Stanley on the *de se*

Comments on Stanley, *Know How*, Pacific APA, 4/5/2012

Seth Yalcin
yalcin@berkeley.edu

1 Why think know how is a *de se* attitude?

(1) is not is not truth-conditionally equivalent to (2):

(1) Bekele expects to win the race.

(2) Bekele expects Bekele to win the race.

—for Bekele may expect himself to win without realizing that the person he expects to win is he himself. In such a case (2) is true, but (1) may be false. Knowledge-wh ascriptions have a similar structure, with an infinitival clause beneath an attitude verb:

(3) John knows how to hit a ball hard.

Moreover the syntax of both (1) and (3) are widely thought to involve a phonologically null element PRO:

(1′) Bekele expects PRO to win the race.

(3′) John knows how PRO to hit a ball hard.

and we often find *de se* readings where we find PRO, at least under attitudes.

But as Stanley notes, there is an important syntactic difference between (1) and (3), which is that the PRO in (1) but not (3) is obligatorily controlled by the matrix subject.

What exactly that means depends on one's theory of control and of PRO, but for now the apparent theoretical upshot is that there should be at least two interpretative options:[1]

(3′) (a) **Control reading:**
John$_i$ knows how PRO$_i$ to hit a ball hard.
"John knows how he himself (*de se*) can hit a ball hard."

(b) **Arbitrary reading:**
John knows how PRO$_{arb}$ to hit a ball hard.
"John knows how one can hit a ball hard."

---

[1]In fact there clearly are more interpretative options, as observed by Stanley and Williamson [2001], Bhatt [2006], among others; notably there is at least one reading where the tacit modal is deontic *ought*. But for now I merely want to fix on the apparent possibility of these two options.

Now the question arises,

> How can we test for whether any given case of knowledge-wh is a case where the PRO is in fact controlled, and hence a case where a proper understanding of the *de se* reading of PRO is probably going to matter?

I did not notice a test in *Know How*. Considering the matter directly, it seems quite hard to tease these readings apart, in many cases anyway. Naively, one might expect to be able to force the supposed *de se* reading by adding a sentence which negates the arbitrary reading. But this seems not to work:

(4) ? John knows how to hit a ball hard, although he does not know how one can hit a ball hard.

Perhaps this is because the *de se* reading of (3) is sufficient, but not necessary, for the truth of the arbitrary reading? But then, constructions expressing this aren't very felicitous, either:

(5) ? John knows how one can hit a ball hard, but he doesn't know how to hit a ball hard.

Is the problem just that these two kinds of knowledge normally go together? Perhaps, but a skeptic might begin to worry that there really are these two distinct possible readings of (3). Perhaps there is just the arbitrary reading. Or perhaps most of the time the overwhelmingly natural reading is the arbitrary reading.

So it is not completely clear to me why Stanley holds that know how is, or is usually, a *de se* attitude. The book proceeds under the assumption that in all of the relevant target cases, the knowledge-wh in question involves controlled PRO. What is the argument for this view?

(Is there an argument *against* the assumption? Perhaps there is, if knowledge entails belief, and we adopt a Stanley's conception of the kind of content implicated in know-how ascriptions. For there seem to be cases of know how in the absence of (what Stanley's view predicts is) the corresponding *de se* belief.)

2 Stanley against the Lewisians: empirical data

Predicational view. that in cases of subject-controlled PRO, the infinitival clause headed by PRO generally has a semantic value equivalent to that of a property of individuals. Where the matrix verb in such cases is an attitude verb, we model the corresponding attitude state along the lines of Lewis [1979].

Propositional view. PRO does not receive receives a bound-variable reading in subject control constructions. Instead, PRO is assumed on this view to "inherit its reference from its antecedent", i.e., the matrix subject. This account does not require the assumption that attitude verbs sometimes express relations to properties.

2.1 Puzzling Reinhart sentence

Reinhart sentence:

(15) John wants [PRO to become a doctor], but his mother doesn't want that.

Stanley:

> On one natural reading of (15), what Johns mother doesn't want is *that John becomes a doctor*. On this reading, the use of "that" clearly denotes a proposition about John. This is evidence for the propositional theory, over the predicational theory, since the propositional theory has no problem accounting for this reading. (80)

This is trouble. However, a slight change to the example generates an analogous problem for the propositionalist. Consider:

(6) Most of Hannah's children want [PRO to go into medicine], but Hannah doesn't want that.

Not clear what the proper analysis of these constructions are. But it does seem clear that the propositionalist theory Stanley has in mind is not going to be not adequate for these examples in general. On such data, the predicational and propositional views seem at best to draw.

2.2 Sloppiness and *de se* readings of indexicals

Stanley:

> ... the desire to explain *de se* readings, and the desire to account for the fact that a *de se* pronoun like PRO only gives rise to sloppy readings in Verb Phrase ellipsis... come apart. A pronoun with a *de se* reading can easily give rise to strict (invariant) readings in Verb Phrase ellipsis, as in (16):

(16) I believe I am a philosopher, and you do too.

> ... But it is natural to take the belief-ascription in (16) as a *de se* believe ascription. ... [the predicational view] predicts that (16) only has the interpretation (18b), and not the interpretation (18a):

(18a) I believe I am a philosopher, and you believe I am a philosopher.

(18b) I believe I am a philosopher, and you believe you are a philosopher.

> ... To capture reading (18a), we must treat the *de se* pronoun "I" in (16) as an expression that refers to the speaker. Therefore, the property of being a *de se* pronoun, and the property of contributing a λ-abstract are distinct. It is hard to see how any theory that treats "I" as anything but a referring expression in (16) can account for these facts. (80)

But:

- The predicational view is not a view about first-person indexicals, but a view about the interpretation of controlled PRO. So it makes no predictions at all about (16).

- The predicationalist can, like the propositionalist, start with an ordinary Kaplanian semantics for first-person indexicals. The problem is then to explain the possibility of the reading (18b); and as far as this problem goes, the predicationalist and the propositionalist are on a par.

- Moreover, if the predicationalist wishes to recognize a special *de se* reading of first person indexicals, she can certainly help herself to semantic mechanisms beyond abstraction over PRO. One widely discussed recent move is to recognize context-shifting operators (Schlenker [2003], Anand and Nevins [2004], Anand [2006], Kratzer [2009], Ninan [2010]).

- It should be pointed out that it is a open question whether there are such things as "*de se* pronouns", strictly speaking. Given that first-person features are clearly neither necessary nor sufficient for the availability of *de se* readings, we should not assume without argument that first-person pronouns optionally give rise to such readings.

  (Specifically: it is not obvious that (16) has two distinct readings, one of which is *de se*, as opposed to merely having one reading which divides into two possible ways of being true.)

3 Stanley against the Lewisians: foundations

3.1 Mooring the notion of self-ascription

Stanley also has a criticism of a more foundational character, focused on Lewis's notion of self-ascription. He writes:

> The sense that the self-ascription framework provides an explanation of the *de se* is due merely to the fact that it uses the *de se* vocabulary in the metalanguage. What was to be explained is the fact that "John wants to win" only has a *de se* reading, and cannot be true if John merely wants someone to win who in fact unbeknownst to John turns out to be John himself. The self-ascription framework claims to explain this fact, by appeal to the fact that John wants to self-ascribe a property. But this is not an explanation. What we desired was an explanation of what it means to self-ascribe, not just using that vocabulary in the metalanguage. (89)

But talk of "self-ascription" within Lewis's framework is a dispensable heuristic, not something carrying an explanatory load, and not something essential to understanding the proposal.

(Here it helps to compare the situation to the possible worlds model Lewis means to be upgrading.)

The foundational story for Lewis's model of *de se* content builds on the idea that we are constitutively rational: the intentional mental states of an agent have the content that they have largely in virtue of the fact that the dispositions of the agent to act can be explained as rational on the hypothesis that the agent's mental states have that content.

> For an agent x to be in a total state of belief whose content is given by the set of centered worlds P is (at least) for x to be disposed to act in ways that would tend to satisfy x's desires in worlds w such that $\langle w, x \rangle \in$ P.[2]

Crude, obviously, but it does tells us a great deal about what it is supposed to mean for a centered world to be compatible with one's beliefs—what it means, on Lewis's model, to self-ascribe a property. It is (largely) to be in a state of mind with this sort of functional structure.[3]

---

[2] Note added 2/19/2013: this statement is confused. The following is closer to what I had in mind: "To be believe B and desire D, where B and D are sets of centered worlds, is (at least) to be disposed to perform the sorts of actions which would, if performed by an agent located in B, tend to bring it about that that agent is in D." (Thanks to Herman Cappelen and Josh Dever for calling my attention to this mistake.)

[3] The "largely" leaves room for the role of eligibility in Lewis's account: "Self-ascribed properties may be 'far from fundamental'... but not too far. Especially gruesome gerrymanders are prima facie ineligible to be contents of belief and desire" ([Lewis, 1994, 320]).

As compared to the possible worlds model, the centered worlds model serves to widen the array of possible functional/behavioral states agents can be in, enabling the possibility of behavioral differences between pairs of agents who take the world to be the same way and who also want the world to be the same way.

This is the part of Lewisian story which addresses the question what it means to be in a state of mind with centered-worlds content (or indeed, any content). Talk of self-ascription is not doing work here, and nor is it intended to do work; on the contrary, such talk is itself moored in the same general way.

Stanley's claim that self-ascription talk isn't explanatory is thus correct. But the claim is intended as an objection, and in this respect it is misguided.

3.2 Thinking of oneself as oneself

The preceeding sketches what I take to be Lewis's approach to explaining what self-locating content is. This brings us to another aspect of Stanley's foundational criticism. Stanley claims that any account of what it is to self-ascribe a property "will include an account of what it is to think of something as oneself". He reads Lewis as shirking off this mandatory question.

The question seems obscure as stated, however, and I don't see why the defender of Lewis's view is obligated to address it.

Still, we can ask, does the Lewisian story contain something which could be called an account of what it is to think of something as oneself?

If the question is whether it supplies an account of some putative mode of presentation of the self figuring as a constituent in *de se* thought, then the answer is obviously negative, but to expect that would of course be to beg the question.

It does, however, give an account of what it amounts to to have beliefs *de se*: it is to think thoughts which are properties, properties which distinguish between individuals in at least some worlds. It is to think thoughts whose correctness turns on where one is in fact located in in the space of possibilia. And what it is to think thoughts like these is in turn fundamentally explained by Lewis's brand of functionalism about such states, and the associated metasemantic story, sketched already above.

4 Stanley's Fregeanism about the *de se*

Stanley's Fregean account is, compared to predicational rivals, very underspecified. Stanley wants to remain agnostic on significant details of the formal semantics. This is because he takes himself to have a highly limited objective: he only wishes to establish that *de se* ascription can be explained with some or other

Fregean analysis of *de re* ascription. But he wishes to remain agnostic on how exactly the Fregean should model *de re* ascription. His objective is to show that the *de se* creates no problems for the Fregean beyond those already presented by the *de re*.

But it is not the case what Stanley does say about the *de se* is compatible with just any Fregean account of the *de re*. His view works only with certain Fregean accounts of the *de re*. Thus he tacitly ends up with commitments about how the Fregean must analyze the *de re*—commitments that go largely undefended.

It is, moreover, arguable that the sort of Fregean account of the *de re* Stanley ends up with is less plausible than rivals Fregean accounts, accounts which would *not* underwrite his commitments about the *de se*.

Two of Stanley's central claims about controlled PRO attitude ascriptions:

1. The PRO_i in these constructions is coindexed with the matrix subject, and contributes a definite description which is analyzed as a generalized quantifier over first-person ways of thinking. PRO_i semantically contributes a function from monadic properties of senses to truth-values:

$$[\![\mathrm{PRO}_i]\!]^g = \lambda f.\iota x : \mathrm{SELF}_{g(i)}(x) \wedge f(x)$$

   $\mathrm{SELF}_{g(i)}$ is a function from the value of $g(i)$ (an individual) to the property of being identical to $g(i)$'s first-person mode of presentation. (Thus if $g(i)$ is Bekele, $\mathrm{SELF}_{g(i)}$ is the property of being identical with Bekele's non-publically accessible first-person way of thinking, and PRO_i would be roughly equivalent to something like "the first personal mode of presentation of Bekele.")

2. These attitude ascriptions are a species of *de re* ascription. Specifically, the descriptive quantifier introduced by PRO_i takes scope above the attitude verb and binds across it.

In explaining his Fregean picture, Stanley suggests that his preferred treatment of (1′) (repeated below) is structurally analogous to a the *de re* reading of (7) where 'his brother' takes scope between 'John' and 'thinks':

(1′) Bekele expects PRO to win the race.

(7) John thinks his brother is kind.

Both PRO and 'his brother' are quantifiers taking scope between their matrix subjects and the attitude verb, and they are both quantifying in.

If the analogy is to hold, however, 'his brother' in (7) must be a quantifier over senses, like PRO. But it is not clear why we should think this is plausible, or what semantic mechanism is supposed to trigger this interpretation. It seems difficult to reconcile with simple examples, for instance:

(8) The person who John thinks is a spy is bald.

Now one view at this point could be that any Fregean theory of the *de re* already has exactly these problems; hence the *de se* is not creating new issues here. I take it this is Stanley's view in *Know How*.

But this view is incorrect. One can give a Fregean theory of the *de re* that does not assume that 'his brother' is a quantifier over senses. Here is an example. For concreteness, assume a syntax for (7) like this:
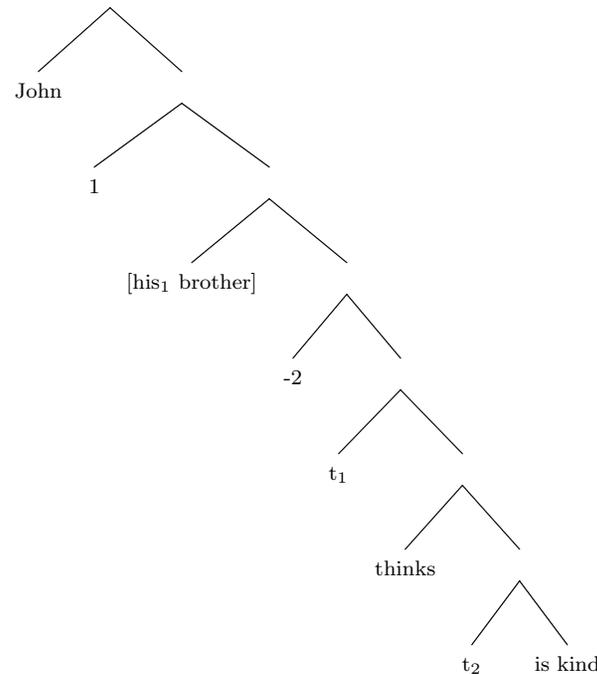
A syntax along these lines is fairly standard. The Fregean twist comes in the introduction of a new device of abstraction.

Fregean predicate abstraction

If α is a branching node with β and γ as daughters and β dominates only a negatively marked numerical index $-i$, then for any variable assignment a,

$[\![\alpha]\!]^a = \lambda x.\exists m : m$ is a mode of presentation of x and $[\![\gamma]\!]^{a^{m/i}}$.

This account does not require 'his brother' to express a quantifier over

4

senses; rather, we can suppose it quantifies over individuals, as it superficially appears to. So no special problems with (8).

Turning back to PRO, we can add that this account also seems to fair better with cases of subject control which do not obviously involve hyper-intensionality:

(9) John arrived PRO exhausted.

(10) The plant started PRO to die.

It of course remains to look at much more data, but *prima facie* this kind of Fregeanism about the *de re* strikes me as more promising.

Now the point I wish to make is that Stanley's Fregean account of the *de se* cannot be seen as a special case of this Fregean view of the *de re*. This is because Fregean predicate abstraction merely existentially quantifies over modes of presentation. It does not require, as Stanley's view of PRO does, that the relevant modes of presentation quantified over be first-personal in character. Thus one cannot apply the present Fregean account to the structure:

Bekele

1

PRO$_1$

-2

t$_1$

expects

t$_2$ to win
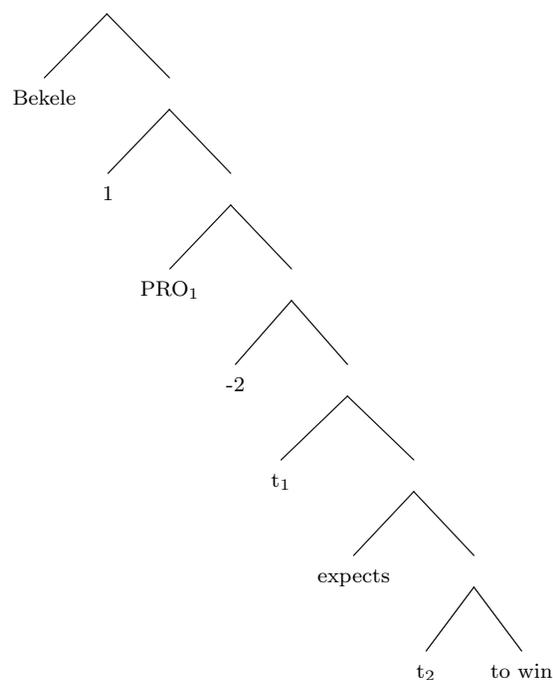
and get Stanley's desired result. If one wanted to adjust this Fregean account in the direction of Stanley's truth-conditions, one would require at least an additional abstraction rule dealing explicitly with the movement of subject control PRO out of attitude contexts. At this stage, one begins to notice (if one hasn't already) how much simpler the predicational view looks from a semantic point of view.

Stanley hoped to motivate a Fregean view of the *de se* without taking a stand on the proper Fregean analysis of the *de re*. But these points suggest it is just very difficult to say anything substantive about the shape of a Fregean view of the *de se* without taking on commitments about the proper Fregean analysis of the *de re*.

So Stanley's discussion seems to have a lacuna. Given his objectives, the absence of concrete detail in *Know How* about how the view is to be semantically implemented is not justified. And without the details, it is hard to see how the view can yet be reasonably seen as a serious competitor to the versions of the predicational view currently available (again, for instance, Anand and Nevins [2004], Anand [2006], Kratzer [2009], Ninan [2010], Stephenson [2010], among others).

References

Pranav Anand. *De De Se*. PhD thesis, Massachusetts Institute of Technology, 2006.

Pranav Anand and Andrew Nevins. Shifty operators in changing contexts. *Proceedings of SALT XIV*, 2004.

R. Bhatt. *Covert modality in non-finite contexts*, volume 8. Mouton de Gruyter, 2006.

A. Kratzer. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry*, 40(2):187–237, 2009.

David K. Lewis. Attitudes de dicto and de se. *Philosophical Review*, 88(4):513–43, 1979.

David K. Lewis. Reduction of mind. In S. Guttenplan, editor, *A Companion to the Philosophy of Mind*. Blackwell, 1994.

Dilip Ninan. De se attitudes: Ascription and communication. *Philosophy Compass*, 5 (7):551–567, 2010. ISSN 1747-9991.

Philippe Schlenker. A plea for monsters. *Linguistics and Philosophy*, 26:29–10, 2003.

J. Stanley and T. Williamson. Knowing how. *Journal of Philosophy*, 98:411–444, 2001.

T. Stephenson. Control in centred worlds. *Journal of semantics*, 27(4):409–436, 2010.