Workflow: Annotated pdf

PROOF COVER SHEET

Author(s): Manuel García-Carpintero
Article title: The philosophical significance of the De Se
Article no: SINQ 1262003
Enclosures: 1) Query sheet
                2) Article proofs

Dear Author,

Please find attached the proofs for your article.

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication

Please limit changes at this stage to the correction of errors. You should not make trivial changes, improve prose style, add new material, or delete existing material at this stage. You may be charged if your corrections are excessive
(we would not expect corrections to exceed 30 changes).

For detailed guidance on how to check your proofs, please paste this address into a new browser window: http://journalauthors.tandf.co.uk/production/checkingproofs.asp

Your PDF proof file has been enabled so that you can comment on the proof directly using Adobe Acrobat. If you wish to do this, please save the file to your hard disk first. For further information on marking corrections using Acrobat, please paste this address into a new browser window:http://journalauthors.tandf.co.uk/production/acrobat.asp

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your names will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Manuel	García-Carpintero	

Queries are marked in the margins of the proofs, and you can also click the hyperlinks below.

AUTHOR QUERIES

General points:

1. **Permissions:** You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. For further guidance on this topic please see:
   http://journalauthors.tandf.co.uk/copyright/usingThirdPartyMaterial.asp

2. **Third-party material:** If there is material in your article that is owned by a third party, please check that the necessary details of the copyright/rights owner are shown correctly.

3. **Affiliation:** The corresponding author is responsible for ensuring that address and email details are correct for all the co-authors. Affiliations given in the article should be the affiliation at the time the research was conducted. For further guidance on this topic please see: http://journalauthors.tandf.co.uk/preparation/writing.asp.

4. **Funding:** Was your research for this article funded by a funding agency? If so, please insert 'This work was supported by <insert the name of the funding agency in full>', followed by the grant number in square brackets '[grant number xxxx]'.

5. **Supplemental data and underlying research materials:** Do you wish to include the location of the underlying research materials (e.g. data, samples or models) for your article? If so, please insert this sentence before the reference section: 'The underlying research materials for this article can be accessed at <full link>/ description of location [author to complete]'. If your article includes supplemental data, the link will also be provided in this paragraph. See <http://journalauthors.tandf.co.uk/preparation/multimedia.asp> for further explanation of supplemental data and underlying research materials.

6. The **CrossRef database** (www.crossref.org/) has been used to validate the references. Changes resulting from mismatches are tracked in red font.

AQ1	Please provide the missing city and country for affiliation.
AQ2	The spelling of Name for 'Böer and Lycan (1981)' has been changed to 'Boër and Lycan (1981)' to match the entry in the references list. Please confirm that this is correct and provide revisions if needed.

AQ3	Please note that the use of 'op cit' is against journal style. Please replace these with the relevant reference details throughout.
AQ4	The reference 'García-Carpintero & Torre (2016)' is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details following journal style.
AQ5	The reference 'Papineau (2007)' is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details following journal style.
AQ6	The reference 'Peacocke (2014)' is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details following journal style.
AQ7	The disclosure statement has been inserted. Please correct if this is inaccurate.
AQ8	If 'Biggs and Wilson (Forthcoming)' has been published, please give details for references list following journal style.
AQ9	The CrossRef database (www.crossref.org/) has been used to validate the references. Mismatches between the original manuscript and CrossRef are tracked in red font. Please provide a revision if the change is incorrect. Do not comment on correct changes.
AQ10	If 'García-Carpintero (Forthcoming)' has been published, please give details for references list following journal style.
AQ11	Please provide missing page number and volume number for the 'Guillot (2016)' references list entry.

How to make corrections to your proofs using Adobe Acrobat/Reader

Taylor & Francis offers you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can mark up the proof directly using Adobe Acrobat/Reader. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

1. Save the file to your hard disk.

2. Check which version of Adobe Acrobat/Reader you have on your computer. You can do this by clicking on the "Help" tab, and then "About".

   If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

3. If you have Adobe Acrobat/Reader 10 or a later version, click on the "Comment" link at the right-hand side to view the Comments pane.

4. You can then select any text and mark it up for deletion or replacement, or insert new text as needed. Please note that these will clearly be displayed in the Comments pane and secondary annotation is not needed to draw attention to your corrections. If you need to include new sections of text, it is also possible to add a comment to the proofs. To do this, use the Sticky Note tool in the task bar. Please also see our FAQs here: http://journalauthors.tandf.co.uk/ production/index.asp.

5. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. If you have more than one file, please zip them together and then upload the zip file.

   If you prefer, you can make your corrections using the CATS online correction form.

Troubleshooting

Acrobat help:http://helpx.adobe.com/acrobat.html
Reader help:http://helpx.adobe.com/reader.html

Please note that full user guides for earlier versions of these programs are available from the Adobe Help pages by clicking on the link "Previous versions" under the "Help and tutorials" heading from the relevant link above. Commenting functionality is available from Adobe Reader 8.0 onwards and from Adobe Acrobat 7.0 onwards.

Firefox users: Firefox's inbuilt PDF Viewer is set to the default; please see the following for instructions on how to use this and download the PDF to your hard drive:
http://support.mozilla.org/en-US/kb/view-pdf-files-firefox-without-downloading-them#w_using-a-pdf-reader-plugin

Routledge
Taylor & Francis Group

The philosophical significance of the *De Se*

Manuel García-Carpintero

Departament de Lògica, Història i Filosofia de la Ciència, Universitat de Barcelona   AQ1

ABSTRACT
Inspired by Castañeda, Perry and Lewis argued that, among singular thoughts in general, thoughts about oneself 'as oneself' – first-personal thoughts, which Lewis aptly called *de se* – call for special treatment: we need to abandon one of two traditional assumptions on the contents needed to provide rationalizing explanations, their *shareability* or their *absoluteness*. Their arguments have been very influential; one might take them as establishing a new 'effect' – new philosophical evidence in need of being accounted for. This is questioned by the skeptical arguments in recent work by Cappelen & Dever and Magidor, along lines that a few discrepant voices had already announced earlier. Skeptics content that the evidence does not really call for revising traditional theories of content. I will discuss their challenges – first and foremost, concerning action explanations – aiming to make the case that the '*De Se* effect' is no illusion: *de se* attitudes require us to revise one of the two tenets of traditional views.

1. Preamble: the *De Se* 'Effect'

Inspired by Castañeda (1966, 1968), Perry (1979) and Lewis (1979) argued that, among singular thoughts in general, thoughts about oneself 'as oneself' – first-personal thoughts, which Lewis aptly called *de se* – call for special treatment, in a way I explain below. Their arguments have been very influential. If I may allow myself an autobiographical aside, I usually mention them as good examples

CONTACT Manuel García-Carpintero ✉ m.garciacarpintero@ub.es

of progress in philosophy, with the following considerations. Sometimes scientific progress comes not so much from advancing new theories as from establishing new 'effects' – as psychologists call new evidence (*new* relative to the evidence and views held at the relevant time) in need of being accounted for. New *explananda* can in fact be even more lasting than new theories. In physics, for instance, Archimedes' alleged discovery that gold is soluble in *aqua regia* has survived as evidence to be explained by an adequate chemical theory much longer than the theories put forward to account for it. Similarly, although the tides were of course known in the eighteenth century, they were only classified as evidence together with other gravitational effects by Newton, and this (not just the tides, but the relatively theory-laden classification) has survived longer than his own account. Thus, having their name associated with a new effect, such as the McGurk effect or the Müller-Lyer illusion, may be something that researchers covet more than providing a theoretical account for it.

In philosophy, we can find illustrations of the same phenomenon. Kripke's (1980) rigidity data, or his Gödel–Schmidt thought experiment for his 'semantic' argument are good examples; they are here to stay, whatever one thinks of the different 'direct reference' theories purporting to account for them. For a second illustration, consider a footnote in that very work, in which Kripke (1980, 51) complains: 'Logicians have not developed a logic of vagueness'. The work of Fine, Kamp, Williamson and others has mended this in the ensuing years. But again, I have more confidence in the 'effects' that have been discovered through these debates (such as the back-and-forth intuitions regarding where cut-off points lie in 'forced-March sorites' cases, pointed out by Raffman (1994)), than in any of the theories advanced to explain them, even the one I take to be correct. In some cases (Gettier comes to mind) very influential philosophical 'effects' are made public without even mentioning any supporting theory.

It is against this background that I think Lewis' and Perry's work can be mentioned as a good example of progress in philosophy: it establishes a new 'effect'. But this is questioned by the skeptical arguments in recent works such as Cappelen and Dever (2013)

AQ2

– 'C&D' henceforth – and Magidor (2015).[1] Let me elaborate on what I take to be the relevant 'effect' at stake in these debates. In order to describe *explananda*, we need characterizations as independent as possible of the extant rival theories purporting to account for them; we thus need a characterization compatible with the claims of *de se* skeptics, like the aforementioned authors, who would otherwise deny the very existence of *de se* attitudes. I suggest we adopt the sort of approach that Hill (2009, 19–22; cf. also Kriegel (2015, 52–53)) advances when confronting a similar problem – that of providing an initial characterization of what theories of *qualia* are about, acceptable even to qualia-skeptics who would otherwise deny their very existence.[2] Since Nagel (1974), philosophers have typically resorted here to 'what it is like' locutions, but it is unclear that this turn of phrase, in its ordinary usage, can adequately serve the purpose.[3] Hill instead introduces *qualia* as properties *it is reasonable to think of* as fully graspable only from the perspective of the conscious subject, such as pains, orgasms or tastes, so that such and such arguments (here specific examples should be provided: those in Kripke (1980) or Jackson (1986) will do) have been advanced to argue for their irreducibility to the physical. What we might call the 'qualia effect' is then the (relatively theoretical) one of making it apparent (or reasonable) that the mind cannot be reduced to the brain and that physicalism is false.

Similarly, let us take *de se* attitudes to be those typically expressed by means of sentences including first-personal pronouns, such as (here paradigm examples should be mentioned, like those in Perry's Lingens', Heimson's, Messy Shopper's, and Lewis's Two Gods thought-experiments), so that philosophers including Perry and Lewis have argued that traditional theories of content could not properly characterize them. By analogy with the previous case, what we might call the '*De Se* effect' will be that of making it apparent (or reasonable) that certain attitudes lack *objective* contents: they lack either *absolute* or *shareable* contents (in a

---

[1]Although (as C&D (2013) exhaustively document) most philosophers share my appraisal of Lewis' and Perry's contribution, as usual in philosophy the consensus is far from universal, and some of the more recent skeptics' critical points were anticipated by, among others, Boër and Lycan (1981), Stalnaker (1981) and Millikan (1990).

[2]Ninan (2016, 89–90) has a related discussion and similar proposal to the one I'll make. The example of *qualia* is not just a convenient one; as it will become clear, phenomenal consciousness plays an important role in my discussion.

[3]Cf. Snowdon (2010) for a good critical discussion.

sense momentarily to be developed), so that 'content-objectivism' is false.

In Hill's account, *qualia* are not said to be 'subjective' (fully grasp-able only from a first-personal perspective), merely that they are typically *thought* to be so; and they are not said to be irreducible to physical properties – it is merely that some have argued that they are so, on the basis of their alleged subjectivity. Similarly, we do not exclude *de se* attitudes to have objective contents, nor consequently to be also expressed, even typically expressed, by non-first-personal sentences – nor do we assume that thinkers to which they occur need to express them; and we do not assume Lewis's and Perry's claims on contents to be correct. Hence, we are not begging any questions against the skeptics. For I take the main skeptical view in this area to be that these claims by Lewis and Perry *are wrong*. According to them (C&D 2013, 16; Magidor 2015, 272), we can explain all we need to explain about propositional attitudes, including *de se* attitudes, by assigning them contents as traditionally understood; as the latter puts it: 'the category of de se attitudes (if there is one) does not … require any special amend-ment of our general account of propositional attitudes' (Magidor 2015, 272).

What are the relevant features of this traditional understanding of thoughts? (I am using this notion interchangeably with *propo-sitional attitude*.) Perry (1979, 36) stresses two (given the assump-tion that they have contents, *propositions* to which the subject holding the attitude is related): (i) they have *absolute* truth values (given a full description of the world, they can be assigned one), as opposed to having truth values only relative to subjects or times; and (ii) they are *shareable* – if constituted by concepts, these, like Fregean senses, can be fully grasped by different individuals. The conception of propositions as classes of possible worlds meets these conditions, and it is the explicit target of Lewis (1979, 515); but there are other views on propositions that also meet them, such as Fregean accounts. Ninan (2016, 100–102) mentions two considerations to explain why propositions have been tradition-ally understood to have those two features: they allow for a nat-ural characterization of agreement in belief (i.e. holding beliefs with the same contents) and the lack thereof; and they allow for non-vacuous generalizations when propositional attitudes such as

SINQ 1262003
21 November 2016
Initial
CE: XX   QA: MS
Coll:XX   QC:XX

INQUIRY   5

beliefs, desires, and intentions are mentioned in the explanation of action (*ceteris paribus*, all those with the same attitudes would act equally).

Perry's and Lewis's work (anticipated by Castañeda) were thus produced against a theoretical background that assumed a well-motivated account of what propositions are like, and they provided considerations based on ingenious examples which, according to them, proved these assumptions wrong. This is what I take to be the philosophical 'effect' they established, *the De Se effect* – the illustration of philosophical progress their work provides. It will be my main contention in this paper that it *is* an effect (so that we can safely continue to use it as an example of philosophical progress), not – as the critics argue – an illusion.

Perry and Lewis came up with different theories to account for the *De Se* effect. Lewis's abandons feature (i) of traditional propositions. He takes contents to be *properties* instead; Sosa (1981) and Chisholm (1981) advanced essentially the same view. Contents would be entities that are true or false, given a full characterization of a way for the world to be, only relative to a subject and a time. Alternatively, the contents of propositional attitudes are, or at least select, not classes of worlds, but rather classes of *centered worlds*: worlds together with a designated subject and time. For its part, Perry's theory abandons the second feature: we should distinguish, he contends, the *content* of a belief from the *belief-state* through which it is accessed. The content is a proposition – say, that Perry is making a mess. The state is a condition of the subject, by being in which the content is believed. Contents respond to the role that propositional attitudes constitutively have in appraising the rationality of the subject, the adequacy of his beliefs to his evidence and of his actions to his beliefs and desires, the value of his desires, etc. But only, on Perry's view, in a coarse-grained way: a full account of rational action must mention not just contents, but also the state through which they are accessed; and states are mental particulars that cannot be shared by different individuals, or the same at different times.

In previous work (García-Carpintero 2016a), I have questioned the details of Perry's proposal, and provided an alternative one that is nonetheless very much in the same spirit. My main worry concerned the need to explain how it is that states have by themselves

the capacity to contribute to rational explanations along the lines that contentful states do.[4] Perry usually talks at this juncture about the different 'classificatory practices' we have for representational states/acts, relative to different purposes. He argues that states are mental particulars that may be classified by their 'official contents', but also by a plurality of other finer-grained propositional contents, useful for different explanatory purposes. This might suggest an unwelcome instrumentalist stance. We would like to understand why *de se* thoughts in themselves, given the constitutive features our account ascribes them, have the rationalizing virtues they are supposed to have.

On my proposal, summarized in Section 3 below, the appeal to states boils down to locating propositional attitudes in a specific network of other contentful attitudes, including some that cannot be shared. But the details of my own account of the *De Se* effect, or Perry's and Lewis's, do not matter here, and as suggested above I am much less confident about them than I am about the effect itself. So, this is my agenda for the rest of the paper. I will discuss challenges posed by the skeptics about three different issues that in my view help to make the case that the *De Se* effect is no illusion, first and foremost, action explanations that require *de se* attitudes; I will also mention the phenomenon of *immunity to error through misidentification*, and *logophoric* expressions (those which indicate *de se* attitudes) such as PRO in natural languages. I will sketch considerations for my own account that I have given elsewhere in more detail;[5] but the focus will lie in trying to provide answers to the skeptics, using my particular take on *de se* phenomena only to show one way (among others) in which the *De Se* effect might have come about.

2. Action explanation and the *De Se* effect

In the previous section, I have provided a characterization of *de se* thought neutral vis-à-vis the debates confronting believers in the *De Se* effect and deniers. Let me use henceforth 'genuine *de se* thought' or 'true *de se* thought' to refer to *de se* thoughts as they would be if the *De Se* effect were real, i.e. if a proper account of

---

[4]Cf. also Zemach (1985, 187) for this sort of criticism.
[5]Cf. García-Carpintero (2016a, forthcoming).

SINQ 1262003
21 November 2016
Initial
CE: XX QA: MS
Coll:XX QC:XX

INQUIRY 7

them should abandon either feature (i) or feature (ii) above in the traditional conception of the attitudes. I take the agency-based arguments for the essential role of first-personal thoughts (the famous thought-experiments involving characters such as Perry's Messy Shopper, Lingens the amnesiac lost in Stanford Main Library, Castañeda's amnesiac soldier reading his biography, or Lewis's two gods) to offer the best considerations for the existence of the *De Se* effect, i.e. for the genuineness of the *de se* thoughts that figure in them. This section is devoted to developing this contention.

Let us focus for concreteness on Perry (1979, 33) Messy Shopper case:

> I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. But I seemed unable to catch up. Finally it dawned on me. I was the shopper I was trying to catch.

Before his epiphany, the shopper has a belief about himself (under the individuating concept or mode of presentation *the shopper with the torn sack*) to the effect that he is making a mess. As Perry notes, he might also have a *de re* belief about himself, on most current ways of elaborating this notion, to the same effect:

> Suppose there were mirrors at either end of the counter so that as I pushed my cart down the aisle in pursuit I saw myself in the mirror. I take what I see to be the reflection of the messy shopper going up the aisle on the other side, not realizing that what I am really seeing is a reflection of myself.

None of this, however, is sufficient for him to have the *de se* belief that he would express in accepting 'I am making a mess', the one that leads him to rearrange the torn sack in the cart. The question is then whether this *de se* belief (which he might acquire from a benevolent shopper uttering to him, 'you are making a mess' while pointing to his torn sack) is a genuine *de se* one, as Perry argues.

Let me first outline the way I suggest that we should think of agency-based arguments, before going into the details. I think of them as modest, but nonetheless compelling causal (thought-) experiments. Imagine, for illustration, a situation in which there is a standing ball, B, which starts moving after being hit by another, A. We believe B's movement to have been caused by A's hit. I take

this to be an ordinary application of Folk Physics, underwritten by (a folk version of) Mill's method of difference (more on this presently). We compare the many situations in which neither A's hit nor B's movement occurs, with the situation in which both the
5 hit and the movement occur, and conclude that the hit makes the difference, thereby constituting the relevant causal factor of the movement: without it, the latter would not have occurred. I take this to be a straightforward ordinary application of the method for causal inference that Mill called *Method of Difference*,
10 'MD' henceforth – his 'second canon', which he considered uniquely reliable in making such inferences (Mill 1843, III.VIII.§2).[6]

Similarly, we compare (consciously putting ourselves in his shoes) the previous condition in which the Messy Shopper lacks the first-personal thought – even though he might well share
15 with his latter self any other *de dicto* and *de re* thoughts about himself, to the effect that he is making a mess – all other relevant folk-psychological conditions also obtaining (in particular, the pro-attitude to prevent sugar-spilling in supermarkets), with the situation in which he comes to have it (again, consciously putting
20 ourselves in his shoes), and conclude (crucially appealing to our conscious introspective awareness of our mental states in one and another condition) that a *genuine* first-personal thought is a required causal factor. I take the thought-experiments to make plausible the general point that a true *de se* thought is always caus-
25 ally needed to obtain a folk-psychologically complete behavior rationalization (which, following Davidson (1963), I take to be a form of belief/desire causal explanation), by applying Mill's MD in the causal-explanatory framework of Folk Psychology to a case involving a *phenomenal contrast* that we can consciously expe-
30 rience. Of course, we run the relevant causal experiments in our imagination; but this is neither here nor there, as we do the same with physical causal thought – in the way Galileo ran many of his celebrated 'experiments'. And it is not unheard-of to establish causal claims by the methods of philosophical inquiry: witness the already mentioned Davidson (1963) on reasons as causes, Grice

---

[6]We cannot go here into the thorny issue of what causal factors are, although I myself like to think of them in terms of laws, like Mackie's (1974) INUS conditions: Insufficient Necessary part of an Unnecessary Sufficient nomic condition for the occurrence of the effect.

(1961) on the causal theory of perception, or Putnam (1967) on mental states as functional roles.[7]

The two most important features of this sketch are (a) the appeal to Mill's MD, and, more specifically, (b) to a case involving a
5   phenomenal contrast. I will now develop these two features, trying to show how they help to answer the considerations in chapters three and four of C&D (2013) and the very similar ones in the first section of Magidor (2015). The messiness of the details (the several points at which the account can be challenged) suffices to show
10   that these skeptical works are among the truly valuable ones in their genre, forcing us to reflect much more carefully on significant philosophical issues than we would otherwise have done.

I need to make a preliminary point, before going into the details of the two features. C&D (2013) contrast action explana-
15   tions involving *de se* attitudes with others that don't, i.e. *impersonal* explanations (cf. Magidor 2015, 259). Thus, François might carry out the action whose content he would describe as *that I duck under the table* because he has the belief whose content he would express as *I am about to be shot* and the intention whose content
20   he would express as *that I not be shot*, together with a connecting instrumental belief. A corresponding impersonal explanation would replace 'I' with 'François' in the previous content-expressions: François believes that François is about to be shot, desires that François not be shot, which, given the connecting instrumental
25   beliefs, explains the action that François ducks under the table. C&D then suggest that their opponent would need to defend a claim they call the *Impersonal Incompleteness Claim*, IIC: *Impersonal action rationalizations are necessarily incomplete because of a missing indexical component*.

30   Now, C&D acknowledge at several points in their discussion that IIC might be considered a red herring (op. cit., 38), and this is exactly what I think. As the previous outline showed, in my view a good defense of my main claim needs to appeal to considerations based on Folk Psychology, which essentially concern rational
35   beings like us, perceiving and acting in the way we do: who knows

---

[7]Biggs & Wilson (forthcoming) convincingly argue that the use of abductive methods is compatible with the *a priori* justification of the thereby supported judgments; cf. also Bonjour (1998) and Peacocke (2004).

SINQ 1262003
21 November 2016

Initial

CE: XX   QA: MS
Coll:XX   QC:XX

10   M. GARCÍA-CARPINTERO

what sort of rationalizations are needed to account for the behavior of angels and Gods, or the beings with telekinesis powers that Millikan (1990, 727) invokes to make similar skeptical claims? I do not think we need to engage in unconstrained speculation about such matters. Fortunately, they (op. cit., 40) consider a more to the point restricted version of IIC, IIC2: *as a matter of deep physical necessity, indexical beliefs are needed in order for human beings, constituted as they are, to act*. Assuming a broad notion of the physical on which Folk Psychological facts are part of it, this is what I'll be concerned to defend – on the understanding that an 'indexical belief' is a true *de se* one.[8]

(a) For the application of Mill's MD in the strict way he envisaged (Mill 1843, III.VIII.§2), we need to compare two cases meeting these two conditions: (i) the candidate effect occurs in the second, *experimental* condition, but not the first, *control* condition; while (ii) the circumstances of the two cases differ *only* with respect to one factor: the presence of the candidate cause in the experimental condition but not in the control. Now, having confidence enough that condition (ii) obtains in any particular case involves a huge idealization (or leap of faith), which has led to a debate about whether MD is just an instance of *Inference to the Best Explanation*, IBE.[9]

We do not need to go into this debate here, but we'll come back to it because it bears deeply on our discussion. For I take it that both C&D and Magidor would question the truth of (ii) in our cases. They grant that the examples, in particular Perry's Messy Shopper, involve the comparison of two cases (the shopper before the epiphany – control condition – and after it – experimental condition) in which agents share all relevant attitudes toward propositions meeting the two traditional conditions of absoluteness and shareability, and which differ in the effect – only after the epiphany is the subject prepared to self-ascribe the state by uttering 'I am making a mess', to look for the torn sugar sack inside

---

[8]On a standard way of understanding two-dimensional semantics, the distinction between epistemic and metaphysical necessity it allows is not itself metaphysical (García-Carpintero and Macià 2006): it is not that there are epistemically possible worlds that are not metaphysically possible, or the other way around, just that we allow ourselves different epistemic resources (*a priori* or empirical) to access one or the other possibility. Likewise here: I should not be understood as allowing for non-folk psychological rationalizations that do not involve indexical beliefs, only as restricting myself to folk psychological resources in making claims about possible rationalizations. (Thanks to Daniel Morgan for this point.)

[9]Cf. Lipton (2004, 121–128), and Scholl (2015) for a good up-to-date discussion.

the cart, and so on. The application of MD to conclude that a true *de se* reason was present in the experimental condition requires, however, *that everything else is equal* – that *cetera* are *paria*. And this is what they reject:

> Everyone agrees that the representational level gives out somewhere. For [this] motivation […] to be persuasive, it needs to come with some reason to think that it gives out at exactly the right spot: while the self is still involved in the representation.[10] (C&D 2013, 43)

> More generally, an agent will have indefinitely many third-person beliefs and desires about the world that don't result in or rationalize action by the agent. So we need to explain which ones produce action and which ones don't; we need a filtering mechanism of some sort. Call this 'the Selection Problem.' […] One way to put our basic point is to say that the Selection Problem is solved by identifying a category of *actionable contents*. Most likely there isn't much systematic and general to say about that category. What constitutes actionable content for an agent will depend on the kinds of sensory inputs the agent has – for example, whether he moves around the world only with smell and taste or using sonar and sounds, eyes and ears and fingers, etc. It will depend on the agent's physical setup: whether he has fingers, claws, a trunk, wings, feet, fins, etc. It will depend on whether the agent's skills include steering cars out of skids or balancing plates on poles, and whether the agent's casual causal reach includes having employees publish Kripke's manuscripts … There has to be some part of the agent's action-generating architecture that recognizes the contents as representing potential actions that are within actionable reach, but that doesn't have to rise to the level of a belief or be represented as being within actionable reach. (C&D 2013, 49–51)

---

[10]Note that Lewisians would agree with Cappelen and Lepore that the self is *not* involved in the causal-explanatorily significant rationalizations that these arguments are intended to unearth, because the crucial feature of this account of genuinely *de se* attitudes is that the self is not part of the content – it is brought to bear when the attitude is evaluated at a circumstance, exactly like the world of evaluation is under standard assumptions on contents. In fact, when it comes to the genuinely *de se* attitudes explaining the most primitive basic actions, Perry would agree – cf. e.g. Perry (2002). So I take it that what C&D mean here is that their opponents need to show that representational states give out *at a point where genuinely de se attitudes are still involved*. I find a bit annoying, from the point of view of scholarship, that here and in several other places C&D appear to overlook this crucial aspect of Lewis's (and Perry's) views: 'on our view, the agent doesn't need to be represented in … an adequate rationalization', 37 – but both Perry and Lewis, to make a contrast with whom this remark is intended, would concur; 'what is not trivial is that … we *must* represent ourselves in relation to the objects we engage with', 43 – but this is not what they claim: on the contrary, they would agree that we mustn't. As indicated, I assume that these are just infelicitous expressions, which can be simply ignored by interpreting the claims as suggested.

suppose in order to lift my hand, I must first shift my shoulder blade. It does not seem that I cannot intentionally lift my hand without also having the belief that in order to lift my hand I need to shift my shoulder blade: even intentional actions are ultimately achieved via other actions, ones that are more basic or direct, rather than being the result of some process of thought. (Magidor 2015, 259)

although beliefs certainly play an important role in explaining action, the connection between belief and action must also be mediated by some more hard-wired, physiological constraints. For example, suppose John's arms are paralyzed and Jane's are not. Even if John and Jane share all their beliefs (or if you want, all their centered-beliefs), and Jane's beliefs lead her to lift her arm, those same beliefs will not lead John to lift his arm – simply because he is unable to. Similarly, while Jack's beliefs can lead Jack to lift his legs and flee the scene, Jill's beliefs cannot lead her to lift Jack's legs (at least not in the same manner), simply because she is physiologically unable to do so. (Magidor 2015, 262)

The objection is *prima facie* well taken: we can never be certain that the second condition in any application of MD obtains – the so-called 'multiple differences' problem (Lipton 2004, 128; Scholl 2015, §3). In particular, in cases involving rationalizations it may well be that MD cannot be applied to derive the intended inference because *cetera* are not *paria*, i.e. because some of those non-representational conditions also required for the production of the effect (such as the ones in the examples that Magidor considers) differ. To put it using C&D's *Action Inventory Model* (op. cit., 50) outlined in the quotation above, it might well be that the difference between the messy shopper before and after the epiphany has just to do with the fact that the relevant behaviors were not *actionable* before and have come to be so afterward (i.e. to be in the 'inventory'), while this does not involve any change in the attitudes.

In his highly illuminating piece on the skeptical arguments, Ninan (2016, 106–107) discusses and rejects this objection with considerations that I do not find convincing; this becomes clearer when we assume that the arguments at stake essentially appeal to Mill's MD. In order to appreciate the role of this first main feature of my proposal, it will be helpful to stop for a moment to critically discuss Ninan's otherwise similar one.

Ninan focuses on the fact that, consistent with their devotion to impersonal explanations, C&D typically discuss the explanation of actions with agent-specific contents such as François's action

of ducking under the table. Instead of focusing on the Messy
Shopper case, Ninan considers examples involving two subjects, as
in another example of Perry's in which someone ('DN') is attacked
by a bear and a friend is trying to help. Here, the relevant differ-
5   ences in actions to be explained involve DN's disposition to utter 'I
am attacked by a bear', which the friend lacks, or DN's disposition
to curling up, which again the friend lacks. Ninan grants that, if
the goal is to explain why DN carries out the agent-specific action
type *that DN curls up*, while the friend doesn't, C&D's claim that
10   the explanation is not intentional, but has just to do with the fact
that that action is *actionable* for DN (i.e. in his 'action inventory')
but not for the friend is plausible. But on his view this is only so
because the friend *is not able to perform that action*, in the circum-
stances. What about *agent-neutral* action types, such as the action
15   of curling up, Ninan asks? He argues that we also explain instances
of action types like this, and that C&D need to acknowledge this,
but that their account would be implausible in this case. For, while
the action type *that DN curls up* is intuitively available at the time
to DN but not to the friend, the action *of curling up* is available to
20   both, because both are able to perform it.

      But consider again the Messy Shopper case, or any other
thought-experiment involving just one subject. There the differ-
ence between agent-specific and agent-neutral characterizations
of the action to be explained is irrelevant, because the agent is one
25   and the same; both the action *that the shopper looks into the cart*,
and that *of looking into the cart*, are ones that the shopper is able
to do, both before and after the epiphany, and in that sense they
are both available to him. So the intuitive idea of what 'agents can
do' is not what is behind C&D's notion of 'actionable contents' in an
30   'action inventory', because they are providing the same account
also for these cases: they want to say, I assume, that both those
actions come to be in the inventory of the shopper after the epiph-
any, but were not there before.

      What is truly behind C&D's model is made clear when we bear
35   in mind the 'multiple differences' difficulty in applying Mill's MD:
what they are invoking in their behalf is the plain fact that many
other things, in addition to attitudes, contribute to causing the
actions explained by means of them in rationalizations. As C&D put
it in the first quotation above, 'the representational level gives out

somewhere'; or, as Magidor says, 'the connection between belief and action must also be mediated by some more hard-wired, physiological constraints'. The account of why an action is produced in one case but not the other, in spite of the fact that it could in principle receive a rationalizing explanation ('the subject is capable of doing it') in both, and all relevant attitudes are present in both, might just have to do with some of those other, non-intentional conditions. What C&D suggest about the Messy Shopper case is, as indicated, that both the agent-specific action that the shopper looks into the cart, and the agent-neutral action of looking into the cart, are actionable (i.e. in the inventory) after the epiphany but not before, in spite of the fact that the shopper is 'capable' of doing both of them at the two times. Why? They do not think that they need to answer this; which other non-intentional conditions should be in place for an action to be in the inventory, as C&D say in the quotation provided above, is a messy affair about which nothing systematic or general can be said, at least not at the present state of knowledge. Like Ninan, I think there is something fishy here, but I do not think he put his finger on what it is. I'll come back to what I take it to be after I elaborate on the second crucial feature of my response to their challenge.

(b) In my view, it matters very much that the contrast that the thought-experiments highlight is a *phenomenal* one. C&D's and Magidor's point is in general well taken, and relevant to the argumentative structure of the thought-experiments – in my reconstruction, an appeal to Mill's MD. However, their *prima facie* well-taken alternative suggestion is implausible in these cases, because they are also designed to show that the relevant difference is one of which the subject (and we with her, in our imaginative recreation of the episode) is *consciously aware*. This is what the *epiphany* is, in the Messy Shopper case (and in all others): it is a *conscious realization*. So this is the second feature of my elaboration of the cases: they are not just appeals to Mill's MD, but they are also 'phenomenal contrast' arguments.

A famous example of these arguments is the appeal to the contrast between listening to speech in a given language with and without understanding of what is being said (cf. Kriegel (2015, 30) and Chudnoff (2015, 1–2) for further examples and references). The contrasts are put to different uses, prominent among which

SINQ 1262003
21 November 2016
Initial
CE: XX QA: MS
Coll:XX QC:XX

INQUIRY 15

is that of establishing that there is a *proprietary phenomenology* to cognitive attitudes such as *considering*, *realizing*, or *judging*. This is a controversial issue, regarding which we do not need to take a stand.[11] It only matters for our purposes that *there is* a cognitive phenomenal contrast between the two cases involved in all thought-experiments, whether or not their phenomenal character is proprietarily cognitive.

Putting together the two strands in my reconstruction, this is what I take to be the shape of the arguments that establish the *De Se* effect by invoking the thought-experiments:

(1) There is a causally significant difference between the two cases.

(2) There is no difference in mental states individuated by absolute and shareable contents according to tradition.

(3) The only differential factor in the two cases is the phenomenally accessible presence and absence, respectively, of a *de se* thought.

∴ The differential causal factor is a genuinely *de se* thought.

In order to justify the third premise, it will be useful to adopt a proposal made by Chudnoff (2015, 17–18). He defends a type of phenomenal contrast argument for the irreducibility of cognitive phenomenology that he describes as 'glossed', in that the crucial premise establishing the phenomenal contrast includes a gloss on what the contrast is. It does not merely claim that there is a contrast, but, say, one such that the subject becomes aware of an abstract state of affairs. Our cases of phenomenal contrast could similarly be glossed, in this way: the distinctive thought the shopper has after the epiphany *is about himself*. This gloss links the appeal to a phenomenal contrast with another debate to which such cases are put, that about the relations between phenomenal character

---

[11]For the record, I am convinced by arguments such as the ones given in the works just mentioned, my own epiphany about this came through the combination of a particular example of phenomenal contrast I had found puzzling for a while, and the arguments in Marta Jorba's 2013 dissertation *Cognitive Phenomenology: A Non-Reductive Account*. The example is what we might call the 'Euro effect' – the difference in your impression of the value of an ordinary object such as a meal (in a cheap–expensive scale) measured in a given currency before and after you have become familiar with it, a process people went through when the Euro was introduced. (Which one experiences also when one travels to a country with a different currency, or comes to be familiar with the 'significance' of large amounts in one's own currency.).

and intentionality. In particular, there is a debate concerning the claim that Chudnoff (op. cit., 8) calls *Phenomenal Intentionality*: AQ3 the intentionality of some states is grounded on (or determined by, or supervene on) their phenomenal features.[12] This is a claim I
5 find compelling, but once more we do not need to take a stand on it; we only need to assume something less controversial, namely that some aspects of what one is phenomenally aware of are intentional properties of the states, such as what they are about. So the difference-making thought of which we are phenomenally aware
10 is also first personal in this phenomenal sense, in addition to being so in the (neutral) sense introduced in the previous section and invoked in the argument. This gloss thus elaborates on the specific nature of the contrast that (3) claims to exist, and helps thereby establishing the conclusion by an application of MD.
15   C&D consider a much less developed version of this reconstruction of the arguments (op. cit., 41–42), and complain that a particular case cannot establish a general necessity claim such as IIC2:

> this is at best super-weak inductive support for a strong thesis such as
20 > IIC. Without further argument, this won't justify a necessity claim like
> [IIC2]. *That an ignorance of the kind the agent starts out with in Perry's*
> *examples can block rationalization of agency, doesn't show that it* has
> to. Cases can be used as counterexamples to universal claims, but
> can't be used to (deductively) establish the latter.

25   This overlooks the abductive strength that I am suggesting the experiments have, for they are designed to help us convince ourselves that we could easily recreate them in any particular case amenable to folk psychological explanation.[13] Thus, consider C&D's impersonal explanation above of François's action of ducking. It
30 is obviously not sufficient. It overlooks that in the ordinary cases we are led to imagine in which François carries out the action, he would have (as we are aware of, by putting ourselves in his shoes) the first-personal believe *I am about to be shot*. And it is essential that he does have it, to properly explain the action. Because, once
35 more, we can imaginatively compare two cases in both of which François has the relevant impersonal beliefs and desires, but only

---

[12]Cf. Smithies (2013) for a good presentation of the debate, and Horgan and Graham (2012) for an interesting application.
[13]As indicated above, I agree with Biggs & Wilson, Bonjour (1998) and Peacocke (2004) that reliance on abduction does not deprive the justification for a claim of its *a priori* character.

in one of which he ducks – because, intuitively put, only in one does he believe that he is François, i.e. has relevant attitudes about himself as such.

5   C&D and Magidor might still insist on the 'multiple differences' point: for all we know, it might be that what explains the differential effects in the two cases is some or other independent difference, in our case a non-intentional one, different in different cases. But this is what tobacco companies take refuge in about the experiments pur- portedly showing that nicotine causes cancer, and it is as irrational:
10   this is what, I believe, we find fishy in the skeptics' suggestions. There are different proposals in the literature to elaborate on the nature of this irrationality – i.e. on why MD is a good inferential procedure, when adequately applied. Mill himself (1843, III.VIII.§3) suggested that the clue lies in the intervention involved in the experimental
15   condition. Lipton (2004, 128) doesn't find this sufficient, and insist that we need to appeal to what he takes to be the crucial virtues of the explanations we find best in invoking IBE, what he calls their 'loveliness': in our case, the intentional explanation makes the effect understandable, provides a precise mechanism through which the
20   effect is produced, unifying it with related cases. Scholl (2015, 98), in defense of the virtues of MD independently of IBE, suggests that the clue lies instead on the comparison, in a sufficient number of cases, between the *control* and the *experimental* condition: if there were other relevant causal factors, they would now and then show
25   up in the control situation, producing in those cases the effect. As I said, this is not the place to adjudicate between these proposals. It is enough for my purposes to view the thought experiments as forceful applications of MD, thereby putting the skeptics in com- pany bad enough for their points to loose the strength they might
30   otherwise have thought to have.[14]

---

[14]To be fair, a full defense of the *De Se* effect would need to consider everything relevant to evaluate the explanations. C&D would insist that reliance on objective contents has great merits that we relinquish to our peril in abandoning them. They would be obviously right: we need an alternative account of communication to the simple one that it is just transmis- sion of thoughts (this is what most of the papers in García-Carpintero and Torre (2016) deal with), of disagreement, and more in general of validity and other logical relations. I take the skeptics' views here to be close to Williamson's on epistemicism about vagueness: they prefer to posit mysteries (unknown, perhaps unknowable variegated action explanations; unknowable cut-off points) for the sake of preserving the strengths of the semantic tradition, classical logic included. I feel more committed to what I take to be pieces of evidence like the *De Se* effect than to theoretical assumptions whose standing I find shakier (such as the semantic tradition), but I admit that it is unfair rhetoric on my part to classify the skeptics' views together with straightforwardly irrational attitudes about causal inference.

AQ4

SINQ 1262003
21 November 2016

Initial

CE: XX   QA: MS
Coll:XX   QC:XX

18   M. GARCÍA-CARPINTERO

Skeptics might instead reconsider their acceptance of premise (2). This is what another strand in their considerations suggests, the 'these are just Frege cases' point, cf. C&D (2013), Magidor (2015, 258, 265). I'll go into this a little further in the next section; the basic point I make there is this: given what we have established, they are 'Frege cases' which differ from other such cases in a crucial respect: while the other cases by themselves do not establish the *De Se* effect (i.e. they fail to establish that the relevant Fregean thoughts violate one of the two traditional conditions, absoluteness and shareability), these do establish it. Nothing in what C&D or Magidor say alters this, as far as I can tell.[15]

3. The wider picture: accounts of the *De Se* effect

This concludes my defense of the position that the *De Se* effect is real against the skeptical claims that it is only an enticing illusion. The claim might, however, be bolstered if we have some idea of how the effect comes about and what sort of account might explain it, whether or not we believe the explanation. With this goal in mind, I will now outline my own views.

Of course, there is a familiar model that does purport to explain the *De Se* effect, namely, Lewis (1979) account that abandons the absoluteness of traditional contents by taking them to be properties, or sets of centered worlds. Now, as anticipated, I myself do not think this is the best account; but this is not the problem that drives me here, because choosing among theoretical accounts of the *De Se* effect (as in any other interesting case) will be a holistic

---

[15]Although the main lines of my take on skepticism about the action explanation arguments for genuine *de se* thoughts were articulated in discussions with C&D with the occasion of the course they gave at LOGOS based on the book in June 2014, and some aspects of it featured in the ensuing debates in the LOGOS blog, http://www.theblogos.net/, Ninan (2016) has greatly helped me to shape it. I disagree with some of the concessions he makes to C&D, and with part of his own proposal, as indicated above. (The fact that purported rebuttals of *de se* skeptics disagree so much with each other is further proof that this is skepticism worth taking seriously.) My suggestions are closer to Torre's (ms), who like me considers cases in which the two actions are equally available to the two subjects (or to the subject at different times), and claims that it is differences *in what they are aware of* that are explanatorily relevant. Still, he would need to deal with C&D's (op. cit., 51–52) related point: 'Sometimes the actionable contents are represented as being within actionable reach, the agent believes that they are, but in those cases they are representations of the relationship between the agent and the actionable content. The agent doesn't have to think of him- or herself, the actionable content or the relation between him- or herself and the content in any special way'. A fully convincing reply needs, I think, the appeal to phenomenal contrast in the framework of MD in the way I have suggested.

affair, involving considerations about many (perhaps at first sight unrelated) issues, something in any case beyond what I can sensibly aim to achieve in this piece. What is relevant is that, in spite of the efforts of some Lewisians (Recanati 2007 in particular), I
5   do not think that this view helps with the task just set up for this section – i.e. to help bolster the main claim of this paper.

    Possible worlds propositions need to be considered as semantic values among a plurality of others in semantic and psychosemantic theories, just like extensions, because a compositional account
10   of the contribution of some operators depends on them. Hence, even if we opt for a finer-grained view of contents, we should still countenance related possible worlds propositions. This aside, for a good formal account that is technically convenient for many of the needs we have when devising adequate compositional theories in
15   the semantics of natural language, possible worlds propositions currently offer the best deal. Given the arguments in the previous section, we should thus accept Lewis's suggestions that it is sets of centered worlds and not just worlds that we should thereby take into consideration.[16] But I do not believe that thinking of contents
20   in this way helps us to understand why first-personal contents are not to be adequately modeled with traditional propositions – why they are unable to capture first-personal 'perspectives'.

    Let me then outline the Perrian account I have provided elsewhere (García-Carpintero 2016a), referring the reader there for
25   details, arguments, and further references. The main problem with Perry's own account mentioned above was this. The appeal to states, in contrast to contents, cannot put aside contents altogether, because rationalizing explanation is essentially explanation in terms of contents. For one thing, rationalizing states constitu-
30   tively stand in logical relations: entailment, contradiction, and so on; the simple account of disagreement that Ninan (2016) mentions in favor of the traditional account of propositions is just a particular case of this. As Perry himself and many others have pointed out, contents represent those relations in the way numbers model
35   the relations between quantitative properties that we measure

---

[16]See nonetheless Moss's (2012) proposal on how to stick to traditional possible world propositions, even acknowledging the force of the previous point, which fits nicely with the presuppositional view outlined below. (According to her, traditional possible world propositions are still adequate to model contents, *conditional on what we know with certainty*.)

by means of them. Perry is not insensitive to this fact, because he ultimately characterizes the relevant rationalizing properties of states by ascribing to them contents in addition to the ones he officially counts as such. The problem with his proposal lies in properly understanding the relationship between the official content of a given state, and the unofficial ones also correlated with it that help provide the significant rationalization in first-personal cases. It is at this point that Perry typically resorts to the unhelpful instrumentalist-sounding considerations I questioned above.

My own Perrian proposal has two ingredients that are relevant for present concerns: a quotational acquaintance account of introspection of conscious states' phenomenal features, wedded to a demonstrative account of quotation, and a token-reflexive presuppositional account of indexicals in general and demonstratives in particular. I'll outline them, in reverse order.

(i) *Indexicals, token-reflexivity and presuppositions*. In order to properly understand the interrelations between the two contents, I have resorted to the not just analogous but in my view straightforwardly related case of the relation between the contents of assertions and other speech acts and the contents of the ancillary presuppositions we sometimes express in making them. Although the states of information we end up in by accepting that 'John stole the camera' and 'it was John who stole the camera' are the same, these two sentences pack the information they convey in different ways. The second, cleft sentence *presupposes* that someone stole the camera, but the former, plainer sentence does not. For present purposes, we can think of presuppositions along the well-known lines that Stalnaker (1978) has suggested.[17] Speech acts like assertions take place relative to a *common ground*, a set of already accepted propositions. Linguistic presuppositions are requirements on the common ground, whose satisfaction should be checked at an ideal time just after the utterance is made (because it might well be that it is the utterance itself that generates the common ground information that satisfies them: see the next paragraph), but before acceptance of the resulting assertion

---

[17]My own views, although strongly influenced by him, differ at some points (cf. García-Carpintero 2016b). I take *presupposing* and *referring* to be ancillary speech acts, and I think of such acts as constitutively normative; in particular, I think of presuppositions as constituted by normative requirements that their contents are already common knowledge. I also think that some linguistic presuppositions are semantically triggered.

is assessed. If accepted, it goes to conform the common ground, licensing further presuppositions in the ensuing discourse. The difference between our two sentences lies in the fact that an utterance of the it-cleft sentence will feel inappropriate at *pre-*
supposition evaluation time if it is not common ground by then that someone stole the camera. But the state of information we move to by accepting either will be the same.

Consider an utterance of 'he is hungry'. My view agrees with direct reference theorists such as Kaplan (1989) that the asserted content is a singular proposition, *x is hungry*, for some contextual assignment to *x*. It is expressed, however, in a context in which another singular proposition is presupposed – in this case, one semantically triggered by something akin to a Kaplanian character for 'he' – which, Kaplan's invectives notwithstanding, we should grasp in token-reflexive terms thus: *x is the male picked out by the demonstration associated with he*, where the bold-faced 'he' refers to the relevant token. These semantically triggered presupposi-tions are about linguistic expressions – reflexively about tokens of the very linguistic expressions conveying them.[18] They will typically be supplemented by further pragmatic presuppositions, specify-ing additional features of the demonstrated referent, perceptually accessible or accessible from previous discourse. (These presuppo-sitions illustrate the point made in the previous paragraph about the time at which presuppositions are to be appraised.)

In the proposal I am outlining, a distinction like this between presupposition and other speech acts to which they are ancil-lary obtains in the mental realm too, in the shape of a distinction between occurrent mental states such as judgments and inten-tions, and relevant background beliefs. In what ways are they rel-evant? They would be relevant for the appraisal of the main states vis-à-vis their constitutive norms. In the case of judgments, they would be relevant for the doxastic justification of the beliefs they manifest.[19]

---

[18]They are *de lingua beliefs*, as explained by Fiengo and May (2006, 86), exactly like their *assignments*. Proposals along these lines are quite standard nowadays in the linguistic semantics literature; cf. Maier (2016), where, like here, they are extended to the mental realm.

[19]*Doxastic* justifications are reasons why one has the beliefs one does; *propositional* justi-fication, reasons for which one's beliefs would be justified, if one were to form them in a non-deviant way on their basis.

(ii) Gertler (2012, 99) defines acquaintance views of introspection thus: some introspective knowledge consists in judgments that (1) are directly tied to their truthmakers; (2) depend, for their justification, only on the subject's conscious states at the time of the judgment; and (3) are more strongly justified than any empirical judgments that do not meet conditions (1) and (2). The *directness* in (1) consists in that 'metaphysical reality and the epistemic intersect' in such judgments, a metaphor that she (op. cit., 104–109) elaborates thus. First, the relevant mental features (a phenomenal type, an instance thereof) constitutively (not merely causally) contribute to the content of the judgment: the judgment is about them in virtue of a direct, constitutive relation between the mental reality and the phenomenal judgment. Second, the way things *epistemically* seem to the subject is constituted by the way they *phenomenally* seem to her.

On some more specific versions of this sort of view (cf. Papineau 2007, 120–124), the model is elaborated along quotational lines, assuming a demonstrative view of quotation (García-Carpintero 2004). On the latter views, quotation marks are demonstrative expressions, and the material quoted inside them plays the role of a *demonstratum*; the referent is something saliently related to it. Now, as suggested above, when the messy shopper after the epiphany makes the judgment that he would express by uttering 'I am making a mess', he is consciously aware of his thought being about himself. He is thereby consciously aware of a feature of the representing state he is in: its having a first-personal character. The referent of the first-personal concept in it is fixed by a token-reflexive rule analogous to the linguistic rule for 'I', which we could articulate as follows: *the subject of the thought with this first-personal phenomenal feature*. When the shopper makes the epiphanic judgment that his acceptance of 'I am making a mess' expresses, he is on this view judging a singular content, *x is making a mess*, with him assigned to *x*, and he is presupposing (in the indicated sense, i.e. as a background belief of his, relevant for the epistemic evaluation of the judgment) another singular proposition about him, to the effect that he meets that condition.

AQ5
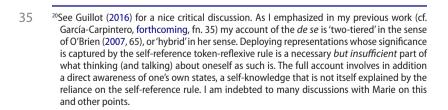
This raises all sorts of issues (say, about its being too intellectualistic an account of the involved states) that I will not address here.[20] I just hope that the outline makes it clear why the view rejects the shareability condition on traditional propositions, constituting a (hopefully mild) version of 'limited accessibility' conceptions of *de se* contents, such as Frege's own. For the presupposed material constitutively includes a feature of the very thought of whose content it is a part. Structurally, the account of Frege cases (the fact that 'Hesperus' and 'Phosphorus', 'Superman' and 'Clark Kent', cannot be respectively substituted *salva veritate* by one another in some sentences) would be exactly parallel, so the skeptics are right that the cases invoked in arguments for the *De Se* effect are 'just Frege cases'. But they are special, in that they motivate abandoning the shareability of contents in a way that, by themselves, other cases don't.

This concludes my defense that the *De Se* effect is genuine, here to stand. A full reply to the skeptics would require going into other relevant issues they tackle, but their discussion of the action explanation cases is in my view the crucial one. I have addressed some of them elsewhere. Thus, I agree with Peacocke (2014, Chapters 5 and 6) that an adequate account of the first person should explain epistemic phenomena such as the Cartesian *cogito* (the particular privilege of judgments such as *I exist*) and *immunity to error through misidentification*. The latter is the apparent *datum* that, on the one hand, when we judge *I once was in Athens* on the basis of a picture of what we take to be ourselves in the Acropolis, our evidence is such that it could still justify the existential 'part' of our judgment, *that someone was in Athens*, even if we were wrong that that was us in Athens; while, on the other, this scission cannot occur when we judge the same on the basis of our first-personal, episodic memories. García-Carpintero (forthcoming) marshals the account just outlined to this task, answering in passing skeptical objections

AQ6

---

[20]See Guillot (2016) for a nice critical discussion. As I emphasized in my previous work (cf. García-Carpintero, forthcoming, fn. 35) my account of the *de se* is 'two-tiered' in the sense of O'Brien (2007, 65), or 'hybrid' in her sense. Deploying representations whose significance is captured by the self-reference token-reflexive rule is a necessary *but insufficient* part of what thinking (and talking) about oneself as such is. The full account involves in addition a direct awareness of one's own states, a self-knowledge that is not itself explained by the reliance on the self-reference rule. I am indebted to many discussions with Marie on this and other points.

SINQ 1262003
21 November 2016
Initial
CE: XX   QA: MS
Coll:XX   QC:XX

24   M. GARCÍA-CARPINTERO

by C&D and Magidor. Similarly, linguists have contended that Castañeda (1968) was right after all that there are in natural languages 'logophors', i.e. expressions used to indicate specifically *de se* ascriptions, although not the ones that Castañeda envisaged; this would be 'big PRO' uses such as the implicit subject of the infinitival clause in ascriptions of attitudes like 'he expects to win the election', in contrast with 'he expects that he (himself) wins the election' (cf. Schlenker 2011). Both C&D and Magidor question this by offering examples in which these constructions are not used in a *de se* way. However, one can reply in defense of the additional indirect support that the existence of logophors would provide for the claim in this paper, as I have done (García-Carpintero 2015, §2), that their examples have purely 'pragmatic' explanations.[21]

A full vindication of the well standing of the *De Se* effect could hence only come from a global account of all the related phenomena that acknowledges it and can be defended as the best overall.[22] I hope I have said enough to somehow reassure those whose confidence in the *De Se* effect was shaken by recent skeptical arguments, in spite of the power that I grant they have.

Acknowledgment

Disclosure statement

---

[21]Thus, Magidor (op. cit., 274) offers an example in which 'Jack wants to be given some water' sounds adequate even if it does not report a *de se* attitude. (Jack is in a hospital bed, one of five marked one to five; unbeknownst to him, he occupies number five. On a screen he sees live footage of the beds, including the numbers; he notices that patient in bed five urgently needs water (without realizing that it is he himself), and shouts 'Water for bed number five, please!'. A nurse then makes the report. However, to the extent that we share the intuition that the report is adequate, we could explain it in that 'want' is not then used to report a propositional attitude, but with the sense of 'being in need of'.

[22]A full account of this kind would need to address the very interesting points in Section 3 of Magidor's paper, where she critically examines the broader 'descriptivist' picture of the attitudes in which Lewis (1979) frames his own account. My own views are very close to Lewis's here; in fact my preference for the alternative Perrian framework to vindicate the *de se* derives precisely from how they both handle the wider picture, so I think we definitely need to confront the points she makes.

SINQ 1262003
21 November 2016
Initial
CE: XX QA: MS
Coll:XX QC:XX

INQUIRY 😊 25

Funding

References

Biggs, Stephen, and Jessica Wilson. Forthcoming. "The Apriority of Abduction." *Philosophical Studies*. AQ8

Boër, S. E., and W. Lycan. 1981. "Who, Me?" *Philosophical Review* 89: 427–66.

Bonjour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge U.P.

Cappelen, Herman, and Josh Dever. 2013. *The Inessential Indexical*. Oxford: OUP.

Castañeda, Hector. 1966. "'He': A Study in the Logic of Self-Consciousness." *Ratio* 8: 130–157.

Castaneda, Hector. 1968. "On the Logic of Attributions of Self-Knowledge to Others." *The Journal of Philosophy* 65: 439–456. AQ9

Chisholm, Roderick M. 1981. *The First Person*. Minneapolis, MN: University of Minnesota Press.

Chudnoff, Elijah. 2015. *Phenomenal Contrast Arguments for Cognitive Phenomenology*. Philosophy and Phenomenological Research: Advance online publication. doi:10.1111/phpr.12177.

Davidson, Donald. 1963. "Action, Reasons and Causes." *Journal of Philosophy* 60, 685–700. Also in his *Essays on Actions and Events*. Oxford: OUP, 1980, 3–19.

Fiengo, R., and R. May. 2006. *De Lingua Belief*. Cambridge, MA: MIT Press.

García-Carpintero, Manuel. 2004. "The *Deferred Ostension* Theory of Quotation." *Noûs* xxxviii: 674–692.

García-Carpintero, Manuel. 2015. *Oxford Handbooks Online*, edited by Sandy Golberg. Accessed November 12, 2015. doi:10.1093/oxfordhb/9780199935314.013.61.

García-Carpintero, Manuel. 2016a. "Token-reflexive Presuppositions and the *De Se*." In *About Oneself*, edited by M. García-Carpintero and S. Torre, 179–199. Oxford: Oxford University Press.

García-Carpintero, Manuel. 2016b. "Accommodating Presuppositions." *Topoi* 35 (1): 37–44. doi:10.1007/s11245-014-9264-5.

García-Carpintero, Manuel. Forthcoming. "*De Se* Thoughts and Immunity to Error through Misidentification." *Synthese*. doi:10.1007/s11229-015-0817-y. AQ10

García-Carpintero, Manuel, and Josep Macià. 2006. "Introduction." In *Two-dimensional Semantics*, edited by M. García-Carpintero and J. Macià, 1–21. Oxford: Oxford University Press.

Gertler, Brie. 2012. "Renewed Acquaintance." In *Introspection and Consciousness*, edited by D. Smithies and D. Stoljar, 93–128. Oxford: Oxford University Press.

Grice, H. P. 1961. "The Causal Theory of Perception." *Aristotelian Society Supplementary Volume* 35: 121–168.

Guillot, Marie. 2016. "Thinking of Oneself as the Thinker: The Concept of Self and the Phenomenology of Intellection." *Philosophical Explorations*.   AQ11

Hill, Christopher. 2009. *Consciousness*. Cambridge: Cambridge University Press.

Horgan, Terence, and Peter Graham. 2012. "Phenomenal Intentionality and Content Determinacy." In *Prospects for Meaning*, edited by R. Schantz, 321–344. Berlin: De Gruyter.

Jackson, Frank. 1986. "What Mary Didn't Know." *Journal of Philosophy* LXXXIII: 291–295.

Kaplan, David. 1989. "Demonstratives." In *Themes from Kaplan*, edited by J. Almog, J. Perry, and H. Wettstein, 481–563. Oxford: Oxford University Press.

Kriegel, Uriah. 2015. *The Varieties of Consciousness*. Oxford: Oxford University Press.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88, 513–543. Also in D. Lewis. *Philosophical Papers*. Vol. 1. Oxford: Oxford University Press.

Lipton, Peter. 2004. *Inference to the Best Explanation*. London: Routledge.

Mackie, John. 1974. *The Cement of the Universe*. Oxford: Clarendon Press.

Magidor, Ofra. 2015. "The Myth of the *De Se*." *Philosophical Perspectives*, Epistemology, 29: 249–283.

Maier, Emar. 2016. "Why My I Is Your You: On the Communication of *De Se* Attitudes." In *About Oneself*, edited by M. García-Carpintero and S. Torre, 220–245. Oxford: OUP.

Mill, J. S. 1843. *A System of Logic*. London: John W. Parker.

Millikan, Ruth. 1990. "The Myth of the Essential Indexical." *Noûs* 24: 723–734.

Moss, Sarah. 2012. "Updating as Communication." *Philosophy and Phenomenological Research* lxxxv: 225–248.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83: 435–450.

Ninan, Dilip. 2016. "What is the Problem of *De Se* Attitudes?" In *About Oneself*, edited by M. García-Carpintero and S. Torre, 86–120. Oxford: Oxford University Press.

O'Brien, Lucy. 2007. *Self-knowing Agents*. Oxford: OUP.

Peacocke, Christopher. 2004. *The Realm of Reason*. Oxford: Oxford University Press.

Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs* 13, 3–21. Also in his *The Problem of the Essential Indexical and other Essays*, 1993, 33–50. Oxford: Oxford U. P. from which I quote.

Perry, John. 2002. "The Self, Self-knowledge and Self-notions." *Identity, Personal Identity, and the Self*, 189–213. Hackett: Indianapolis.

Putnam, Hilary. 1967. "The Nature of Mental States." In *Art, Mind, and Religion*, edited by W. H. Capitan and D. D. Merrill, 37–48. Pittsburgh, PA: Pittsburgh University Press.

Raffman, Diana. 1994. "Vagueness Without Paradox." *The Philosophical Review* 103: 41–74.

Recanati, François. 2007. *Perspectival Thought*. Oxford: Oxford University Press.

Schlenker, Philippe. 2011. "Indexicality and *De Se* Reports." In *Semantics. An International Handbook of Natural Language Meaning*, edited by K. Heusinger, C. Maienborn, and P. Portner, 1561–1604. Berlin: Morton de Gruyter.

Scholl, Raphael. 2015. "Inference to the Best Explanation in the Catch-22: How Much Autonomy for Mill's Method of Difference?" *European Journal for Philosophy of Science* 5: 89–110.

Smithies, Declan. 2013. "The Significance of Cognitive Phenomenology." *Philosophy Compass* 8 (8): 731–743.

Snowdon, Paul. 2010. "On the What-it-is-like-ness of Experience." *The Southern Journal of Philosophy* 48: 8–27.

Sosa, Ernest. 1981. "Propositions and Indexical Attitudes." In *On Believing*, edited by H. Parret, 316–332. Berlin: Walter de Gruyter.

Stalnaker, Robert. 1978. "Assertion." In *Syntax and Semantics*. Vol. 9, edited by P. Cole, 315–332. New York, NY: Academic Press.

Stalnaker, Robert. 1981. "Indexical Belief." *Synthese* 49: 129–151. Also in R. Stalnaker. 1999. *Context and Content*. Oxford: Oxford University Press, 130–149. To which I refer.

Zemach, Eddy M. 1985. "*De Se* and Descartes." *Noûs* 19: 181–204.