2

What's What's Said?

UNA STOJNIC & ERNEST LEPORE   JANUARY 26, 2013

0. Introduction[1]

There is hardly a consensus on what a semantic theory for a natural language is supposed to achieve. There may never have been one. A condition so sufficiently abstract that it may please almost all players requires the theory to interpret all the expressions of the language – simple and complex. Unfortunately, this seemingly unexceptional constraint raises more questions than it answers; in particular, what counts as a correct interpretation? Imagine competing semantic theories for language L that, even if constructed in the same theoretical framework, disagree on their interpretation for some sentence S. Which if either is to be preferred? A natural answer is that the theory that interprets S according to what its utterances intuitively say is to be preferred over one that does not. Though this condition is obviously not the only measure of success, its rationale should be obvious: insisting on a tight connection between interpreting a sentence as p and its users intuitively saying that p helps to explain why, normally, those linguistically competent with a language upon hearing its sentences uttered, can discern what's said.[2] In a simple but clumsy form, this constraint comes to:[3]

---

[1] Thanks to David Braun, Gil Harman, John Hawthorne, Christopher Hom, Kirk Ludwig, Adam Sennett, and especially Matthew Stone.

[2] We are *not* using 'what's said' and 'what's uttered' in any technically loaded sense.

[3] Note, ST is only a necessary condition on adequacy, and as such, is compatible with further constraints on full adequacy.

> *Said That (ST)*: A semantic theory T for a language L should assign p as the semantic content of an utterance u of a sentence S in L iff "The speaker said that p" is a true report of u.[4]

Since Galileo in uttering "La terra si muove" makes "Galileo said that the earth moves" true, it follows by ST that an adequate semantics for Italian must assign to "La terra si muove" as its semantic content *the earth moves*. Any theory that does not have this consequence, then, is unsatisfactory.

A rather neat picture emerges from an ST-constrained semantics: it's intuitive, connecting a theoretical notion of semantic interpretation with a pre-theoretical notion of what's said; and it's useful, providing clear grounds for all sorts of semantic theorizing, including criteria for when two sentences *same-say* each other, and partially, even for when they translate one another. ST, had it been true, would have made life easy, but, unfortunately it is not (Cappelen and Lepore, 1997, 2005). To get a feel for what has gone wrong, consider data of the sort that ST critics like to highlight.

Should Prof X utter (1a), then A, in reporting this utterance to the shortest student, might very well use (1b):

(1a) The shortest student should sit in the front row.

(1b) Prof X said that you should sit in the front row.

Should Prof X utter (2a) in response to A's asking, "Did Alice pass your exam?", then A, in reporting this utterance to Alice's adviser, might very well use (2b):

(2a) No one failed.

(2b) Prof X said that Alice passed.[5]

Should B want to report A's utterance of (3a) to some movers, knowing that when A spoke only one table was in Room 211, but another with flowers on it has since been added, B might very well use (3b):

(3a) The table in front of Room 211 has to go.

---

[4] We will be using "interpretation" and "content" interchangeably throughout; we hope this doesn't create any confusions.

[5] Some argue (e.g. Farkas & Brasoveanu, 2007.) that reports like (1b)–(2b) are *generally* infelicitous. All that we need to register our point, though, is that there are contexts in which such reports are both licensed and perfectly felicitous. And indeed, there are (as we show below).

(3b) A said that the table in front of Room 211 *without flowers on it* has to go.

The obvious fact that emerges from these sorts of data is that the felicity of say-that reports often depends on non-linguistic considerations (e.g. whom you are talking to, what you are trying to accomplish with your report, how have conditions changed since the original utterance, etc.).

To take two further examples, though (4b)-(5b) might sound odd as reports of (4a)-(5a) when offered out of the blue, it is easy to imagine appropriate contexts in which each is felicitous, as in (4b')-(5b'):

(4a) A: John and Mary first went to the movies, then they had dinner together and then they went to the party. They had a great time.

(4b) B: A said that John and Mary had a great time at the party.

(4b') B, when being asked whether the party was any good: A said that John and Mary had a great time at the party.

(5a) A: Bill bought a sandwich. It cost $7.

(5b) B: A said that the sandwich cost $7.

(5b') B, when asked about the cost of the sandwich: A said that the sandwich cost $7.

Since felicitous *say-that* reports can, and often do, depend on shared non-linguistic beliefs about the contexts of utterance and of the report, such reports look rather suspect as guides to isolating semantic content. Additionally, since distinct utterances of the same sentence often license radically different reports, to insist on ST would require the sort of massive contextualization[6] most of us are unwilling to tolerate.[7,8] In short, life turns out to be not so good, after all; but, then, where do we go from here?

---

[6] We believe, but won't argue here, that it's easy enough to extend these cases in a way that suggests almost any sentence can be used to say *anything* given the appropriate context (of the utterance or/and of the report).

[7] Of course, in a quest for adequacy, it is slightly uncomfortable to reject ST on these grounds. The obvious rejoinder is: inadequate according to which criterion? We might reply that any theory predicting such massive context-sensitivity would not only be surprising and inelegant, but simplicity considerations strongly favor simpler theories. Of course, obviously,

One strategy tries to restrict ST to *literally/strictly speaking say that* reports; the idea is that we should replace ST with:

> *Literally Said That (LST):* A semantic theory T for a language L should assign p as the semantic content of an utterance u of a sentence S in L iff "The speaker literally/strictly speaking said that p" is a true report of u.

LST proponents concede that some *say that* reports fail to isolate *semantic* content, but take solace in thinking that the corresponding *literally/strictly speaking say that* reports do. In cases (1)-(3) above, it is not unreasonable to protest that these reports, though true, are *not* what the speaker literally said. In uttering (1a), Prof X literally said the shortest student should sit in the front row. In uttering (2a), Prof X literally said that no one failed. In uttering (3a), A didn't *strictly speaking* say the table without the flowers on it in front of Room 211 has to go, but rather that the table in front of Room 211 has to go. So, at least for these cases, the addition of *strictly speaking/literally* salvages the spirit of ST. Unfortunately, as attractive as LST is, it is not an obvious advance over ST.

The main problem with LST is that *literally/strictly speaking said that reports* exhibit no less context sensitivity than the simpler *says that* ones, and so, LST still winds up predicting more *semantic* context sensitivity than many of us are prepared to swallow.

---

simplicity alone would not establish adequacy either, since two equally simple theories can disagree in the assignments of the semantic contents.

[8] A distinct criticism of ST (not pursued until §2 below) derives from the fact that certain aspects of interpretation that many theorists are inclined to include under semantics seem not to register on what's said. Consider (6)-(7):

> (6a) A: Mary stopped smoking.
>
> (6b) #B: A said that Mary used to smoke.
>
> (7a) A: John's sister lives in New Jersey.
>
> (7b) #B: A said that John has a sister.

(6a)-(7a) presuppose the complement clauses of (6b)-(7b) respectively, and these presuppositions are (arguably) linguistically triggered, and therefore, many theorists conclude that somehow these presuppositions are a part of the semantics of (6a)-(7a), but, just the same, most theorists would argue they are not said by A. Accordingly, (6b)-(7b) would be deemed false.

There's linguistic support for this conclusion; if you want to deny what A said with her utterance in (6a) or (7a), you can do so by protesting, "No, I disagree" or "No, that's false", but your denials do not deny that Mary used to smoke nor John having a sister. These data suggest that, however presupposition is linguistically encoded in (6a) or (7a), it is not a part of what's said. If this is right, it would seem to follow that not all the semantic properties of a sentence track or are part of what's said by its utterances.

Suppose a speaker utters, "John put on his shoes and left the room." Consider the report "The speaker said that John left the room." Ask yourself whether the speaker literally said that. The intuitive answer, at least intuitive to us, is that she has. However, most theorists would not want to conclude that "John left the room", uttered in c, has the same semantic content as "John put on his shoes and left the room", uttered in the same context. Rather, one entails the other. Likewise, if a speaker utters, "Anne bought a new red dress" and someone reports this utterance with, "The speaker said that Anne bought a dress", intuitively this is a correct report of what the speaker literally said. At least it is, unless by asking for a literal report, we are asking for a direct quote of what the speaker said. That we are not should be obvious as soon as we consider utterances of sentences with indexical expressions; namely, if a speaker utters "I am happy", then the report, "The speaker said that I am happy" is false in all cases in which the speaker and reporter are non-identical (and in the cases where they *are* identical, such reports are still pretty odd).

Furthermore, tightening LST by an appeal to acts of retraction (claiming that a speaker has not literally/strictly speaking said that p if, when the report is challenged, the reporter can retract to a weaker position, e.g. "Well, the speaker did not quite say that…") won't establish much progress. For one, which retraction is available with (4b) and (5b)? If challenged, the reporter would be perfectly entitled to stick to her guns. Or, suppose A utters, "I had dinner and went to the party." B can perfectly well report, "A said that she went to the party". When challenged, B cannot retract to "Well, A did not quite say that; she only said that she had dinner and went to the party." Most theorists hold that the semantic contents of "A had dinner" and "A had dinner and went to the party" are distinct, and moreover, that the second *strictly speaking* entails the first. But for obvious reasons, it would be difficult to say that a semantic theory T is adequate *only if* it assigns p to an utterance u of S iff "S said that q" is a true report of u, where q entails p.

Of course, stipulating a special meaning of *literally* or *strictly speaking* to figure in LST will not help either; that would obviously get the project backwards. The point of adhering to indirect reports in the first place is to find an *intuitive* adequacy test. If we tailor a particular meaning of *literally says that* to fit a favored semantic theory, how, then, could ensuring the truth of *such* reports have any bearing whatsoever on the adequacy of the theory?[9]

---

[9] Note that for the same reason it won't do to tailor a particular (artificial) notion of *literally says that*, it also won't do to tailor a particular notion of *says that* either. We cannot rely on a *theoretical* notion as an intuitive guide for isolating semantic content.

At this stage, two choices remain available: either reject ST and LST and look for something altogether different, *or*, conclude that they are, though *naïve*, basically on the right track, and so, continue the search for a constraint that will do the trick. In §1, we will explore the second strategy, looking at refinements of ST; in §2, we will consider the first strategy, divorcing semantics from reporting practices entirely.

1. Indirect Reports in a State of Ignorance

Denying *any* connection between semantic interpretation and indirect speech seems *prima facie* unsatisfying; why be interested in semantics if it has nothing to do with what is normally communicated by utterances of sentences? And what better way to get at what's communicated than through what's said? And what better way to get at what's said than through felicitous indirect reports? These seem to be working assumptions in most of the literature. But since there are serious reasons to be dissatisfied with both ST and LST, what's left?

We believe the most promising strategy along these lines is to restrict ST not to what's *literally/strictly speaking said* but rather to cases of *ignorant indirect reporting*. Our motivation should be obvious: re-consider utterances (1a)-(2a), and assume reports of them by someone proficient in English but ignorant of the circumstances surrounding their production; all this reporter knows is that these utterances were produced in some context or other, by some speaker of English or other:

> (1a) Prof X: The shortest student should sit in the front row.

> (2a) Prof X: No one failed.

With (1a), this restriction amounts to assuming the reporter ignorant of who is sitting in the front row, and so, in no position to use (1b).

> (1b) A, to the shortest student: Prof X said that you should sit in the front row.

Similar considerations thwart using (2b) to report (2a).

> (2b) A: Prof X said that Alice passed her exam.

Namely, A *ex hypothesi*, does not know that Alice is in Prof X's class.

But even in this state of ignorance, a reporter can *still* use (1c) in reporting (1a); and (2c) in reporting (2a):

(1c) Prof X said that the shortest student should sit in the front row.

(2c) Prof X said that no one failed.

These intuitive transitions between utterances and their indirect reports suggest a novel restriction on ST; indirect reports in situations of *ignorance of extra-linguistic information* fix semantic content.

We, of course, want to insulate indirect reports from coloring by non-linguistic information about the reporter's circumstances and his audience as well. Here is why. Suppose A uttered (8a):

(8a) Vermillion is everyone's favorite color.

A reporter knowing how limited an audience's color vocabulary is might opt to report what the original speaker said, not with a color word, but with a description like "the color of my pen," holding up a vermillion pen. We would not, however, conclude that the complement clause in (8b) semantically interprets A's utterance of (8a).

(8b) A said that the color of my pen is everyone's favorite color.

To avoid such pitfalls, we recommend restricting ST to *complete* non-linguistic ignorance, allowing only for the exploitation of information one gains *qua* competent speaker, i.e. *linguistic knowledge*.

> *Ignorant Said That (IST)*: A semantic theory T for a language L should assign p as the semantic content of an utterance u of a sentence S in L iff "The speaker said that p" is a true report of u by someone ignorant of all the circumstances surrounding u as well as the circumstances surrounding the report.

The reports in (1c)-(2c) satisfy IST, and so, the idea is, the semantic interpretations of (1a)-(2a) are specified by their complement clauses.

The rationale behind IST is intuitive enough: reports in circumstances of ignorance abstract away from all those features wedded to context – whether the context of the utterance or the context of the report of the utterance – and thus, they move closer to capturing what's common to every utterance of the sentence (-type). And it's natural to think what's common is what's semantically encoded. Since in ignorant reporting, the only knowledge to draw on is linguistic knowledge, it would seem to follow that

such reports provide the best intuitive guide to content. IST, thus, re-establishes a connection between semantics and what's intuitively said.

Unfortunately, IST runs into trouble with *genuine* linguistic context sensitivity. If Harry utters (9a), how would someone ignorant of all of the extra-linguistic facts report him?

> (9a) I am Harry.

It would seem that the best a reporter could do would be (9b):

> (9b) The speaker said that the speaker is Harry.

Or, if someone uttered (10a), it would seem that the best a reporter could do would be (10b):

> (10a) It's raining here.

> (10b) The speaker said that it is raining at the location of the utterance.

Similarly, the best a reporter could do for an utterance of (11a), when in a state of extra-linguistic ignorance, would be (11b):

> (11a) That's lovely.

> (11b) The speaker said that the object demonstrated is lovely.

(We leave it to the reader to extend the strategy to "he," "she," and other familiar context sensitive expressions.)

According to IST, then, we should conclude, assuming that these intuitive indirect reports are accurate, that the full semantic content of (9a) is *that the speaker is Harry;* the full semantic content of (10a) is *that it's raining at the location of the utterance;* and the full semantic content of (11a) is *that the object demonstrated is lovely*. But there are familiar reasons why many theorists have thought this may not be such a good idea.

According to Kaplan (1989), when Harry utters (9a), he semantically expresses the necessary truth that he is Harry, but the complement clause of (9b) does *not* semantically express a necessary truth. The speaker, Harry, might have remained silent, or he might have been mute. Someone else might have spoken instead. The point is, as Kaplan famously argued, indexicals and demonstratives are *directly referential* and *rigid*, whereas descriptions, e.g. "the speaker," are not directly referential (though some are rigid).

Thus, "I" and "the speaker" do not share the same modal profile. Conclusion: "I", does *not mean the same as* "the speaker", and so, IST must be wrong.

Kaplan's critical point generalizes. If A utters (12a) at time t_1, then the best anyone can do in reporting A, assuming extra-linguistic ignorance, is (12b):

(12a) It's raining now.

(12b) The speaker said that it's raining at the time of the utterance.

But, of course, the time of utterance might have differed from t_1; the speaker might have spoken later. Nevertheless, time t_1 could not have differed from itself. So "the time of utterance" and "now" do not share the same modal profile, and so, cannot share meaning. Similarly, if A utters (10a), at location l_1, then the best an ignorant reporter can do is to report A with (10b). However, while l_1 cannot fail to be itself, the speaker might have chosen to speak somewhere other than at l_1. And so, "the location of the utterance" and "here" do not share a modal profile. And so on for other recognized context sensitive expressions.

The point is familiar: the *modal profiles* of "the speaker," "the time of utterance," "the object demonstrated," "the place of utterance" are all distinct from that of "I", "now," "that", and "here" respectively; but, so Kaplan's argument continues, only expressions with the same modal profiles can share *semantic content*. Nothing, in general, could satisfy "A" without satisfying "B," if "A" and "B" are synonyms. IST seems to require us to violate this common background assumption.

It is worth pointing out that Kaplan is not denying that competent English speakers know that the first person pronoun "I" always picks out the speaker. Nor that uses of "now" pick out the time of utterance, and "here" the place of utterance;[10] and uses of "that" the demonstrated object. And so, Kaplan is not denying that a linguistic theory should encode (i)-(iv) *somewhere:*[11]

      (i)   Every use of "I" picks out its user.
      (ii)   Every use of "now" picks out its time of use.

---

[10] Both of these claims – concerning "now" and "here" respectively – have been challenged, but for reasons irrelevant to our discussion: the fact is that "now" and "here" also have a demonstrative use as well as an indexical use.

[11] Though there might be counterexamples to (i)–(iv), they are not relevant for our discussion here, so we set them aside.

(iii) Every use of "here" picks out its place of use.

(iv) Every use of "that" picks out what is demonstrated by its user.[12]

However, Kaplan *is* denying that (i)–(iv) should be captured as a matter of *semantic content*. (Kaplan himself distinguishes two levels of "content" – character and content – and thereby, seems to manage to devise a theory that encodes (i)–(iv), while avoiding the modal objections. However, giving the honorific "semantic content" solely to what he calls "content," and not to what he calls "character" without argument is somewhat arbitrary. This is a topic for the next section.)

One might wonder whether Kaplan's take-home lesson should be endorsed. One might worry that the argument goes astray, since when a reporter learns that someone uttered (9a) without knowing who, there is more than one way to report her speech act. One is the way we have been doing it; using what we might call a *de dicto* report, namely, (9b). As we have seen, this sort of report, according to Kaplan, fails to capture the semantic content of the source speech act, since its complement clause lacks adequate modal properties: the report expresses a general proposition, whereas in Kaplan's and most other's considered opinion the reported utterance expresses a singular proposition.

However, another way to report the relevant speech act in these circumstances is with (9c):

(9c) The speaker said *of himself* that he is Harry.

That is, we might use what we may call a *de re* report. One might try to argue that in this way, we avoid the objection from differences in modal profiles, since the proposition that makes this report true would be singular (viz. *that x is happy* [where x = the speaker (and, assuming the source speech act is true, x = Harry)]. Under this construal, the modal profiles of the source speech act and the proposition that makes the complement of the report true are identical – for any speaker x, x's utterance of (9a) is true (with respect to a possible world in which x exists) iff x = Harry.[13]

This story obviously generalizes to other cases. If a speaker utters (11a) and a reporter overhears this utterance without knowing what the demonstrated object is (or who the speaker is), the reporter can still resort to the "*de re*" report (11c):

---

[12] Whatever the relevant notion of the demonstration is. We shall not fuss about that here.

[13] Thanks to David Braun, Kirk Ludwig and Matthew Stone for this suggestion.

(11c) The speaker said of the demonstrated object that it is love-ly.[14]

Likewise, if someone overhears a speaker saying (13a) without knowing who the speaker is, or whom the speaker is referring to, we can report this utterance with (13b):

(13a) She is nice.

(13b) The speaker said of the demonstrated/salient woman that she is nice.

The point is the same in all these cases; the complement clauses of these reports, according to the suggestion, attribute singular propositions to the speaker, and the reports share modal profiles with their source speech acts. Thus, this suggestion concludes, on the assumption that Kaplan is right, the only thing that the objection from the difference in modal profiles shows is that we were looking at the wrong reports.

In the context of our ambition in this paper, we are less then persuaded by this line of defense of IST. First of all, since all of these "*de re*" reports involve *quantifying in*, it is less than clear that their complement clauses in the cases of interest in fact deliver singular propositions, rather then general ones. The semantics for constructions of this form is notoriously hard to tract, but it seems to us that no promising account would actually deliver the desired result, namely, that the proposition expressed by the complement clause of e.g. (9c), (11c) or (13b) expresses a singular proposition, given that in all of these cases the antecedent for the relevant pronoun in the com-plement clause is a definite description ('the speaker', 'the demonstrated object', 'the demonstrated/salient woman'). Given these problems, we think the proposal is a non-starter. But even if we were to set these worries aside, and suppose that the complement clause of the "de re" reports does in fact express a singular proposition in all the right cases, there are further worries with this proposal. None of these "de re" reports actually specifies the se-mantic contents of the original speech acts, since they all involve *quantify-ing in* – i.e. they are all of the form, "The speaker said of x that x is F" – where it is not known to the reporter who (or what) x is. Thus, these reports in effect merely *describe* the content of the original utterance. And so, on

---

[14] One might be a little uncomfortable with this use of "object" here, since arguably we can use demonstratives to refer to things we wouldn't naturally call objects [e.g. events, etc.]. Per-haps, the more neutral "thing" would be better.

their bases, we can merely infer that there is *some* (singular) proposition the speaker expressed, but, crucially, we cannot retrieve what it is. This is *not* sufficient if the aim is to retrieve the semantic content of the source speech act.

In short: even if invoking "*de re*" speech act reports avoids the objection from differences in modal profiles, we still fail to articulate a satisfactory criterion of adequacy.

Perhaps, one could attempt the following rejoinder. Even though in the aforementioned cases, "*de re*" reports do not *reveal* semantic content, but merely describe it, still this might be sufficient, if these "*de re*" reports nevertheless manage to uniquely capture semantic content. And one might argue that in all the aforementioned cases, (9)–(13), the "*de re*" report in question describes a unique semantic content; that is, the truth of each establishes that there is only one (singular) proposition in each case, that uniquely renders the report true.

Even if this were true, would it vindicate the IST? We think not. Remember, we are trying to find a criterion of adequacy on a semantic theory. Now, (granting, for the sake of argument that it is sufficient that "de re" reports merely describe semantic content) here is our current situation with *IST*. As we have seen, there are two types of *says that* reports we could look at – "*de re*" ones and "*de dicto*" ones. The reason to be suspicious of *de dicto* reports was that in the cases of linguistic ignorance, if Kaplan and Kaplaneans are right, looking at those reports would predict the wrong results for (9)–(13). For, the argument goes, the modal profiles of the source speech acts in (9a)–(13a) do not match the modal profiles of the complement clauses of the reports in (9b) – (13b).

But, the problem is that the argument from differences in modal profiles already presupposes we somehow have a direct insight into the semantic content of (9a)–(13a). However, these intuitions cannot be (based on) the intuitions about what is said. Since both "*de dicto*" and "*de re*" reports in the relevant cases are true, merely by looking at what's said, we have no more reason to think one type of report tracks what's said better than the other. And so, if what's said is supposed to afford us insight into semantic content, we have no more reason to think "*de re*" reports track semantic content better than "*de dicto*" ones. So, in these cases, modal intuitions, rather then what is said, are doing all the work.

This is obvious once we appreciate that in other types of cases, we would say *de dicto* reports capture semantic content rather than "*de re*" reports (granting *for the sake of the argument* that "*de re*" reports do in fact have their complement clauses express a singular proposition in the relevant cases, which we have seen is as a matter of fact dubious), as in (14)-(16):

(14a) The speaker utters, "The tallest building in the world is in Dubai."

(14b) The speaker said that the tallest building in the world is in Dubai.

(14c) The speaker said of the tallest building in the world that it is in Dubai.

(15a) The speaker utters, "The top ranked male tennis player in July 2011 is from Serbia."

(15b) The speaker said that the top ranked male tennis player in July 2011 is from Serbia.

(15c) The speaker said of the top ranked male tennis player in July 2011 that he is from Serbia.

(16a) The speaker utters, "The fountain of youth is hard to find."

(16b) The speaker said that the fountain of youth is hard to find.

(16c) The speakers said of the fountain of youth that it is hard to find.

In these cases, invoking modal intuitions would suggest that "*de dicto*" reports more adequately capture semantic content. So, it seems, according to this line of thought, that in some cases "*de re*" reports better capture semantic content, but in others "*de dicto*" reports fare better. How do we tell when to rely on one and when to rely on the other? The most natural thought that comes to mind is – by appealing to modal intuitions: ignorant *say that* reports in tandem with modal intuitions serve as an intuitive guide to semantic content.

However, this suggestion is too quick. Consider the following:

(17a) The speaker utters, "The smallest prime is everyone's favorite number."

(17b) The speaker said that the smallest prime is everyone's favorite number.

13

(17c) The speaker said of the smallest prime that it is everyone's favorite number.

Modal intuitions are silent in this case between "*de re*" and "*de dicto*" reports. And the relevant distinction we want to make – namely, between directly referential and non-directly referential terms comes from within the theory, and so, is of no help when the criterion of adequacy is in question. So, it seems we are back to square one – LST does not get us what we want.

Up to here, we have focused exclusively on intuitions about *say that* reports made in various ways and under various conditions. The problems we have run into repeatedly derive from the fact that intuitions about the felicity of these reports are apparently *not* guided by judgments about semantic content *alone*. They can also be informed by knowledge (or a lack thereof) of the circumstances surrounding both a report and the reported utterance. These reliances are so strong that the more we attempt to restrict their impact (by adding modifiers like "strictly speaking" or "literally", or by imposing extra-linguistic ignorance), the more difficult it becomes for us to respect a commitment to the spirit of ST.

The bottom line is that intuitions about indirect reports (as well as about other sorts of attitudinal attributions) are invariably guided by considerations irrelevant to semantics; as observed, they can be guided by knowledge of the context of the speaker, or of the reporter, and by how much information the relevant participants have about either of these contexts or about the world in general. Exploiting these sorts of information is foundational to our reporting practices. Agents say what they say for a reason, and it's hard to divorce our acts of reporting their speech acts from the influence of knowledge of their and our own intentions and motivations, and goals and purposes.

To the extent that this is right (and it is), we need to identify a better way to separate *semantic* considerations from other sorts of consideration that a reporter may add to the mix. *Say that* reports, in any incarnation, are either too restrictive or too permissive to settle semantic adequacy: ignore context entirely, and semantic adequacy becomes elusive; let it in, and it becomes too liberal. Either way, ST and kin are *not* able to capture all and only semantic content. Our favorite version, IST, had the advantage of capturing only conventionally (linguistically) encoded information, but as we have seen, is still less then satisfactory. For these reasons, we are indeed very pessimistic about the prospects of using *say that* reports as guides to semantic content. We thus propose to drop this line of inquiry altogether and turn to a different kind of methodology. We believe that a solution to the problem of finding a way of abstracting distinctively *all* and *only* se-

mantic content is locatable in Lewis' twin ideas of convention (1969) and the conversational record (1979), to which we shall now turn.

2. Lewis on Coordination on the Conversational Record

We begin with Lewis's (1979) notion of the conversational record, i.e. an abstract 'scoreboard' that tracks interlocutors' contributions to their interaction; the record includes, among other things, the environment of the conversation, what the conversation is about, what information they are attempting to distribute in the course of the conversation, as well as those assumptions, presuppositions, and other items implicitly or explicitly acknowledged, e.g., the referents of names (Thomason 1990), a 'home base' to accommodate words like 'come' and 'go' (Fillmore 1975). New utterances naturally force updates and changes to the scoreboard. In this regard, the record is a running database – sometimes items are added; sometimes they are removed. Its topic might change; its presuppositions might be challenged; and its participants might change their minds about items previously recorded.

Placing information on the conversational record does *not* require the speaker or audience to believe or desire, or to have come to believe or desire, or to intend to do anything. Just the same, the record develops so that, all things being equal, the contributions a speaker makes are treated as if they were true if possible (at least for the purposes of the conversation). If someone utters, "Mary stopped smoking," then unless another party objects, the presupposition that Mary used to smoke automatically enters the record; as does the at-issue proposition that Mary no longer smokes.

Why is any of this pertinent to the task of identifying an adequacy condition for semantic theory? As we construe Lewis, in *some* cases, information is placed on the record because it has been signaled to the audience by the speaker's utterance in virtue of shared linguistic *conventions*; and, in other cases, information is put there by virtue of being suggested, or indicated, or revealed through broader background knowledge.[15] Illustrations will help to clarify these differences.

Suppose Harry utters, "I'm happy." Then minimally it enters the record that the speaker made this utterance. Additionally, interlocutors can track that it was Harry who uttered "I". These contributions obviously go on the record as a matter of brute observable fact. Other publicly available infor-

---

[15] This is intended as the broadest distinction. Subdivisions on both sides are possible. Different types of background knowledge might impact the conversational record in different ways. Likewise, different types of information might get on the record in virtue of an extant linguistic convention, in different ways. We do not pursue the possible subdivisions here.

mation can enter in a similar way – if Harry is smiling, and all the conversational participants mutually recognize as much, then in normal circumstances, it will become a part of their records that the speaker is smiling. But now suppose that Harry utters, "Trenton is in New Jersey" to a previously uninformed audience. Hearing this utterance may very well result in their adding to the record the proposition *that Trenton is in New Jersey*. In *this* case, the information that enters does so as a result of the participants exploiting shared linguistic conventions. (The exact notion of convention is to be clarified shortly. All that need be noted thus far is that in order to interpret Harry's utterance about Trenton the audience needs to invoke the knowledge that they have as competent speakers of their shared language.) If they don't exploit their knowledge of these conventions, then this particular information would not wind up on the record.

The obvious question that concerns us is how to distinguish among various means of information distribution, namely, those that invoke background or public knowledge *from* those that exploit shared linguistic conventions. To this end, we can exploit Lewis (1969) in conceptually separating the different kinds of situations interlocutors meet in a conversation when deciding which information to enter on the record, by invoking Lewis' key explanatory notion of *coordination*, in terms of which he proposes to analyze the notion of convention.

Coordination can occur when agents face a coordination problem. These sorts of problems crop up wherever there are situations of interdependent decision by at least two agents, where coincidence of interest predominates, and where there are at least two coordination *equilibria*, i.e. at least two ways that participating agents can coordinate their actions for their mutual benefit. Agents solve a coordination problem when each acts so as to achieve equilibrium. They do so *by coordination* when, confronted by multiple options for matching their behaviors, they exploit their mutual expectations in settling on one equilibrium (where each agent does as well as he can given the actions of others) to the exclusion of all others.

Lewis illustrates this sort of practice with Hume's example of two men, A and B, in a rowboat: to move, they must coordinate their rowing patterns. There are almost a limitless number of speeds at which each can row, but to row effectively, they need to settle on a single speed, which, interestingly, they can achieve without an explicit agreement. They may stumble on to it; or one might mimic the other. But, should A row at a certain speed because A expects B to do so; and should B row at a certain speed because B expects A to do so; and so on, such that each does his part because he expects the other to do his, then they, thereby, reach an equilibrium through coordination.

The practice of updating the conversational record so as to register specific information also poses a coordination problem. After all, there is *no non-arbitrary* connection between an utterance and what a speaker can use it to register on the record (other than that the speaker made the utterance). But if the speaker's strategy is to use a particular utterance to get his audience to register particular information on the record, and if he expects his audience to respect this strategy, and if the audience should happen to respect a like-strategy in tracking the information that the speaker is attempting to place on the record, and if the audience expects the speaker to respect this strategy, and so on, then the speaker and audience will end up, *through coordination*, with identical updates of the conversational record.[16]

The way in which agents reliably solve coordination problems is by adhering to a particular scheme implicit in their tendencies or mutual expectations. The key to understanding how coordination functions in solving coordination problems is to appreciate the surprisingly underappreciated role that *conventions* play.

A *convention* is a regularity observed by agents, but, of course, not *every* regularity constitutes a convention; eating and breathing are *regularities* that we each follow but they are *not conventional*. Someone adheres to a convention just in case his *reason* for acting in accordance with a certain equilibrium solution to a coordination problem is that he expects others will act in accordance with this same solution to the problem, and that they will do so only if they expect him to act in accordance to the same solution, and he further has some reason for expecting them to act in accordance to the same solution (Lewis 1969:42).

A group of agents are said to share a convention, then, just in case each member does her part in regularity X because she expects everyone else in the group to do their part in X, and each party prefers to do their part in X conditional upon others doing so. Had anyone expected everyone to do their part in another alternative pattern Y, she would have done her part in regularity Y (and not in X).[17]

A convention, in short, is simply a self-perpetuating solution to recurring coordination problem. A group is reliably good at solving a coordination problem only if its members either share patterns of behavior or background knowledge that enables them to choose one pattern over viable others. Since *interlocutors* are apt at retrieving contributed information from heard utterances, and since each conversation creates a coordination prob-

---

[16] Of course, there needs to be mutual recognition as well. See Lewis (1969).

[17] It's crucial for Lewis' idea that Y exists. That follows from how coordination problems are defined.

lem for its participants, it follows, by definition, that the participants are exploiting linguistic conventions.[18]

The lesson we take away from Lewis (1969) on convention/coordination combined and Lewis (1979) on the conversational record is how to devise a proposal for semantic adequacy; in particular, one that builds on the idea that for some utterances a speaker intends for the audience to add to their conversational records particular information *as a matter of coordination*. For this to be successfully achieved, the speaker and the audience need to draw upon their shared knowledge of linguistic conventions. We propose, then, to say that a semantic theory is adequate just in case it specifies the conventional knowledge that goes into determining this, and only this, information. So construed, the proposal for semantic adequacy becomes *Coordination* (CRD):

> *CRD*: A semantic theory T for a language L should assign as semantic
> content to an utterance u of a sentence S of L whatever u of S contributes
> to the conversational record in virtue of coordination.

CRD, unlike ST and its kin, is very permissive; according to it, any aspect of *conventionally* encoded information contributes to semantic content;[19] and not only whatever conventionally encoded information goes into determining what a speaker has said with her utterance.

We welcome semantic liberalism; IST was appealing precisely because it stripped utterances of all non-linguistic information in the service of attempting to isolate all and only the information that is recovered in virtue of invoking linguistic convention alone. Lacking an adequate notion of a con-

---

[18] Interlocutors without a shared convention can still solve coordination problem, but it would be plain luck or an innate alignment that accounts for their success because there's no reason *except for convention* to choose one regularity over another in facing a coordination problem (i.e., communication is "a consequence of conventional signaling" (Lewis 1969: 150)).

[19] Perhaps, this might include expressive content, conventional implicatures, presuppositions and other non-at-issue information. Though we will not argue here that any one, or all, of these aspects are conventional, we leave it open whether some (or all) of them might be. There also remain interesting questions about how, if these aspects are conventional, this framework can explain, for example, the difference between it entering the record that Harry is in pain after he utters "Ouch!" vs. its entering the record after his uttering "I am in pain." Or, how can it explain the difference between its entering the record that Mary used to smoke after an utterance of "Mary stopped smoking" vs. an utterance of "Mary used to smoke"; or its entering the record that there's a contrast between being French and brave, after an utterance of "Dan is brave but French" vs. an utterance of "There's a contrast between being brave and French." It might be that there are many ways for the speaker and the audience to coordinate (as a matter of convention) on the same proposition, even if these different ways do not encode meaning in the same way. How to explain, or even to describe, this in Lewis' framework is a topic for another discussion.

vention (as well as its commitment to indirection in accessing conventionally encoded information), IST fell to problems with context sensitive expressions. As we will show below, CRD succeeds in reconciling these goals – it captures all and only conventionally encoded information, whole avoiding the pitfalls of IST.[20]

To illustrate what CRD determines, consider first a non-context sensitive (ignoring tense) case, where a speaker wants to inform her audience that Trenton is in New Jersey; first, she needs to identify an utterance that she is confident that, in her circumstances, will put the proposition that Trenton is in New Jersey on their conversational record. She must choose an utterance that in the context of the conversation at that stage provides the audience with evidence for registering this proposition (and not another) on the record. Convention enters the calculation because the speaker and audience can solve the coordination problem they confront (i.e. how to end up with the same conversational record) only by coordinating their mutual efforts in tracking contributed information.[21] In this case, what enables them to do so is their shared convention that a speaker utters, "Trenton is in New Jersey" (in this context, in this manner, with their shared common ground) only if she wants the audience to add to the record the proposition that Trenton is in New Jersey; and, likewise, the audience infers, as a matter of convention, that the speaker utters, "Trenton is in New Jersey" (in this context and manner) only if she is putting this proposition on the record, and thus, they achieve an equilibrium to their coordination problem. (We leave it as an exercise for the reader to go back over earlier examples to convince yourself that they are also captured by CRD.) Of immediate interest to us is whether similar considerations extend to cases involving context sensitive expressions.

Suppose Harry opts to convey that he is happy to his audience by uttering, "I am happy." With this utterance, he is confident, in his circumstances, that he will get his audience to put the proposition that he is happy on the record. The linguistic convention Harry is adhering to is that a speaker X utters, "I am happy" (in this context, in this manner, and with this shared common ground) only if he wants his audience to add to the record the proposition *that he is happy*; and his audience infers, by appealing to the same convention, that Harry utters, "I am happy" (in this context and man-

---

[20] Note that we are not trying to settle, in the present paper, how linguistic conventions come about. Nor are we claiming that they cannot change with time. We are only interested in when a semantic theory captures the information that an adequate semantic theory should capture.

[21] Solutions that would occur by mere luck would, obviously, be irrelevant, so we set them aside.

ner) only if he is putting this proposition on the record, thus achieving an equilibrium to their coordination problem.

This doesn't automatically mean that the separate proposition *that the speaker is happy* won't also enter the record in the similar fashion, i.e. in virtue of a shared linguistic convention.[22] It is perfectly compatible with all we have said that there are cases where with a single utterance more than one proposition enters the record in virtue of extant conventions.[23] What goes on the record as a matter of coordination and what doesn't comes down to which information is and which isn't linguistically encoded. CRD essentially constrains a semantic theory, stating that it is adequate *iff* it captures all and only what's linguistically conventionally encoded.[24] That much is unsurprising. The merit of CRD over ST, LST, and IST is that it offers a direct way of capturing what's conventionally encoded, by appealing to the twin notions of coordination and the conversational record. In this way, it avoids the problems that the previous proposals were stuck with.[25]

An important residual worry is how CRD accommodates cases where the audience, ignorant of non-linguistic information, overhears an utterance of a sentence containing a context sensitive expression. For the sake of concreteness, suppose an audience overhears an utterance of "I am happy," but has no idea who made the utterance. What happens in these circumstances to the conversational record? Our answer is – nothing special. It certainly becomes part of the record that this speech act occurred, i.e., that some speaker uttered this sentence, and if they understand English, it also enters the record that the speaker (of the utterance in question) is happy. However, since it is unbeknownst to the audience that Harry spoke, it will not become part of the record that Harry is happy. This is perfectly in accord with CRD.

This case is problematic for IST, since the overarching hope and promise guiding IST is that we can strip ourselves of all non-linguistic information, and still isolate *all* of what is semantically encoded with indirect

---

[22] In fact, this is in essence what some of the proponents of semantic two-dimensionalism or other types of semantic pluralism would advocate.

[23] However, we do not, at present wish to commit ourselves to this claim.

[24] Of course, the semantic content of an utterance u of the sentence "I am happy" is *not* independent from the state of the conversational record at the time of utterance. Had it been a part of the record that someone other than Harry, say Bill, was the speaker, then given that knowledge, in virtue of the same linguistic convention, the interlocutors would have coordinated on the proposition that Bill is happy. Similar considerations hold for other examples. More on this below.

[25] Indeed, we are not claiming anything surprising by claiming that what an adequate semantic theory should capture is all and only what's conventionally linguistically encoded. Most theorists would agree. What's more important in our claim, and what has been missed in the debate thus far, is that by invoking an appropriate notion of a convention (analyzed in terms of coordination) and the conversational record we gain a direct route to semantic content.

reports. However, it is precisely this appeal to ignorance that renders IST unsatisfactory once context sensitivity is included into the mix. CRD faces no such problem. Of course, in ordinary linguistic practice, occasionally we find ourselves, as a matter of fact, ignorant of all, or nearly all, non-linguistic information. But no one ever said that in every case, for any utterance, a competent speaker can retrieve *all* of the semantic content (we certainly never said that). In fact, such claims are blatantly false.

The problem with IST is not simply that in some instances of non-linguistic ignorance it is impossible to retrieve all of the semantic content, but rather that by virtue of its essential appeal to non-linguistic ignorance, IST is rendered incapable of explaining why competent speakers *can* and *do coordinate* on certain propositions (e.g., that Harry is happy), while other competent speakers (the ones facing non-linguistic ignorance) cannot. With its self-imposed limitations, IST cannot account for the complete semantic contribution of indexicals and other context-sensitive expressions. No such problem confronts CRD. For, crucially, CRD permits interaction between knowledge of linguistic conventions and background, non-linguistic knowledge. And, in so doing, it correctly predicts that there shall be cases in which full interpretation is rendered impossible since access to relevant knowledge is blocked.

3. Conclusion

We began by considering a string of possible criteria of adequacy on a semantic theory, where each tries to capture the connection between what's linguistically encoded and what's intuitively communicated by focusing on speech act reports. Such attempts are prevalent in the literature. We argued that even the most promising one – *IST* – fails to deliver adequate results.

By appealing to Lewis' twin ideas of the conversational record and convention (through coordination), we saw that we can get around the problems facing *IST*. This, of course, is no accident. Intuitively, an adequate semantic theory should be concerned with underwriting all of the knowledge that speakers have in virtue of linguistic competence. And that includes nothing more and nothing less than knowledge of the extant conventions governing linguistic usage. *IST* was on the right track by virtue of dispensing with non-linguistic knowledge, and thereby attempting to isolate all and only linguistically encoded (i.e. conventional) information. However, by trying to isolate this conventional information *indirectly*, through ignorant speech act reporting, it imposes too severe restrictions on the interplay between conventional knowledge and non-linguistic knowledge – re-

strictions that would ban the full semantic effects of context-sensitivity altogether.

CDR skirts these problems by isolating conventional content *directly*, through the practice of coordination. Thus, it need not impose any strong and implausible restrictions on the interaction between the conventional and the non-conventional. When Harry utters, "I am happy", his audience can draw on the convention that someone utters this sentence only if s/he wants us to add the proposition that s/he is happy to the record. In this case, since they posses the background knowledge that it was Harry who uttered the sentence, they can add that Harry is happy to the conversational record. For them to do so, they obviously have to draw upon the common ground (but non-conventional) knowledge that Harry is the speaker. These two pieces of knowledge interact, and together they permit the audience to coordinate with the speaker on recovering the proposition that Harry is happy. This is precisely as it should be.