

Initialization and Self-ascription

Daniel Skibra

EMAIL:danielskibra@gmail.com

Zukunftskolleg, University of Konstanz

ECAP11

University of Vienna, August 25, 2023

LINK: <https://tinyurl.com/SkibraECAP11>



Roadmap

The ideal roadmap for today's talk

- 1 Review about the *de se*
- 2 Lewis's account of the *de se*
- 3 a recurring objection to Lewis's account
- 4 a recent defense of Lewis
- 5 some issues the defense brings into view

Roadmap

The realistic roadmap

- 1 ~~Review about the *de se*~~
- 2 Lewis's account of the *de se* (*quickly!*)
- 3 a recurring objection to Lewis's account
- 4 a recent defense of Lewis
- 5 some issues the defense brings into view

This might fit.

A bit of review about the de se

A familiar scenario (Cf. Perry 1979) :



Messy shopper: Perry is shopping in a supermarket and comes across a trail of flour*. He rightly suspects that a shopper's sack of flour has broken or tipped and is spilling flour onto the floor. So he gives chase, following the trail through several aisles in search of the Messy Shopper. After turning up empty, it occurs to him that *he* is in fact the Messy Shopper, at

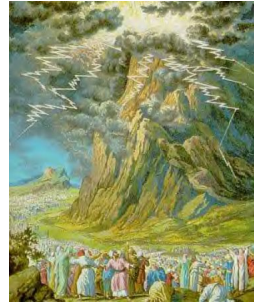
which point he looks to his cart for the offending sack of flour and fixes it upright.

What does Perry come to believe?

A bit of review about the de se

Another familiar scenario (Cf. Lewis 1979):

Two gods: There are two gods living on different mountains in a possible world they co-inhabit. One lives on top of the tallest mountain and throws down manna; the other lives on the coldest mountain and throws down thunderbolts. They are both omniscient in the sense that they know every (possible world) proposition that is true in their world. But each one can still be ignorant about which of the mountains he inhabits.



What's the belief the gods fail to have?

A bit of review about the de se

Several ways of characterizing **what's distinctive** about the kind of attitude(s) illustrated by the previous scenarios. The attitudes...

- seem to be *about oneself* ...,
- are best *characterized in indexical terms* ...,
- are related to the *explanation of action* in some privileged way...

A bit of review about the de se

Several ways of characterizing **what's distinctive** about the kind of attitude(s) illustrated by the previous scenarios. The attitudes...

- seem to be *about oneself* ...,
- are best *characterized in indexical terms* ...,
- are related to the *explanation of action* in some privileged way...

But also — hard to characterize *content* of those attitudes on traditional models of attitude content.

A traditional view of content

Possible worlds account of *belief* (Hintikka 1962; Stalnaker 1984):
 \rightsquigarrow proposition := set of possible worlds

‘S believes p’ is true just in case $\forall w \in \text{DOX}_S : p(w) = 1$,
where DOX_S is the set of worlds compatible with S’s total belief
state.

Problem: given some plausible assumptions (e.g., about
indexicals),

- Perry believes the **same thing** *before* he realizes who is making
the mess as he does *after*!
- there is **no way** to characterize the ignorance of either god!

Lewis's (1979) proposal – property contents

DISCLAIMER: I'll be focusing on Lewis's account of the *de se*. (Of course, there are others.)

Aims to maintain an unstructured account of attitude content, as one that *carves up possibilities*. But, still needs carve up those possibilities *more finely* than possible worlds.

Solution: property contents.

When a person believes *de se* that they are p , they *self-ascribe* the *property* of being p .

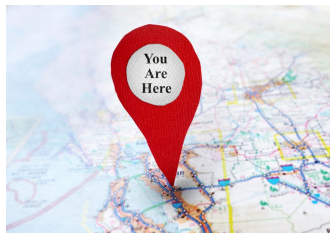
- Before realization, Perry believed the proposition that *that shopper is making a mess*.
- After realization, Perry self-ascribed the property of *making a mess*.

From property contents to centered worlds

We model properties as sets of world-bound objects. Formally: we add to worlds a **center** (usually, an individual-time pair, $\langle x, t \rangle$)

Typical metaphor invoked:
a map with a “you are here” marker.

Self-ascribing property P (at a time) is to locate one's center amongst those objects that have P (at that time).



the Property View

The schema for belief content on the property view is as follows:

- for S to believe *de se* that P is for the content of S 's belief to be:
 - $\{\langle w', \langle x, t' \rangle \rangle \mid P(x, t', w')\}$,
 - or alternately, using lambda-notation: $\lambda x. \lambda t'. \lambda w'. P(x, t', w')$.

Adverting to the earlier belief-reporting schema:

$\lceil S$ believes (*de se*) that $p \rceil$ is true just in case

$\forall \langle w', \langle x, t' \rangle \rangle \in \text{DOX}_S : p(w', x, t') = 1$

where DOX_S is the set of worlds compatible with S 's total belief state.

Messy shopper & two gods

We can now characterize the difference in attitude, *and in attitude content*.

- At first:
 $\forall w' \in \text{DOX}_{JP} : \text{that person}_c \text{ is making a mess in } w'$
- After realization:
 $\forall \langle w', \langle x, t' \rangle \rangle \in \text{DOX}_{JP} : x \text{ is making a mess at } t' \text{ in } w'$

Similar remarks apply to the Two Gods case and allow us to characterize their ignorance.

The de se and the de dicto

But: Lewis's property view is supposed to be *general*, not just for de se attitudes. No problem! Make all attitudes sets of centered worlds. It's just the de dicto ones involve vacuous binding.

More precisely:

- At first:

$\forall \langle w', \langle x, t' \rangle \rangle \in \text{DOX}_{JP} : \text{that person}_c \text{ is making a mess in } w'$

- After realization:

$\forall \langle w', \langle x, t' \rangle \rangle \in \text{DOX}_{JP} : x \text{ is making a mess at } t' \text{ in } w'$

Conceptually: belief de dicto is locating oneself in logical space, belief de se involves locating oneself inside the borders of the world. (Cf. Egan 2006 on interesting and boring centering)

Looming objections?

Two observations:

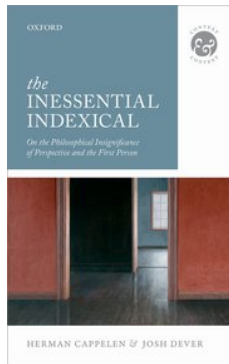
- ① Notoriously hard to give an account of what is distinctive about first personal thought. What is this relation to oneself?
- ② Lewis seems to side-step this problem. No need to say what this relation is; just functionally define it.

So, we're left with an account where *all* attitudes are kinds of *self-ascription* (modulo direction of fit).

Primitiveness: Capellen and Dever 2013

“So to understand the difference between knowing that Zeus is on the tallest mountain and knowing that he is on the tallest mountain, **we need to understand the difference between (for Zeus) Zeus-ascribing and self-ascribing a property.**

But that difference is entirely extrinsic to the theory of content Lewis is proposing.”
(Capellen and Dever 2013:103)



Primitiveness: Stanley 2011



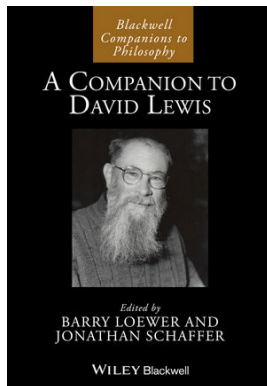
“A principal attraction of a view that treats first-person thought as self-ascriptions of properties is the sense that it appears to allow us to evade the extraordinarily difficult task of explaining what it is to take a first-person perspective on the world. [...] **The self-ascription framework claims to explain [de se readings of want-PRO ascriptions] 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.

(Stanley 2011:88–89)

Primitiveness: Holton 2015

“To self-ascribe a property is not just to ascribe a property to oneself, as one might ascribe it to someone else. That would just give us back the question of what it is to think of oneself in the right way.”

“The idea of primitive self-ascription is an obscure one. Our natural grasp on it is via the general idea of ascribing a property to an object, but this is exactly what we are not allowed to do here.” (Holton 2015:400)



What's the problem?

NB: There's probably more than one objection reflected in these remarks. I'll focus on the most immediately salient.

- Self-ascription (and liberal variable binding) allows Lewis to do give unified account of *de dicto* and *de se* belief.
- But, the complaint goes: we don't know what self-ascription is. It's a primitive of the theory.

Defense via explanation

Openshaw's (2020) defense: These complaints are unfounded. Lewis **does** give us (indirectly) an explanation of self-ascription.

- self ascription is explained by way of the mechanism of initialization that is made use of in the two-dimensional semantics that Lewis advocates for natural language.

TERMINOLOGICAL ASIDE; Lewis doesn't use the term *initialization*, as far as I know, which comes from Belnap et al. 2001. But a very similar idea appears in Lewis's work.

Background: context-index semantics

Lewis (1980) follows Kaplan 1977 in modeling the context-sensitivity of natural language by means of a two dimensional framework involving both *context* and *index* parameters.

- context (*c*) is a sequence of coordinates used to determine the values of indexical expressions (*I*, *here*, or *now*, etc.)
- index (*i*) is a sequence of coordinates that serve jointly as circumstances of evaluation, and can individually be shifted by certain expressions (e.g., modals, tense, etc.)

Why we need both c and i

A sentence may fall in the scope of an index-shifting operator, and yet contain indexicals, whose values are still recovered from the context of utterance, even when the circumstance of evaluation is shifted.

- $\llbracket \text{I was here} \rrbracket^{c,i}$ is true just in case
 - PAST (*I be here*) is true at c, i
 - PAST (*Speaker_c be at location_c*) is true at c, i
 - (*Speaker_c be at location_c*) is true at some i' just like i but w/
 t' of $i' \prec t$ of i

Notice: indexical expressions get values from context even when in the scope of index-shifting operators.*

More about c and i

What's in c ? What's in i ?

- If you're Kaplan:
 - $c \rightsquigarrow$ depends on what indexical expressions there are
 - $i \rightsquigarrow \langle w, t \rangle$ (also: depends on what operators there are)
- If you're Lewis:
 - $c \rightsquigarrow$ time slice of an individual (get sequence from this)
 - $i \rightsquigarrow$ *at least*: worlds, times, standards of precision;
 $\langle w, t, s, p_1, \dots, p_n \rangle$

More about c and i

What's in c ? What's in i ?

- If you're Kaplan:
 - $c \rightsquigarrow$ depends on what indexical expressions there are
 - $i \rightsquigarrow \langle w, t \rangle$ (also: depends on what operators there are)
- If you're Lewis:
 - $c \rightsquigarrow$ time slice of an individual (get sequence from this)
 - $i \rightsquigarrow$ *at least*: worlds, times, standards of precision;
 $\langle w, t, s, p_1, \dots, p_n \rangle$

Semantic value, truth, content

In this “context-index” framework: the input to compositional semantics is a sentence S and a context c , and the output is an assignment of truth values to the semantic value of S at every index i . How do we get to *content* from *truth at every index*?

Who cares??

Why does this matter?

- There's been considerable hand-wringing about this in philosophy of language.
- Recall that Lewis liked unstructured theories of content \rightsquigarrow sets of possible worlds (at least, to begin with).
- But: output of semantics is *not* a set of possible worlds.

What to do???

Who cares??

Lewis (1980): *Don't worry about it!!*

- “we can define the propositional content of sentence s in context c as that proposition that is true at world w iff s is true at c at the index i_c^w that results if we take the index i_c of the context c and shift its world coordinate to w .” (Lewis 1980:94)
- We can get **content** from **semantic values** by giving the index coordinates the values from c .

Initialization and “original indices”

We need context to play this role *anyway*

- shifted index coordinates need to be shifted *from* some value (“otherwise the mobile [*shiftable*] parameters would be useless” Belnap et al. 2001:148–149)
- Because the context determines these initial index values, Belnap et al. talk of context “initializing” the index.
- In subsequent remarks, Lewis called these initialized indices, “original indices” (“...in which the shiftable features are as determined by the context which appears as first coordinate” (cf. Lewis 1983:231))

Back to self-ascription

This is the mechanism Openshaw appeals to as that which would help provide an explanatory basis for Lewis's notion of self-ascription:

“Talk of self-ascription is then just an intuitive gloss on work performed by the stipulative process of initialization...”

“[Self-ascription] is a phrase used to express the relation that a believer, S , bears to a property-content which is true at c where S is the subject of c .” (Openshaw 2020:§3)

Not done yet... I

There is still a conceptual gap between the mechanism of *initialization* and the “*relation a believer... bears to a property content...*”

Let's close the gap

- initialization is an instruction to take a semantic item and convert it to another one – namely one that can serve as content.
- self-ascription tells you how to evaluate content (once you have an item of content)

Not done yet... II

Let's close the gap

- Given initialization, though, we can just take context as default circumstance of evaluation.
- **The result:** when we allow contents to be **properties**, it is a set of circumstances of evaluation which look like “possible contexts”
- Since contexts for Lewis are (time-slices) of individuals, we get a set of possible individual time-slices

Self-ascription explained?

So, are we all good??

- ✓ definition of self-ascription
- ✓ explanation of self-ascription that doesn't rely on relation to self
- ✓ general account of attitudes



This is looking pretty good.

Wait!! Not so fast!

I think Openshaw is basically right about how Lewis intended us to understand self-ascription. But the “primitiveness” people still have reason to worry.



There's a **general** theoretical worry, and a more **specific** worry.

General worry I

First, the general worry.

This...

Given initialization, though, we can just take context as default circumstance of evaluation...

...is actually a pretty substantive commitment.

- Recanati (2007) calls this the “Reflexivity constraint”
- Is this commitment justified? What’s the source of this commitment?

General worry II

- Remember: initialization is really just a mechanism for converting one semantic object into another.
- It's a separate question *what values to initialize with*.
- Initialization only helps explain self-ascription if *context* gives us default circumstance of evaluation.

The “primitiveness” people can still legitimately complain we don't have an explanation of *that*.

Specific worry I

The specific worry

- Account is supposed to be *general*.
- Holds for *de dicto* as for *de se* belief. (And, *de dicto* is just a special case of the *de se*.)
- Also general across **all attitudes**: belief, desire, imagination, etc. They would all have some analog of self-ascription (*modulo* direction of fit) as the relation the attitude holder has to content.

This commits the account to a **prediction**: *all* content is evaluated with respect to context-qua-circumstance of evaluation. Call this the “proximal circumstance of evaluation”. Are there really no attitudes which evaluate content at “distal circumstances of evaluation”?

Specific worry II

Final remarks are impressionistic, but suggest (I think) that [there are distal circumstances of evaluation](#). (*If so*, then self-ascription cannot be basic in the way Lewis needs.)

Consider **desire**: what is the analog of *self-ascription* for wanting?

- Let's call it "self-prescription".
- Self-prescribing a property P means something like: preferring P 's obtaining at the context (to it not obtaining/ to an alternative obtaining).
- Desire contents still look like satisfaction conditions.
- Looks OK so far....

Different kinds of desires I

But:

Consider desires whose **satisfaction is conditional on their persistence**. (Cf. Gordon 1986; Parfit 1984; McDaniel and Bradley 2008)

- Example: I want to eat an Apfelstrudel.

Then, it gets so hot that I lose my appetite. If I ate the Apfelstrudel anyway, the desire wouldn't be satisfied. Why? I need to still have the desire for it to be satisfied.

Different kinds of desires II

There are also desires whose **satisfaction is not conditional on their persistence**. (Parfit 1984)

- Example: I want my children's lives to long surpass mine.

By definition, if these satisfaction conditions obtain, my mental state won't then persist.

Lots of reasons to worry about such desires if you are a moral philosopher. For our purposes, suffices to say it looks like a case of an attitude with a distal circumstance of evaluation.

Other potential examples

- “selfless” desires (Nolan 2006)
- “counterfactual” *de re* attitudes (Ninan 2012)
- episodic memory (Recanati 2007)

Why this matters: If there *are* attitudes with content evaluated as distal circumstances of evaluation, then self-ascription *cannot* be basic/ default relation an attitude holder has to attitude content.

Fin.

Thank you!

References I

- Belnap, N. D., M. Perloff, and M. Xu (2001). *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford University Press on Demand.
- Capellen, H. and J. Dever (2013). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.
- Egan, A. (2006). Secondary qualities and self-location. *Philosophy and Phenomenological Research* 72(1), 97–119.
- Gordon, R. M. (1986). The circle of desire. In *The Ways of Desire*, pp. 101–114. Precedent.
- Hintikka, J. (1962). *Knowledge and Belief*. Ithaca: Cornell University Press.
- Holton, R. (2015). Primitive self-ascription: Lewis on the de se. In B. Loewer and J. Schaffer (Eds.), *The Blackwell Companion to David Lewis*. Blackwell.
- Kaplan, D. (1977). Demonstratives. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes From Kaplan*, pp. 481–563. Oxford University Press.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review* 88(4), 513–543.

References II

- Lewis, D. (1980). Index, context, and content. In S. Kanger and S. Öhman (Eds.), *Philosophy and Grammar*, pp. 79–100. Reidel.
- Lewis, D. (1983). Postscripts to *Attitudes de dicto and de se*. In *Philosophical Papers, Volume 1*, Volume 1. Oxford: Oxford University Press.
- McDaniel, K. and B. Bradley (2008). Desires. *Mind* 117(466), 267–302.
- Ninan, D. (2012). Counterfactual attitudes and multi-centered worlds. *Semantics and Pragmatics* 5(5), 1–57.
- Nolan, D. (2006). Selfless desires. *Philosophy and Phenomenological Research* 73(3), 665–679.
- Openshaw, J. (2020). Self-ascription and the *de Se*. *Synthese* 197(5), 2039–2050.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Perry, J. (1979). The problem of the essential indexical. *Noûs* 13(December), 3–21.
- Recanati, F. (2007). *Perspectival Thought: A Plea for Moderate Relativism*. Oxford University Press.

References III

Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.

Stanley, J. (2011). *Know How*. Oxford: Oxford University Press.