gives a more specific statement of the notion of epistemic counterparthood. He identifies epistemic counterparthood with counterparthood by acquaintance, where the notion of ac-

generally used, within possible worlds theories of mental content, to ‘locate’ actual world individuals within belief worlds of subjects.

For illustration, consider again the Lingens-Herman example. Model the information in Herman’s belief state as a set of possible worlds, namely the worlds that are compatible with what Herman believes. In all those worlds, Herman meets an individual in the library. In some of those worlds, that individual is Lingens; in others it is Lauben. But all of them have certain features in common: in particular, they all look a certain way, have a certain conversation with Herman, and so on. Those individuals are Lingens’s epistemic counterparts in Herman’s belief state.

In short, then, I suggest that we export the notion of an epistemic counterpart from possible worlds accounts of mental content to possible worlds semantics for epistemic operators. Bound indexicals range over epistemic counterparts of their referent in the actual context. For example, *I* ranges over epistemic counterparts of the actual speaker; *you* ranges over epistemic counterparts of the addressee; and so on.¹⁴

Let me clarify what commitments come with my proposal. I am committed to the claim that a notion of epistemic counterparthood is hard-wired in the semantics of epistemic operators. But I am not committed to any claims about mental representation and mental content. In particular, I am not committed to the idea that the possible worlds model provides the best, or even a good way of representing mental content. There is an obvious analogy, both formal and conceptual, between possible worlds accounts of attitudes and possible worlds semantics for attitude reports. I exploit this analogy, but I can do this without endorsing a possible worlds account of attitudes.

Similarly, I am not committed to a specific way of cashing out the notion of an epistemic counterpart. I have defined epistemic counterparthood by appealing to a notion of a subject having beliefs about a specific object. This notion is obviously, and conveniently, vague. Spelling out this notion requires taking sides in a number of debates about content and *de re* attitudes. For example, it requires choosing between a reductionist and a non-reductionist position about *de re* attitudes. (On the former, *de re* attitudes turn out to be a special kind of self-locating attitudes, while on the latter they don’t: see, among others, Lewis (1983) and Ninan (2008).) Obviously this debate might have an impact on the semantics of epistemic operators; but this impact doesn’t need to be assessed now. Here I state a general semantic framework that can be combined with different views about epistemic counterparthood.

Moreover, I am not committed to a metaphysics that employs counterpart theory

acquaintance captures a kind of direct epistemic relationship between a subject and an object. To get his definition, we should substitute clause (a) of my definition with the clause: (a’) *S* bears the acquaintance relationship *R* to *x* and believes that he is *R*-acquainted with *x*. I prefer my definition because it doesn’t commit me to a specific way of cashing out epistemic counterparthood.

¹⁴Interestingly, Hintikka (1969) makes something like this proposal for the case of names in pre-Kripkean times. His idea, though, seems to have fallen out of consideration by the time of Lewis’s (1983).

to model identity of individuals across possible worlds. I advocate the use of certain counterpart relations to model the range of some bound variables in the semantics. This is fully compatible with a background metaphysics where the same individuals exist at different worlds, as argued by Kripke (1980). The problem of trans-world identity concerns the nature of possibilities. The problem of determining the range of bound indexicals concerns how possibilities—once we've established what they are—should be used to model linguistic content. The two issues are to a large degree independent and I need not take a stance on the former here.

Finally, let me stress an interesting consequence of using counterparts. So far I've made the simplistic assumption that each object is paired with a unique set of representatives in the subject's belief state. But, since at least Quine (1956), philosophers are familiar with cases of double vision, where the same object is individuated by a subject in two different ways. I have already given a case of this kind in section 2.4. Suppose that Lingens mistakes a mirror image of himself for Fidel Castro. Then he thinks of himself in (at least) two separate ways—in one case as the person he takes himself to be, in the other as Castro. The apparatus of counterparts allows me to capture these cases quite easily. The only assumption I need is that sets of counterparts should be paired with pronouns, rather than individuals. For example, in the pair of reports

(17) I believe that I am lost in a university library.

(18) I believe that you are not lost in a university library.

the pronouns *I* and *you* appearing in the scope of the belief operator are paired with different sets of counterparts. I'm going to say more about the pairing of pronouns and counterparts in the next sections, though the main idea is that context is ultimately responsible for it.

This gives a general introduction to the view. Now let me be more specific about the new semantics for epistemic operators.

3.3 Assignment shift

The suggestion that *might*, *believe*, and the like bind indexicals is grounded in a more general shift in the way of thinking of informational modals. On the standard view, informational modals are, in essence, quantifiers over possible worlds. On the view I'm advocating, they also encode in their meaning an apparatus that locates real-world individuals within the set of worlds quantified over. Thus on the new picture these modals manipulate a greater amount of information. The classical picture had them quantify over a set of worlds connected to the actual world via an accessibility relation; on the new picture, they quantify in addition over counterparts of actual individuals in each of the worlds in the set.

The new picture is implemented by letting epistemic operators manipulate the assignment parameter, which I construe here as a sequence of individuals.¹⁵ The general idea behind this is quite natural. When we are characterizing an information state, we cannot use referential expressions to pick out directly elements of the context. We need to ‘locate’ the relevant individuals in the information states. Reference within epistemic contexts is always vicarious reference: it always passes through representatives. As a result, modals don’t act only on the world of evaluation, but rather have an effect on the whole referential apparatus of the language (at least, on the apparatus that handles reference to elements of the context). This apparatus is shifted in such a way that indexicals pick out representatives of the relevant individuals at each world.

Notice that ordinary quantifiers like *every* and *some* are also assignment shifters. But *every* and *some* shift the assignment in a piecemeal way, operating on one variable at a time. Epistemic operators force a shift of the whole assignment at once. In addition to this, they coordinate the shift of the assignment with the shift of the world parameter, in such a way that a world is always paired with the assignment involving epistemic counterparts of individuals at that world. Let me explain.

For illustration, I use a standard semantic framework, where the interpretation function is relativized to a context, a world, and an assignment (though, as will become clear in section 4, the context parameter is superfluous on the new semantics). On orthodox assumptions, the effect of *might* is a shift of the world parameter. This means that the clause in the scope of the modal is evaluated at a different world:

$$\llbracket \text{might } \varphi \rrbracket^{c,w,g} = \text{for some accessible } w', \llbracket \varphi \rrbracket^{c,w',g}$$

On the new semantics, modals shift both the world and the assignment parameters:

$$\llbracket \text{might } \varphi \rrbracket^{c,w,g} = \text{for some accessible world-assignment pair } \langle g', w' \rangle, \llbracket \varphi \rrbracket^{c,w',g'}$$

It should be clear how this lets modals have semantic effects on indexicals. On the new picture indexicals are variables. So their referent is determined by the assignment:

$$\llbracket I \rrbracket^{c,w,g} = \llbracket x_1 \rrbracket^{c,w,g} = g(1)$$

Modals shift the assignment, hence they have the potential for shifting indexicals. (I postpone further details about the semantics for indexicals to section 4.)

This establishes how informational modals function compositionally. But it doesn’t settle what the world-assignment pairs accessible from any given point are. This is a key issue, since without settling it we cannot determine truth-conditions.

¹⁵More often, the assignment is modeled as a *function* from syntactic indices of pronouns to individuals. The two formulations are equivalent (since one can just use the order in the sequence to recover the indices). I choose sequences merely to avoid clumsiness.

It is at this point that relations of epistemic counterparthood come in. Consider first a toy case: pretend that the assignment only settles the value of one variable, namely I . In line with some accounts of belief reports in semantics (for example, Heim (1994)), I assume that context supplies a counterpart function f_I which pairs each world in the information state with the counterpart of the speaker. Then we can specify a set of world-assignment pairs:¹⁶

$$\begin{aligned} &\langle \langle f_I(w_i) \rangle, w_i \rangle \\ &\langle \langle f_I(w_j) \rangle, w_j \rangle \\ &\dots \end{aligned}$$

on which the object in the assignment is the counterpart of the speaker in the world. And now drop the pretense that the assignment only handles one variable: the suggestion is that we can use a whole sequence of counterpart functions to generate assignments starting from worlds. Here is the idea: context supplies us with a sequence of counterpart functions, one for each pronoun in the language: $\langle f_I, f_{you}, \dots, f_n \rangle$. This sequence tracks the way the subject thinks of each element of the context: f_I tracks the way she thinks of the speaker, f_{you} tracks the way she thinks of the hearer, and so on.¹⁷ We use these counterpart functions to ‘project’ assignments out of a world. Here are the sorts of world-assignment pairs we get by following this procedure:

$$\begin{aligned} &\langle \langle \langle f_I(w_i), f_{you}(w_i), \dots, f_n(w_i) \rangle \rangle, w_i \rangle \\ &\langle \langle \langle f_I(w_j), f_{you}(w_j), \dots, f_n(w_j) \rangle \rangle, w_j \rangle \\ &\dots \end{aligned}$$

Notice that this mechanism produces a simultaneous shift of the whole assignment at once. Every object in a shifted assignment is an epistemic counterpart of some object in the original assignment.

The net effect is that epistemic operators work as binders of all indexicals occurring in their scope. They shift the whole apparatus of contextual reference of the language at once. This massive effect explains why epistemic shift is replicated systematically with

¹⁶For reasons having to do with the nature of epistemic counterparthood (see Lewis (1983)), these worlds will actually have to be *centered worlds*. For simplicity, I’m going to ignore this point throughout the paper.

¹⁷Notice that using functions of this kind in the semantics doesn’t require the (dramatically unrealistic) assumption that the subject genuinely has a way of thinking of all objects in the context. If the subject has no way of thinking of some element of the context, we just use ‘dummy’ functions that are everywhere undefined. Hence the context must only provide us with counterpart functions for the objects that the speaker actually has attitudes about.

all indexicals in the language.¹⁸

Finally, let me consider how this machinery accounts for the puzzle. Consider again the problematic (5):

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

Make one plausible assumption: in the case of epistemic conditionals, the counterpart function f_I that determines the range of I picks out the possible individuals that the subject of the information state takes himself to be. I.e., in the case of (5), f_I specifies the epistemic counterparts of the speaker, for the speaker himself.¹⁹ Then the truth-conditions that we get for (5) are:

$\llbracket(5)\rrbracket^{c,w,g} = \text{true}$ iff for all worlds w' compatible with the speaker's knowledge that are such that the coin landed tails in w' , the individual that the speaker takes himself to be in w' is in Main Library, Stanford in w' .

which is exactly right: the individual who's relevant for evaluating the conditional is not the actual speaker, but rather the person who's speaking in the circumstances singled out by the antecedent. Notice that the effect of treating indexicals as variables and letting them range over counterparts is that they contribute to truth-conditions a function, rather than just a referent. This is what solves the puzzle. The account also accommodates in a natural way substitutivity puzzles. I already anticipated the main idea; but, for clarity, consider again

- (17) I believe that I am lost in a university library.
(18) I believe that you are not lost in a university library.

Believe helps itself to different counterpart functions for different indexicals, even if these indexicals corefer. So *I* in (17) and *you* in (18) can be associated to different sets of counterparts, despite the fact that they happen to refer to the same individuals; as a result, (17) and (18) differ in truth-conditions.

¹⁸I discuss more extensively the new semantics for modals in the appendix and in Santorio (2011), which is a technical companion to this paper. Let me point out that my semantics shares a number of features with the semantics that Cumming (2008) gives for names. Despite the similarities, there are important conceptual and technical differences; I briefly discuss them in section 5.1.

¹⁹This assumption is independently plausible. f_I must specify a way the subject of the epistemic state thinks of the actual referent of I . But now, by using the first-person pronoun the speaker is flagging that she is aware that she's achieving self-reference. (Contrast my uttering *My pants are on fire!* with my pointing to the mirror and saying *His pants are on fire!*. In the first case, though not in the second, you can infer, from the pronoun used, that I'm aware that I'm referring to myself.) So it's plausible that, at least in normal cases, the counterpart relation used in an utterance of (5) defaults to one that captures the speaker's *de se* way of thinking of himself.

3.4 Reference without rigidity

Turning modals into monsters challenges the doctrine that indexicals are rigid. It's time to make this claim precise and explore some of its consequences.

The rigidity doctrine has it that, relative to a choice of a context and an assignment, indexicals refer to the same individual with respect to all possible worlds.²⁰ Now, on the picture I suggest, all semantic manipulation of indexicals passes just from the assignment. So rigidity is literally preserved on the new picture. Indexicals and free pronouns do pick out the same individual at all worlds, given an assignment; it's just that, contrary to the orthodox picture, the assignment is sometimes shifted.

But this vindication of rigidity is vacuous. There is substantial failure of rigidity. Epistemic operators shift together, and in a coordinated way, both the world and the assignment parameter. In particular, the value of the assignment parameter is determined as a function of the value of the world parameter. The net effect is that indexicals are sensitive to the world parameter, and their semantic value does change as a result of shifts in the world of evaluation. This is clearly unexpected on the standard picture of indexicality. *Contra* the orthodox view, the referent of *I* can be different at different worlds. Correspondingly, the new picture takes a step in a Fregean direction at the level of truth-conditions. In epistemic contexts, the overall contribution of an indexical to truth-conditions is more informative than its referent. The effect of monstrous semantics is that indexicals are paired with (generally non-constant) functions from worlds to individuals.²¹

Let me emphasize that, despite the denial of rigidity, the new picture doesn't retreat

²⁰I ignore the question of obstinacy, i.e. the question whether indexicals refer to the same individuals even in worlds where those individuals don't exist (see Kaplan (1989), section IV). The question is irrelevant for my purposes.

²¹One might wonder whether this is really a major step. Isn't the abandonment of rigidity already implicit in the use of counterpart theory? If our background metaphysics uses counterpart theory to model cross-world identity, indexicals inevitably refer to different individuals in different worlds. The reply is twofold. First, as I pointed out, a semantics which appeals to epistemic counterparthood is fully compatible with a non-counterpart-theoretic metaphysics. In this case, the denial of rigidity is clearly non-trivial. Indexicals embedded under informational modals pick out different individuals at different worlds, even though the same individuals exist at those worlds. Hence embedded indexicals work very differently from unembedded ones, which keep referring rigidly to the same individuals at all worlds. Second, consider the case in which the background metaphysics is counterpart-theoretic. We can still define a notion of rigidity: following Lewis (1988), a term is *quasi-rigid* (with respect to a context and an assignment) just in case in all worlds it refers to metaphysical counterparts of the individual it refers to in the actual world. Denying quasi-rigidity amounts to claiming that different kinds of counterparts relations (epistemic rather than metaphysical) are employed when indexicals occur under informational modals. This still seems an important point: among other things, it introduces a perspectival component in the semantics of indexicals, since epistemic counterpart relations are relative to a subject's beliefs. But I'm happy to grant that ultimately one should find the denial of quasi-rigidity significant only insofar as one finds the switch from metaphysical to epistemic counterparthood significant.

to descriptivism: it doesn't assimilate the functioning of indexicals to the functioning of descriptions. In essence, the monstrous account separates claims about the compositional semantic values of indexicals from claims about the final truth-conditions of sentences involving indexicals. It agrees with direct reference theory about the former. The compositional semantic values of indexicals and other pronouns differs from the semantic value of descriptions. The two kinds of expressions work in different ways and (insofar as their lexical meanings are concerned) appeal to different kinds of parameters. For example, the lexical meaning of *I* mentions the assignment, while the lexical meaning of a description mentions the world of evaluation:

$$\begin{aligned} \llbracket I \rrbracket^{c,w,g} &= g(1) \\ \llbracket \text{the speaker} \rrbracket^{c,w,g} &= \text{the person speaking at } w \end{aligned}$$

One consequence is that the monstrous account agrees with direct reference theory whenever indexicals are not embedded under epistemic operators. However, the monstrous account diverges from the orthodoxy in that it assumes that the semantics of natural language has the resources to manipulate the two parameters together, in a coordinated way. These extra resources, which are built into the semantics of epistemic operators, make it the case that, at the level of truth-conditions, indexicals can be paired with functions rather than just objects. This is exactly what a broadly Fregean approach would predict. As a result, even though their compositional semantic value is different, *I* and *the speaker* can provide the same truth-conditional contribution (this is arguably the case in (5) and (6)).

One lesson of the foregoing is that a semantic notion of cognitive significance can be combined with the basic ideas behind direct reference theory. The monstrous account exemplifies this combination. Independently of whether it is successful, it's significant that a view of this kind is available.

4 Monstrous semantic theory

4.1 Semantics, context, and content

Section 3 presented a semantics that can predict and explain epistemic shift. But this doesn't exhaust the task I have undertaken in this paper. Indexical reference is one of the cornerstones of a theory of context dependence and semantic content. Changing our semantics for indexicals requires changes to the general architecture of this theory; and this, in turn, has potential repercussions for the role we assign to our semantic theory within a broader theory of cognition. In this section I turn to these issues. This will also allow me to complete my account of indexicals, explaining how *I*, *you*, *now*, and the like are anchored to the context of utterance when they occur free.

I already introduced the idea that the semantics of natural language needs double indexing, that is, it needs to keep track of contextual parameters in two different ways. As I said in section 2, orthodox views do this by relativizing the interpretation function both to a context and to an index of evaluation. This version of double indexing is normally combined with a general view of the interactions between context and meaning which is due to Kaplan (1989). Let me illustrate this view in some detail.

Kaplan's key idea is that the distinction between the context parameter and the index corresponds to a distinction between two points of interaction between semantics and context. On the one hand, the semantics accesses contextual information at the lexical level, via the meanings of indexicals. Crucially, this kind of access takes place before compositional interactions. For example, in

(20) I am lost in a university library.

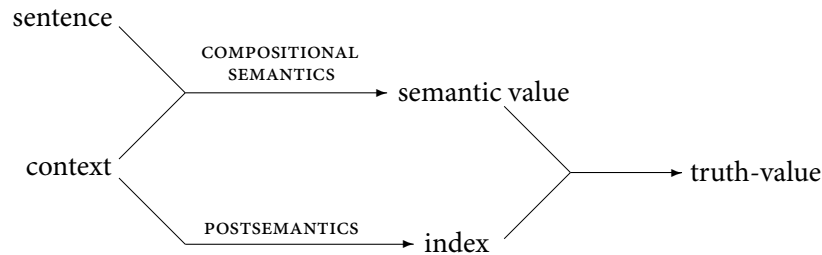
I first 'grabs' its referent, and then feeds it into the process that recursively computes the truth-conditions of the sentence. Hence we need access to contextual information (for example, we need to determine who the speaker is) to determine the input to compositional semantics. We need to look at the context in order to run compositional semantics.

On the other hand, context is also invoked at a different stage. Recall the functioning of operators like tenses and modals. They work by shifting the coordinates at which other expressions are evaluated. For example, in

(1) Juventus was the best soccer team in Europe.

the past tense effects a kind of 'backward shift' of the time at which the clause *Juventus be the best soccer team in Europe* is evaluated. Backward with respect to what? Normally, with respect to the time of utterance. The initial value of the time parameter is set to the time of utterance; it is from that time that the shift effected by operators proceeds. This process of initialization is performed at a different stage of a theory of meaning: following MacFarlane (2003b and 2005), I call this stage 'postsemantics'. Postsemantics is usually placed after compositional semantics, and in any case is independent of it. We don't need to initialize index parameters to carry on a compositional computation of truth-conditions.

Below is a diagram summarizing Kaplan's picture. I use 'semantic value' to denote the kind of meaning that is handled by compositional semantics.



It's worth emphasizing that this general view is not entailed by double indexing; there are alternative ways of implementing the latter. For example, Lewis (1980) points out that we could use functions from contexts and indices to truth-values as the compositional unit. In that case, we would not need to access context at two separate stages. But Kaplan's implementation is often taken as the standard both in the philosophy and the semantics literature.²²

Notice that this general view of context, meaning, and semantics is closely bound up with the prohibition against monsters. The mere setup, in fact, ensures that there cannot be operators that have semantic effects on indexicals. This is easy to see. Kaplan assigns indexicals a referent before they interact compositionally with other expressions. So indexicals can't display the same kind of sensitivity to time, world, or other parameters as, say, descriptions do. The semantic values of descriptions are functions from parameters like times or worlds to referents; but the semantic values of indexicals are just referents: they involve no mention of shiftable parameters. Hence, once the context is fixed, the referents of indexicals are fixed once and for all.^{23,24}

Given that the standard setup stipulates away monsters, monstrous semantics requires rethinking some general aspects of the picture. I do this in two stages. First I focus on indexicals and show how they can be anchored to context on the new view. Then I zoom back out to the big picture, showing how I propose to modify the architecture of the system.

²²An excellent illustration of this is given by the recent literature on truth-relativism: see, among others, MacFarlane (2003a) and (2008), Lasersohn (2005), Stephenson (2007). This literature aims at extending the standard apparatus to a non-orthodox picture: the benchmark view of context, content, and semantic parameters from which this extension proceeds is invariably that proposed by Kaplan in *Demonstratives*.

²³Notice that the formal theory given in *Demonstratives* (section XVIII) doesn't incorporate the prohibition against monsters. But this is just because that theory is not meant to capture in full the philosophical tenets of direct reference theory, as Kaplan himself points out.

²⁴Admittedly, even on the Kaplan picture there could be ways of generating monsters. This would require assuming the presence of special operators that can 'reopen' the slots filled in by indexicals: for example, an operator that, when fed the proposition expressed by 'I am lost in a university library', returns the property *being lost in a university library*. So the claim should be hedged as: on the Kaplan picture, the presence of monsters is a costly stipulation.

4.2 Re-anchoring indexicals

My basic proposal is very simple: indexicals are really just variables. In particular, they are variables with a special marker, which I represent as $[+c]$. For example, *I* is syntactically represented as $x_{1[+c]}$, *you* as $x_{2[+c]}$, etc. Semantically, they work exactly on the model we use for other pronouns. A free pronoun like *she* denotes the object assigned to it by the assignment (for simplicity, I ignore gender):

$$\llbracket \text{she} \rrbracket^g = \llbracket x_5 \rrbracket^g = g(5)$$

Similarly for *I*: its semantic value is just the object assigned to it by the assignment:

$$\llbracket \text{I} \rrbracket^g = \llbracket x_{1[+c]} \rrbracket^g = g(1[+c])$$

The challenge for this proposal, of course, is recapturing the data that originally motivated the standard account. After all, *she* can refer to different individuals within the same context, depending on speakers' intentions or other contextual factors, while unembedded occurrences of *I* are invariably anchored to the speaker. In some way, it is part of the meaning of *I* that it refers to the speaker.

I suggest that the connection between indexicals and context is established not at the compositional stage, but rather at the postsemantic stage. In short, indexicals are anchored to elements of the actual context by the same mechanisms that fix the initial values of index parameters.

Let me get into some detail. Formally, the initialization of index parameters happens via a definition of truth at a context. In the Kaplan framework, this definition says that a sentence is true at a context just in case the semantic value it expresses at that context is true at the index coordinates of the context. For example, taking index coordinates to be worlds and times:²⁵

φ is true at c iff the semantic value of φ at c is true at the world and the time of c

The important observation is that this way of fixing index parameters captures facts that are specifically linguistic. To see this, consider again

(1) Juventus was the best soccer team in Europe.

(1) conveys, as part of its meaning, that Juventus was the best soccer team in Europe at some time before the utterance time. It has no reading on which it says that Juventus is the best soccer team in Europe at (say) some time which precedes some time that the

²⁵See Kaplan (1989), page 522 for an informal definition and page 547 for a formal one.

speaker intends to pick out.²⁶ This fact is captured by letting the initial coordinate of the index be set systematically at the time of the context itself, with no interferences from factors like salience or speakers' intentions.

I claim that the anchoring of indexicals exploits an analogous mechanism. We let the values of indexicals be set to features of the context itself, like the speaker, the hearer, or the time of utterance, via the definition of truth at a context. There are several ways to do this. The simplest one adopts a minimal variant of Kaplan's definition of truth at a context, formulated as follows:

φ is true at c iff the semantic value of φ is true at the world, the time, and the assignment of c

(Notice that this definition, unlike the previous one, uses a notion of semantic value that is not relativized to a context: more on this shortly.) In addition, we specify constraints on what can count as the assignment of the context. This specification will simply list what objects we assign to variables with special indices:

For any context c , an assignment g is the assignment of c only if

$$g(1[+c]) = \text{the speaker of } c$$

$$g(2[+c]) = \text{the addressee of } c$$

...

$$g(5[+c]) = \text{the time of } c$$

...

Notice that, on this setup, the lexical entries of indexicals do not encode all the information that corresponds to their lexical meanings. Lexical entries capture only the compositional component meaning, and indexicals work compositionally as simple variables. This is not an inevitable feature of the monstrous picture. One alternative is to switch to a picture on which the meaning of indexicals involves two independent components. One gets used in the compositional computation in the usual way. The other remains inert at the compositional stage, and is then consulted by the definition of truth at a context to determine the assignment of the context. On this picture, the lexical entry for I would be split into two separate parts:

$$\llbracket I \rrbracket^g = \llbracket x_{1[+c]} \rrbracket^g = g(1[+c]); \{g_c(1[+c]) \text{ is the speaker of } c\}$$

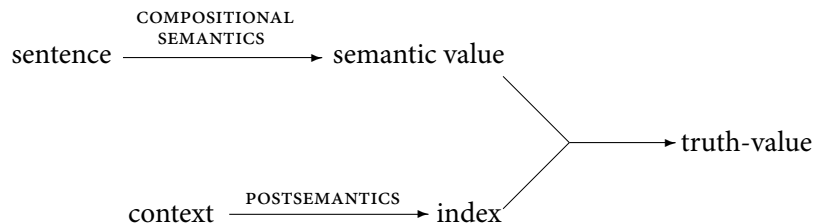
²⁶Of course, setting aside non-standard uses like fictional statements. It's interesting to notice that these cases are paralleled by non-standard uses of indexicals on which the latter do fail to pick out elements of the actual context, like (again) pretense or direct speech reports.

The material in curly brackets is ignored by the compositional computation and is only consulted at the postsemantic level. So far as I can see, this kind of approach is substantially analogous to the one above for present purposes. So I won't take up the task of developing it formally here.²⁷

I want to stress one feature of the postsemantics strategy I have pursued. Despite requiring a *detour* through an extra definition, this strategy exploits mechanisms that are already present in our semantic machinery. I am not introducing new conceptual resources, but rather extending tools that we already have to the case of indexicals. This extension seems appropriate, as on the new picture index parameters and indexicals behave in exactly the same way: both start out anchored to the context and are shifted in the presence of appropriate operators.

4.3 Semantics without the context

How does the new treatment of indexicality modify the architecture of the semantics? The overall effect is an interesting simplification. The interaction between meaning and context happens at a single point, namely at the postsemantic level. No input from context is necessary to run the compositional semantics. So the context parameter becomes superfluous, at least for the compositional part of the theory. We can just run compositional semantics with the index and the assignment.²⁸ The whole theory 'makes contact' with context only at the postsemantic level. This is a diagram representing the new picture:



I'm not claiming that this is the only setup that can accommodate a monstrous semantics. On the contrary, the proliferation of alternatives to the Kaplan system suggests

²⁷For a fully developed semantic framework that implements a somewhat similar idea, see Pott's (2005) extensive study of conventional implicature. As a side observation: notice that, if we went for this option, we would manage to reconstruct something analogous to Kaplan's character into the semantics of indexicals. The main difference would be that, on the new picture, character is only used after compositional interactions.

²⁸Notice that this doesn't mean giving up double-indexing. It is still the case that contextual parameters like time and world are tracked in two distinct ways (though, as I point out in section 3, on the new picture these two ways of tracking contextual coordinates can interact). In this respect, the role of the context is simply taken up by the assignment.

that there might be alternative ways of setting up the general framework that still yield the same predictions about truth-conditions. What I do claim, however, is that this setup is a straightforward and natural way of capturing the functioning of monstrous semantics. Similarly, Kaplan's setup was a straightforward and natural way of capturing the workings of a non-monstrous semantics.

Why is this setup straightforward and natural? Because it puts on the same footing parameters that, on the new account, work fundamentally in the same way. The lesson of sections 2 and 3 is that the functioning of (say) the speaker and the addressee parameters is fully analogous to the functioning of index parameters. To repeat: all these parameters start out anchored to the coordinates of the context and are then shifted by appropriate operators. This setup accommodates the analogy, as it lets all of them be fixed by the postsemantic part of the theory. Formally, we might still be able to revert to a two-tier theory like Kaplan's. But, so far as I can see, this would just be drawing distinctions where there are no significant differences.

The shift of framework is a significant point in itself. But it might also have an impact on questions connecting semantic theory with a general theory of cognition, in particular questions about the border between linguistic competence and general knowledge. These questions fall beyond the scope of this paper, but let me briefly hint at how the new semantics might be relevant for them.

Consider the question: what kind of knowledge is involved in speaking a language? It is broadly agreed that language use (construed in a broad sense, to include all forms of communicative behavior based on language) involves two different kinds of knowledge: on the one hand, linguistic knowledge, which constitutes a specific module of the mind; on the other, general world knowledge, which includes information concerning the context of speech. Syntactic knowledge is knowledge of the former kind; pragmatic knowledge is knowledge of the latter kind. Semantics is a disputed terrain. But Kaplan's picture lends support to one claim: knowledge of the context of speech does mix with linguistic competence to generate meanings for sentences. This just because of indexicals. On the orthodox picture indexicals 'grab' their referents before compositional semantics. Hence we have to use contextual knowledge to settle the referents of indexicals before we go through a compositional computation. On the new picture, this is not the case. Indexicals are assigned referents at an independent stage of a theory of meaning; so compositional semantics can be run without looking at the context. This suggests that the new view of indexicality might lend new support to a modular view, on which compositional semantics has a high degree of autonomy from non-linguistic knowledge.

Of course, this is just a suggestive possibility raised by the new picture. A full assessment of the theoretical consequences of this picture should happen elsewhere.

5 Extensions

Section 4 completes my account. Now I turn to investigating a few natural extensions.

5.1 Names

Consider the following tweak of the amnesiac scenario:²⁹

You and I are watching on a screen a man who's finding his way around a large university library. We decide to dub him 'Herman', and start using this name to talk about him while watching him. Then we are told about the coinflip scenario, and we are informed that the man on screen is the amnesiac who has survived. You say:

- (21) If the coin landed tails, Herman is in Main Library, Stanford.
- (22) If the coin landed heads, Herman is in Widener Library, Harvard.

Unsurprisingly, epistemic shift reappears. (21) and (22) are both true. But if names invariably refer to the object they pick out in actual circumstances, standard accounts miss this prediction. Again, the reason is that there is no single individual in the speaker's epistemic alternatives that is in Main in case of tails and in Widener in case of heads. And again, an intuitive analysis suggests that *Herman* behaves in a shifty way: it picks out whatever individual we named *Herman* in the circumstances individuated by the antecedent.

It's natural to think that we can account for this phenomenon by treating names as variables and letting them be bound by informational modals. This would not be a novelty: recently, Cumming (2008) has suggested just that names work as variables and that attitude verbs are able to bind them via shift of the assignment.

Two qualifications are in order. First, the counterpart relations associated to names do not, in general, work in the same way as those associated to indexicals. In this respect, *Herman* is a rather atypical example. *Herman* is introduced via an explicit stipulation to denote the man you and I are watching on screen (hence its functioning is similar to that of descriptive names; cf. Evans (1985)). Given our background knowledge, the counterpart relation that is most naturally associated to it singles out Lingens in tails-worlds and Lauben in heads-worlds. But now consider a more mundane name, like *Lingens*. *Lingens* is at the center of a complex practice of referring to an individual, Lingens, within a certain linguistic community. (It's not important for present purposes whether this practice involves causal chains or other phenomena.) As a result, one counterpart relation that is made salient by the use of the name is the one picking

²⁹Thanks to Dilip Ninan for suggesting this version of the example.

out the individual named *Lingens* within the relevant community. This difference between the counterpart relations made salient by the use of *Herman* and *Lingens* explains the difference between (23) and (24):

(23) Herman might be Lauben.

(24) Lingens might be Lauben.

In the usual scenario, the speakers know that the individual they name *Herman* might turn out to be *Lingens* or *Lauben*. But it is common knowledge that *Lingens* and *Lauben* denote different people. This explains why (23) is intuitively true and (24) intuitively false.

Let me emphasize, though, that the use of one counterpart relation or the other is ultimately a matter of contextual salience. Different counterpart relations can be employed for different occurrences of the same name, depending on context. To see this, consider

(25) Jason believes that *Lingens* is not *Lingens*.

Jason has convinced himself that the lost amnesiac is *Lauben*. But he's wrong. You know that all along he's been tracking *Lingens*, and *Lauben* is now dead. In this scenario, you can truly utter (25). But of course, the two occurrences of *Lingens* in (25) must be paired with different counterpart relations, lest you want to attribute Jason a contradictory belief.

The second qualification concerns an important difference between Cumming's and my account of assignment shift. Cumming claims that names occurring under informational modals may or may not be bound. By contrast, my account mandates binding of all referential expressions in the scope of these modals. This difference is significant both formally and conceptually: among other things, allowing for non-bound occurrences introduces a distinction between *de re* and *de dicto* readings of names in the scope of informational modals. On Cumming's theory, the *de re/de dicto* distinction is exemplified just by the two occurrences of *Lingens* in (25): the first is *de re* (hence it picks out the actual *Lingens*), the second *de dicto* (hence it picks out *Lingens*'s counterpart in Jason's belief worlds).

A full argument for my position would require extensive discussion, but I can mention two points in its favor. First, the data of section 2 encourage the obligatory binding hypothesis. Shifty conditionals have only one reading: the shifty one. For example, consider again

(21) If the coin landed tails, *Herman* is in Main Library, Stanford.

(22) If the coin landed heads, *Herman* is in Widener Library, Harvard.

There is simply no reading on which one of the two is false (while still having epistemic flavor). This is fully expected on a view like mine, but not on a theory that allows for a *de re/de dicto* ambiguity.

The second point is that, so far as I can see, there is no good way to generalize Cumming's theory to the case of indexicals. We can naturally make sense of a *de re/de dicto* distinction for occurrences of names. Take your favorite metasemantic account of the reference of names: that is, an account of the factors that make it the case that a name refers to a certain object. This account, let's say, claims that a name *n* refers to object *o* in virtue of relation *R* holding between them. (For example, *R* might consist in the existence of the causal chain connecting the two.) Then *de re* occurrences of *n*, which are free variables, are assigned the individual that bears *R* to *o* in the actual world. For example, a free occurrence of *Lingens*—like the first occurrence in (25)—refers to the individual linked to *Lingens* by *R* in the actual world. *De dicto* occurrences, which are bound variables, range over the individuals that bear relation *R* to the name in the worlds within the relevant information state. A bound occurrence of *Lingens*—like the second occurrence in (25)—refers, in each of the relevant worlds, to the individual linked to *Lingens* by *R* in that world. But it's not clear at all how to give an analogous account for the case of indexicals. Neither features of the lexical meaning of indexicals nor features of their metasemantics seem to produce a credible story that covers both cases.³⁰ I have shown how, by contrast, an account based on obligatory binding can accommodate uniformly both names and indexicals.

5.2 Gendered pronouns

Another natural extension of the theory concerns deictic uses of pronouns like *she* and *he*, in two respects. First, there are cases where *she* and *he* seem to pick out different referents depending on the antecedent of an epistemic conditional, as it happens for *I* in (5) and (6). The point should be familiar by now, and I leave it as an exercise to the reader to construct the relevant examples. Second, epistemic shift seems to affect also the functioning of gender in these pronouns. This point is significant because, as I anticipated in section 2, it brings to light a direct connection between epistemic conditionals and belief reports.

Yanovich (2010) has recently pointed out that the meanings of *she* and *he* are indexical in some respect. *she* and *he* invariably specify the actual gender of the individual

³⁰So far as I can see, the only natural suggestion would be to appeal to Kaplanian characters: for example, to let *I* denote whatever individual is speaking in the worlds in the relevant information state. This is essentially what happens in monstrous theories used to account for languages other than English: see, among many others, Schlenker (1999) and (2003) and Anand & Nevins (2004). These theories do in fact allow for both free and bound readings of indexicals under the monstrous verbs. But these theories are, quite obviously, empirically inadequate for English.

they pick out (at least when that individual exists in actuality). To see this, consider these examples:

- (26) If John were a woman, he would be much happier.
 (27) [pointing to a woman] Jason believes that she is a man.
 (28) Tom wishes he were a woman.

Take (26): the gender of the pronoun tracks the actual gender of the individual. The masculine *he* is used to pick out John, despite the fact that in the counterfactual worlds under consideration John is female. Similarly, *mutatis mutandis*, for (27) and (28). This suggests that *she* and *he* involve an indexical component: something in their meaning calls into question facts obtaining in the actual world. It's useful to state explicitly how this is captured in lexical meanings. On standard views, *she* and *he* are analyzed as variables that have a presupposition specifying gender.³¹ For example, *she* denotes the individual assigned to it by the assignment, and presupposes that that individual is female. Making room for the indexical element, the meaning of *she* can be specified as:³²

$$\llbracket \text{she} \rrbracket^{c,i,g} = \llbracket x_5 \rrbracket^{c,i,g} = g(5): g(5) \text{ is female in the world of } c$$

What matters for our purposes is that the pattern exhibited by (26)–(28) is disrupted, once more, in epistemic conditionals.³³ Suppose that I'm talking about a child, Pat, whose gender I don't know. I can say:

- (29) If Pat is a girl, I'll give her a toy bazooka.
 (30) If Pat is a boy, I'll give him a sewing kit.

Once more, we have a kind of epistemic shift. On standard accounts of indexicality, one of (29) and (30) will be infelicitous no matter what the context is, since its presupposition will not be satisfied. Yet the two conditionals are perfectly good sentences.

Here I won't focus on giving a full account of (29) and (30); it should be clear anyway that the monstrous route is a promising way to proceed. Rather, I want to point out how the phenomenon exhibited by (29) and (30) also infects some belief reports, in

³¹The classical presuppositional account of gender is due to Cooper (1983). See the papers in Harbour et al. (2008) for recent literature on the topic.

³²This, of course, is the lexical entry in orthodox frameworks that use a context parameter. In the framework I'm using, it can be reformulated by using a world variable $w@$ that is invariably anchored to the context:

$$\llbracket \text{she} \rrbracket^{i,g} = \llbracket x_5 \rrbracket^{i,g} = g(5): g(5) \text{ is female in } w@$$

³³The data that follow are due to Yanovich, while the observation (in the next paragraph) that the phenomenon generalizes to first-person belief reports is mine.

particular belief reports in the first person. Notice that, rather than (29) and (30), I could say

(31) I believe that either Pat is a girl and I'll give her a toy bazooka or Pat is a boy and I'll give him a sewing kit.

(31) is a good sentence, even though the two pronouns in it cannot both be used appropriately if they are meant to specify the actual gender of the individual. So also in this case there is a kind of epistemic shift. Notice that the fact that (31) is in the first person, rather than the third, plays an important role. Suppose that the speaker, still ignorant about Pat's gender, is reporting Jason's beliefs rather than her own. According to my informants, (32) is better than (33) for this task:

(32) Jason believes that either Pat is a girl and he'll give her or him a toy bazooka, or Pat is a boy and he'll give her or him a sewing kit.

(33) ??Jason believes that either Pat is a girl and he'll give her a toy bazooka, or Pat is a boy and he'll give him a sewing kit.

The contrast between (32) and (33) is important. It shows that the distribution of feminine and masculine pronouns in sentences like (31) is semantically connected to the presence of *believe* (and cannot be explained, say, just by appealing to the properties of disjunction). It matters what worlds are quantified over and how they are related to the speaker's own beliefs: just a difference in the latter factors produces a difference in the acceptability of (31) and (33). So (31) exemplifies a genuine case of epistemic shift and vindicates the idea that we should pursue a unified shifty semantics for epistemic modals and attitude verbs.

5.3 Adverbs of quantification

Nunberg (1993) has pointed out that indexicals can have descriptive-sounding readings when occurring under adverbs of quantification. Suppose that the Pope utters

(34) I am usually Italian. (Recanati (2005), attributed to Nunberg)

(34) is most naturally paraphrased as *the Pope is usually Italian*. Hence the truth-conditional contribution of *I* seems analogous to that of a description. Similarly,

(35) Tomorrow is always the biggest party night of the year. (Nunberg (1993))

has a natural reading on which it says that, for every year, the day falling on a certain date is the biggest party night of that year. Again, *tomorrow* seems to somehow make a truth-conditional contribution analogous to that of a description.

A number of theorists have taken sentences like (34) and (35) as evidence for a kind of descriptivist semantics for indexicals. The most prominent account has been developed informally by Nunberg himself, and has been implemented compositionally (though only for the case of demonstratives) by Elbourne (2008). On Elbourne's account indexicals are syntactically complex entities: crucially, they include a contextually specified descriptive component which takes an argument for situations. This allows them to interact compositionally with adverbs like *always* and *usually*, which, following a longstanding tradition in semantics (see for example Heim (1990) and von Stechow (1994)), Elbourne analyzes just as quantifiers over situations.

It's natural to think that the semantics I've defended should somehow generalize to Nunberg cases. While there are no major technical obstacles, I'm reluctant to take this step at the present stage. The scale and the robustness of Nunberg-type phenomena are still ill-understood; so I'm not in the position to give a theory yielding reliable predictions. In the first place, while (34) and (35) are generally acceptable, they don't sound as good and natural as the data about epistemic modalities that I've introduced in this paper. It's not clear how this asymmetry should be accounted for. Moreover, the productivity of the phenomenon is quite limited; it's easy to find examples where indexicals can't have descriptive readings, even at the cost of making the whole utterance infelicitous. Suppose that, after having survived unscathed a fall from the third floor, I say:

(36) #I guess I was lucky. I usually die!³⁴

The second clause in (36) is obviously infelicitous, even though its structure is analogous to (34) and it's clear what the speaker intends to say. Before giving a theory, it would be important to establish generalizations about the availability of descriptive readings. This task would take me very far from my main argument and is better left to a different occasion.

Nevertheless, it's worth pointing out that the general framework of sections 3 and 4 could easily be extended to Nunberg-type phenomena, if one wanted to do so. Following Elbourne, just assume that adverbs of quantification manipulate a modal parameter, be it a world or a situation. Then their functioning can be assimilated to that of informational modals: we let them shift the assignment and the modal parameter in a coordinated way. The outcome is that indexicals range over counterparts of their referent in the world of utterance, giving rise to descriptive truth-conditions. The resulting theory yields predictions that are basically analogous to Elbourne's theory (and hence the two theories share the overgeneration worries).

³⁴Thanks to Irene Heim for suggesting this example.

The monstrous framework I've presented is general enough to be applied beyond the epistemic domain. But specific constraints and specific aspects of the implementation (for example, the kind of counterpart relation in play under different modals) should be assessed on a case by case basis. Here I'm providing (at best) a general blueprint for an account of indexical shift beyond the epistemic case; substantial work is needed to fill in the details.

5.4 Binding of fake indexicals

It's useful to distinguish the phenomena I've been concerned with from some superficially similar issues. Since Partee (1989) and Heim (1994), it is acknowledged that there are seemingly bound readings of indexicals in certain quantified sentences. The classical example is:

(37) Only I did my homework.

(37) has a reading (the so-called sloppy reading) that can be roughly paraphrased as: *nobody but me did her or his homework*. But if *my* works as a genuine indexical, this reading cannot be derived compositionally. The only reading we get is the so-called strict one, on which (37) says that no other person than the speaker did the speaker's homework. The standard solution (see Kratzer (1998), Heim (2002), and von Stechow (2002) and (2003)) essentially consists in denying that the occurrence of *my* in (37) is a real indexical. Rather, it is a bound variable analogous in all relevant respects to a bound occurrence of *he*. Thus the logical form of (37) is:³⁵

(38) [Only I] $\lambda_i. t_i$ did x_i 's homework.

The bound variable x_i is phonologically realized as a first-person pronoun because of morphological rules of agreement. Essentially, the idea is that the person, gender, and number of a bound pronoun must match the person, gender, and number of the phrase that binds it.³⁶ But semantically the pronoun *my* in (37) has nothing to do with indexicals; the occurrence of something that looks like an indexical in the scope of the quantified phrase *Only I* is a mere morphological accident.

If this account is correct, the phenomenon displayed by (37) is completely distinct from the one I've been analyzing in this paper. I never questioned that the problematic bound pronouns I discussed are genuine indexicals. In fact, my account crucially ex-

³⁵' t_i ' stands for the trace left by the determiner phrase *Only I* after movement. Semantically, this is just another variable.

³⁶These rules can take the form of feature transmission or feature deletion (see Heim (2002) and von Stechow (2002) and (2003)). The difference is irrelevant from a semantic point of view: in both cases, the logical form involves pronouns with features that are uninterpreted semantically.

exploits components of the meaning of *I* to determine the compositional contribution of its bound occurrences. (As I explain in the appendix, when *I* occurs bound, the binding modal checks that the counterpart relation associated to it is the one appropriate for the actual speaker. This process crucially relies on the fact that *I* is anchored to the speaker via the postsemantics.) By contrast, from a semantic point of view, *my* in (37) is simply not an indexical. So (37) is not a case of indexical binding at all.³⁷

One might worry that treating the two phenomena in altogether different ways misses a generalization. But this concern would be misplaced: there is good evidence for holding distinct accounts. First, there are obvious differences in the intuitive truth-conditions associated to the data. The natural paraphrase of

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

involves a descriptive phrase like *the person speaking* in the place of *I*. But, as I just pointed out, in the natural paraphrase of (37) *my* is substituted by a pronoun bound by the quantificational expression *only I*. It seems unlikely that a systematic semantic story could produce analogous truth-conditions for the two cases. Moreover, the shifty phenomena I have investigated seem to appear under a specific class of modal verbs. But the kind of phenomenon displayed by (37) doesn't depend on the presence of a specific operator, not even *only*: it can reappear in virtually any situation in which a pronoun is bound. It's hard to get telling data for the case of person features without using *only*. But the case can be made with number features:

(39) Few men brought their children. (Heim (2002))

their in (39) is a plural pronoun, but it can get a singular reading: the sentence can mean *The number of the x such that x brought his children is small*. It's natural to think that (39) and (37) are instances of the same phenomenon; and indeed, in both cases the Kratzer/Heim/von Stechow analysis gets the right predictions.

In short, a theory of indexical shift should not be stretched to cover cases like (37) and (39). Both the intuitive data and the empirical distribution of the phenomenon look very different in the two cases. At least at the present stage, the prospects for a unified and explanatory theory are dim.

³⁷An anonymous referee points out that this claim should be qualified in view of examples like the following, which the referee attributes to Hotze Rullmann:

(i) Only you prepared a handout for our first meeting.

The referee points out that *our* in (i) should be decomposed into a bound variable and a first-person indexical (hence it works as a 'partially bound' pronoun). For my purposes, the important claim here is that no genuine second person indexical is involved in the decomposition of *our*.

6 Outstanding issues

Let me mention a few questions that, for reasons of space, I can't address in this paper.

Kaplanian alternatives. Throughout the paper, I have concentrated on giving my positive account of epistemic shift. But are there any prospects for giving an account that is more in line with the orthodoxy? Since Kaplan's seminal work on *de re* attitude reports (1968), the semantics literature has tried to account for singular terms in the scope of attitude verbs with the tools he provided. However, all approaches that descend from Kaplan have to rely on important technical stipulations (as is often pointed out by their own proponents).³⁸ I suggest that monstrous semantics is just the way to solve these issues: indeed, it is just the account that Kaplan himself should have given (and in fact shares with it some important features, like the use of epistemic counterparts). I set up a comparison between an up-to-date version of Kaplan's semantics for attitude reports and my account in Santorio (2011).

Metaphysical modals. I've focused on informational modals. But does the new account have an impact on metaphysical modals? One option would be to say 'no' and preserve traditional Kripkean semantics for metaphysical modality. But this seems undesirable. It would come at the cost of ignoring strong evidence that metaphysical and epistemic modals share their semantic architecture. After all, the very same words (*must*, *might*, and *could* are good examples) are often used to express both metaphysical and epistemic modalities. A second, more interesting option is to generalize monstrous semantics to all modals. On this view, all modals would shift the assignment function and let indexicals in their scope range over sets of counterparts of their ordinary referents. The difference in truth-conditions would be explained via a difference in the counterpart relations in play. Epistemic modals would use counterparts by acquaintance, metaphysical modals would use metaphysical counterparts. This view seems worth exploring.

Other monsters. A wave of recent work in semantics (among others Schlenker (2003) and (2010), Anand & Nevins (2004) and Anand (2006)) has shown evidence that languages other than English, for example Amharic and Zazaki, also contain indexical-shifting expressions (though see also von Stechow (2002) and (2003) for an attempt at an alternative, non-monstrous analysis of the data). These monsters are formally quite different from mine and are based on classical work in two-dimensional modal logic. Interestingly, though, the monsters of Amharic and Zazaki are again a kind of informational modal (mostly speech report verbs). I believe there is hope for a general characterization of monstrosity in natural language, on which Schlenker and Anand's monsters turn out to be a special case of my monsters.

³⁸For statements of this sort, see for example Anand (2006), page 26, and Schlenker (2010), page 18.

7 Conclusion

I have argued that, *pace* Kaplan, English turns out to be replete with monsters. *I, you, now*, and the like are systematically bound in the scope of modals like *might* and *believe* and epistemic conditionals. This teaches us interesting facts about the semantics of indexicals and modals, and suggests a shift in the architecture of our semantic theories. One lesson is that many (though not all) of the ideas behind direct reference semantics can coexist with a semantic notion of cognitive significance. Indexicals can contribute to truth-conditions something that is richer than their referents, even though they belong to an altogether different semantic kind from descriptions. A second lesson is that the architecture of our semantic theories should be redesigned: semantics makes contact with context not at the compositional level, but only at the postsemantic level.

I have started my paper with a new empirical puzzle and I have offered my account as a way of solving it. I don't have an argument that monsters are the ultimate or the only way of solving the puzzle: arguments of this sort are hard to come by in empirically driven philosophy of language. But I hope that the interest of the proposal will not depend on whether it provides the ultimate or only account of epistemic shift. The mere fact that there is a monstrous alternative to standard views of indexicality and reference has, I think, an interest of its own. Contrary to common wisdom, there is an alternative way of doing semantics for indexicality. The main goal of this paper has been to convince you that this alternative is worth exploring.³⁹

³⁹For comments, questions, mentorship, and encouragement thanks to Corine Besson, Alex Byrne, Fabrizio Cariani, Jennifer Carr, Andy Egan, Danny Fox, Stavroula Glezakos, Irene Heim, Dilip Ninan, Orin Percus, Alejandro Pérez Carballo, Brian Rabern, Agustín Rayo, Philippe Schlenker, Yael Sharvit, Brad Skow, Bob Stalnaker, Andreas Stokke, Kenny Walden and to audiences at MIT, SPE 2009, the AAP 2009 conference in Melbourne, SALT 20, the Arché Semantics Group in St. Andrews, the 2010 Bellingham Summer Philosophy Conference, and the PhLing workshop at Northwestern University. Special thanks to Irene Heim and Dilip Ninan, whose comments in the early stages of the project helped shape the main ideas, and to Fabrizio Cariani, who gave important feedback when it was most needed.

Appendix: semantics

Setup. The basic functioning of the semantics, including the composition rules, is that given in Heim & Kratzer (1998). In line with classical accounts of the semantics for modality, I assume an intensional system, in which clauses are assigned truth-values relative to an index of evaluation. In line with standard semantics for *de se* attitudes, I assume that indices include at least a world, a time, and an individual parameter. (This allows attitude verbs to manipulate centered worlds.) For simplicity, I represent the index just with the schematic letter ‘*i*’. Because of the reasons pointed out in section 4, there is no context parameter. Thus clauses are assigned a truth-value relative to an index and an assignment:

$$\llbracket \text{Fidel loves Ralph} \rrbracket^{i,g} = 1 \text{ iff Fidel loves Ralph at } i, \text{ relative to assignment } g$$

Indexicals. As I state in the text, indexicals are variables with a special diacritic [+*c*]. Their semantics is analogous to the semantics of ordinary variables of the corresponding types. For example:

$$\llbracket I \rrbracket^{i,g} = \llbracket x_{1[+c]} \rrbracket^{i,g} = g(1[+c])$$

Composition rules. Ordinary clauses denote truth-values, while modals require a clausal argument of higher type. This generates systematic type mismatch. The mismatch is fixed via a new composition rule which enforces lambda-abstraction on the assignment function:

Monstrous Functional Application (MFA)

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

Modals. As is standard in semantics for the *de se* (see Chierchia (1989) and Anand (2006)), the modal base of epistemic modals and attitude verbs consists of a set of centered worlds. Differently from standard semantics for the *de se*, modals quantify over pairs of assignments and centered worlds. Before giving a lexical entry, it’s useful to define the notion of an assignment-centered world pair being compatible with a modal base and a sequence of counterpart functions.

(Notice two assumptions. (a) An assignment is a function from numerical indices to possible individuals. (b) Counterpart functions are matched to variables via their indices, so that for each variable there is a counterpart function with the same index.)

An assignment-centered world pair $\langle g_k, i_k \rangle$ is compatible with a modal base M and a sequence of counterpart functions $\langle f_{n_1}, f_{n_2}, \dots \rangle$ iff

- $i_k \in M$;
- for all indices n , $g_k(n) = f_n(i_k)$

In short, we generate assignment-centered worlds pairs by ‘plugging in’ the centered worlds in the modal base into the sequence of counterpart functions.

Finally, here is the lexical entry for epistemic *must* (using ‘A’ as a type for assignments and ‘ χ ’ as a variable over functions from assignments to functions from indices to truth-values):

$$\begin{aligned} \llbracket \text{must} \rrbracket^{i,g} &= \lambda \chi_{\langle A, \langle i, t \rangle \rangle} \cdot \lambda \psi_{\langle A, \langle i, t \rangle \rangle} \cdot \lambda \langle f_1, f_2, \dots \rangle \cdot \lambda M_{\langle i, \langle i, t \rangle \rangle} \cdot \\ & (f_{1[+c]})(i) = (\text{if defined}) g(x_1[+c]), \\ & (f_{2[+c]})(i) = (\text{if defined}) g(x_2[+c]), \\ & \dots \\ & \text{and } \forall \langle g', i' \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle \text{ and s.t. } \chi(g')(i') = \\ & 1, \psi(g')(i') = 1 \end{aligned}$$

In essence, *must* does two things: first, it ‘checks’ that the counterpart functions provided by context are the right ones. For example, it checks that f_1 , if defined, does in fact specify the way that the subject is acquainted with the speaker. Second, it quantifies over the relevant assignment-centered worlds pairs.

Sample derivation. Finally, I give a compositional derivation of

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

(The derivation of (6) is analogous, *mutatis mutandis*.) I assume that the logical form of (5) is the following:

$$\begin{aligned} \text{MUST [the coin landed tails] } [x_{1[+c]} \text{ is in Main Library, Stanford}] (\langle f_1, f_2, \dots \rangle) \\ (M_{\langle i, \langle i, t \rangle \rangle}) \end{aligned}$$

Three qualifications. First, for simplicity, I treat the name *Main Library, Stanford* as a constant rather than a variable. Second, still for simplicity I ignore tense. Third, I use ‘s’ as a shorthand for $\langle f_1, f_2, \dots \rangle$, which is a sequence of variables ranging over counterpart relations.

Here is the derivation:

$\llbracket \text{MUST} [\text{the coin landed tails}] [x_1[+c]] \text{ is in Main Library, Stanford} \rrbracket (s) (M) \rrbracket^{i,g}$ is true iff
(by Functional Application)

$\llbracket \text{MUST} [\text{the coin landed tails}] [x_1[+c]] \text{ is in Main Library, Stanford} \rrbracket \rrbracket^{i,g}(g(s))(g(M))$ iff
(by Monstrous Functional Application)

$\llbracket \text{MUST} \rrbracket^{i,g} (\lambda g'. \lambda i'. \llbracket \text{the coin landed tails} \rrbracket^{i',g'}) (\lambda g'. \lambda i'. \llbracket x_1[+c] \text{ is in Main Library, Stanford} \rrbracket^{i',g'}) (g(s))(g(M))$ iff (by computing the denotations of the clauses)

$\llbracket \text{MUST} \rrbracket^{i,g} (\lambda g'. \lambda i'. \text{the coin landed tails at } i') (\lambda g'. \lambda i'. g'(1[+c]))$ is in Main Library, Stanford at i' ($g(s)$) ($g(M)$) iff (by the meaning of MUST, and keeping track only of the conditions imposed on the first-person index ' $1[+c]$ ')

$[\lambda \chi. \lambda \psi. \lambda \langle f_1, f_2, \dots \rangle. \lambda M. (f_{1[+c]})(i) = (\text{if defined}) g(x_1[+c]), \text{ and } \forall \langle g', i' \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle \text{ and s.t. } \chi(g')(i') = 1, \psi(g')(i') = 1] (\lambda g'. \lambda i'. \text{the coin landed tails at } i') (\lambda g'. \lambda i'. g'(1[+c]))$ is in Main Library, Stanford at i' ($g(s)$) ($g(M)$) iff (by the definition of λ -notation)

$[\lambda \langle f_1, f_2, \dots \rangle. \lambda M. (f_{1[+c]})(i) = (\text{if defined}) g(x_1[+c]), \text{ and } \forall \langle g', i' \rangle \text{ compatible with } M(i) \text{ and } \langle f_1, f_2, \dots \rangle \text{ and s.t. the coin landed tails at } i, g'(1[+c]) \text{ is in Main Library, Stanford at } i'] (g(s)) (g(M))$ iff (by the definition of λ -notation again)

$(g(f_{1[+c]}))(i) = (\text{if defined}) g(x_1[+c]), \text{ and } \forall \langle g', i' \rangle \text{ compatible with } g(M(i)) \text{ and } (g(s)) \text{ such that the coin landed tails at } i', g'(1[+c]) \text{ is in Main Library, Stanford at } i' \text{ iff (fixing the modal base and simplifying)}$

$(g(f_{1[+c]}))(i) = \text{the speaker and } \forall \langle g', i' \rangle \text{ compatible with what the speaker knows at } i \text{ and the sequence of counterpart relations } (g(s)) \text{ such that the coin landed tails at } i', g'(1[+c]) \text{ is in Main Library, Stanford at } i'.$

References

- Abusch, Dorit (1988). "Sequence of Tense, Intensionality and Scope." In *Proceedings of WCCFL*, vol. 7, pp. 1–14.
- Anand, Pranav (2006). *De De Se*. Ph.D. thesis, MIT.
- Anand, Pranav, and 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 R. Bartsch, J. van Benthem, and van Emde Boas (eds.) *Semantics and Contextual Expression*, Dordrecht, Netherlands: Foris.
- Condoravdi, Cleo (2002). "Temporal Interpretation of Modals: Modals for the Present and for the Past." In D. Beaver, S. Kaufmann, B. Clark, and L. Casillas (eds.) *The Construction of Meaning*, Palo Alto: CSLI Publications.
- Cooper, Robin (1983). *Quantification and Syntactic Theory*. D. Reidel.
- Cumming, Sam (2008). "Variabilism." *Philosophical Review*, 117(4): pp. 525–554.
- DeRose, Keith (1991). "Epistemic Possibilities." *Philosophical Review*, 100(4): pp. 581–605.
- Edgington, Dorothy (1995). "On Conditionals." *Mind*, 104(414): pp. 235–329.
- Elbourne, Paul (2008). "Demonstratives as Individual Concepts." *Linguistics and Philosophy*, 31(4): pp. 409–466.
- Evans, Gareth (1985). "Reference and Contingency." In *Collected Papers*, Oxford: OUP.
- von Fintel, Kai (1994). *Restrictions on Quantifier Domains*. Ph.D. thesis, University of Massachusetts, Amherst.
- von Fintel, Kai, and Irene Heim (2010). "Intensional Semantics." Lecture notes, MIT; available at <http://mit.edu/fintel/IntensionalSemantics.pdf>.
- Frege, Gottlob (1892/1997). "On *Sinn* and *Bedeutung*." In M. Beaney (ed.) *The Frege Reader*, Blackwell Publishers.
- Frege, Gottlob (1918/1967). "The Thought: A Logical Inquiry." In P. F. Strawson (ed.) *Philosophical Logic*, Oxford: OUP.

- Harbour, Daniel, David Adger, and Susana Béjar (2008). *Phi Theory: Phi-Features Across Modules and Interfaces*. Oxford: OUP.
- Heim, Irene (1990). "E-Type Pronouns and Donkey Anaphora." *Linguistics and Philosophy*, 13(2): pp. 137–77.
- Heim, Irene (1994). "Puzzling Reflexive Pronouns in *De Se* Reports." Handout for Bielefeld conference, March 1994.
- Heim, Irene (2002). "Features of Pronouns in Semantics and Morphology." Handout for a talk at USC, January 2002.
- Heim, Irene, and 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, Netherlands: D. Reidel.
- Jackson, Frank (1987). *Conditionals*. Oxford: Basil Blackwell.
- Kamp, Hans (1971). "Formal Properties of 'Now'." *Theoria*, 37: pp. 227–274.
- Kaplan, David (1968). "Quantifying In." *Synthese*, 19(1-2): pp. 178–214.
- Kaplan, David (1975). "How to Russell a Frege-Church." *Journal of Philosophy*, 72(19): pp. 716–729.
- Kaplan, David (1989). "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.) *Themes from Kaplan*, Oxford: OUP.
- King, Jeffrey (2003). "Tense, Modality, and Semantic Values." *Philosophical Perspectives*, 17(1): pp. 195–246.
- Kratzer, Angelika (1981). "The Notional Category of Modality." In H. J. Eikmeyer, and H. Rieser (eds.) *Words, Worlds, and Contexts: New Approaches to Word Semantics*, Berlin: de Gruyter.
- Kratzer, Angelika (1991). "Conditionals." In A. V. Stechow, and D. Wunderlich (eds.) *Semantik: ein internationales Handbuch der zeitgenoessischen Forschung*, Amsterdam: de Gruyter.
- Kratzer, Angelika (1998). "More Structural Analogies Between Pronouns and Tenses." In D. Strolovitch, and A. Lawson (eds.) *Proceedings of SALT VIII*, Ithaca, NY: Cornell University Press.

- Kripke, Saul (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Kusumoto, Kiyomi (2005). "On the Quantification over Time in Natural Language." *Natural Language Semantics*, 13(4): pp. 317–357.
- Laserson, Peter (2005). "Context Dependence, Disagreement, and Predicates of Personal Taste." *Linguistics and Philosophy*, 28(6): pp. 643–86.
- Lewis, David K. (1968). "Counterpart Theory and Quantified Modal Logic." *Journal of Philosophy*, 65(5): pp. 113–126.
- Lewis, David K. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, David K. (1976). "Probabilities of Conditionals and Conditional Probabilities." *Philosophical Review*, 85(3): pp. 297–315.
- Lewis, David K. (1980). "Index, Context and Content." In *Papers in Philosophical Logic*, Oxford: OUP.
- Lewis, David K. (1983). "Individuation by Acquaintance and by Stipulation." *Philosophical Review*, 92: pp. 3–32.
- Lewis, David K. (1988). *On the Plurality of Worlds*. Oxford: OUP.
- MacFarlane, John (2003a). "Future Contingents and Relative Truth." *Philosophical Quarterly*, 53: pp. 321–336.
- MacFarlane, John (2003b). "Three Grades of Truth Relativity." Draft no longer in progress, available at <http://sophos.berkeley.edu/macfarlane/work.html>.
- MacFarlane, John (2005). "Making Sense of Relative Truth." *Proceedings of the Aristotelian Society*, 105: pp. 321–39.
- MacFarlane, John (2008). "Epistemic Modals are Assessment-Sensitive." Forthcoming in *Epistemic Modality*, edited by Brian Weatherson and Andy Egan, OUP.
- Ninan, Dilip (2008). *Imagination, Content, and the Self*. Ph.D. thesis, MIT.
- Nolan, Daniel (2003). "Defending a Possible-Worlds Account of Indicative Conditionals." *Philosophical Studies*, 116(3): pp. 215–269.
- Nunberg, Geoffrey (1993). "Indexicality and Deixis." *Linguistics and Philosophy*, 16(1): pp. 1–43.

- Partee, Barbara (1989). "Binding Implicit Variables in Quantified Contexts." In C. Wiltshire, B. Music, and R. Graczyk (eds.) *Papers from CLS 25*, Chicago: Chicago Linguistics Society.
- Perry, John (1977). "Frege on Demonstratives." *Philosophical Review*, 86: pp. 474–97.
- Potts, Christopher (2005). *The Logic of Conventional Implicature*. Oxford: OUP.
- Quine, Willard V. O. (1956). "Quantifiers and Propositional Attitudes." *Journal of Philosophy*, 53: pp. 177–187.
- Recanati, François (2005). "Deixis and Anaphora." In Z. Gendler Szabó (ed.) *Semantics vs Pragmatics*, Oxford: OUP.
- Santorio, Paolo (2011). "Modals are Monsters: on Indexical Shift in English." In D. Lutz, and N. Li (eds.) *Proceedings of SALT XX*. Ithaca, NY: Cornell University Press.
- Schlenker, Philippe (1999). *Propositional Attitudes and Indexicality (A Cross-Categorical Approach)*. Ph.D. thesis, MIT.
- Schlenker, Philippe (2003). "A Plea for Monsters." *Linguistics and Philosophy*, 26(1): pp. 29–120.
- Schlenker, Philippe (2010). "Indexicals and *De Se* Reports." To appear in the *Handbook of Semantics* edited by von Stechow, Maienborn and Portner, Mouton de Gruyter.
- Stalnaker, Robert (1968). "A Theory of Conditionals." In N. Rescher (ed.) *Studies in Logical Theory*, Oxford.
- Stalnaker, Robert (1975). "Indicative Conditionals." *Philosophia*, 5.
- Stalnaker, Robert (1978). "Assertion." *Syntax and Semantics*, 9.
- Stephenson, Tamina (2007). "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste." *Linguistics and Philosophy*, 30(4): pp. 487–525.
- 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, and S. Kawahara (eds.) *Proceedings of NELS 33*.
- Weatherson, Brian (2001). "Indicative and Subjunctive Conditionals." *The Philosophical Quarterly*, 51(203): pp. 200–216.

- Werner, Tom (2006). "Future and Non-future Modal Sentences." *Natural Language Semantics*, 14(3): pp. 235–255.
- Williamson, Timothy (2006). "Indicative versus Subjunctive Conditionals, Congruential versus Non-hyperintensional Contexts." *Philosophical issues*, 16(1): pp. 310–333.
- Yalcin, Seth (2007). "Epistemic Modals." *Mind*, 116(464): pp. 983–1026.
- Yanovich, Igor (2010). "Indexical Presuppositions: the Data." Unpublished manuscript, MIT.