

Questions of Reference and the Reflexivity of First-Person Thought

Michele Palmira

Complutense University of Madrid

Forthcoming in *The Journal of Philosophy*

Penultimate version - please refer to the published version which will be available at:

<https://www.pdcnet.org/jphil>

Abstract

Tradition has it that first-person thought is somehow special. It is also commonplace to maintain that the first-person concept obeys a rule of reference to the effect that any token first-person thought is about the thinker of that thought. Following Annalisa Coliva and, more recently, Santiago Echeverri, I take the specialness claim to be the claim that thinking a first-person thought comes with a certain guarantee of its pattern of reference. Echeverri maintains that such a guarantee is explained by a fairly flatfooted interpretation of the thinker-reflexive rule. I argue, however, that the explanatory aspirations of the thinker-reflexive rule are fulfilled only if we accept an epistemically loaded gloss on the notion of a thinker of a thought featuring the rule. That gloss is unpacked in terms of the subject's ability to be acquainted with the phenomenal character of their thoughts.

Several philosophers in the analytic tradition have maintained that the first-person indexical 'I' possesses some special property that is not possessed by any other singular concept.¹ However, the lack of consensus on both what such a property is

¹ This position is traditionally credited to Gertrude Elizabeth Margaret Anscombe, "The First Person," in Samuel D. Guttenplan, ed., *Mind and Language* (Oxford: Clarendon Press, 1975), pp. 45-65; Carol Rovane, "The Epistemology of First-Person Reference," this JOURNAL, LXXXIV, 3(March 1987): 147-67; "Self-Reference: The Radicalization of Locke," this JOURNAL, XC, 2 (February 1993): 73-97;

and why it is that only ‘I’ has it,² together with the uprising of a sceptical movement about the distinctive significance of mental indexicality,³ cast a shadow over the fruitfulness of contemporary debates on the first person. In a recent article,⁴ Santiago Echeverri admirably takes on the twofold task of precisely identifying a candidate special property and explaining why the first-person concept possesses it. Expanding on a proposal originally made by Annalisa Coliva,⁵ Echeverri maintains that what’s distinctive about ‘I’ is that if I raised questions about its reference while thinking a first-person thought I would both be incoherent and use ‘I’ incompetently. Echeverri calls this feature GUARANTEE and proposes to explain it via the idea that the first-

Sydney Shoemaker, “Self-Reference and Self-Awareness,” this JOURNAL, LXV, 19 (October 1968): 555-67; Ludwig Wittgenstein, *The Blue and Brown Books* (New York: Harper and Row 1958). Note that while Anscombe, Shoemaker and Wittgenstein focused on the first-person pronoun in language, I will follow the current literature and focus on the first-person concept.

² See for instance Annalisa Coliva, “The First Person: Error through Misidentification, the Split between Speaker’s and Semantic Reference, and the Real Guarantee,” this JOURNAL, C, 8 (August 2003): 416-31; “Error Through Misidentification: Some Varieties”, this JOURNAL, CVII, 8 (August 2006): 403-25; Aidan McGlynn, “Immunity to Error Through Misidentification and the Epistemology of *De Se* Thought,” in Manuel García-Carpintero and Stephan Torre, eds., *About Oneself* (Oxford: Oxford University Press 2016), pp. 25-55; Simon Prosser and François Recanati, eds., *Immunity to Error through Misidentification: New Essays* (New York: Cambridge University Press, 2012).

³ See for instance Herman Cappelen and Josh Dever, *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person* (New York: Oxford University Press, 2013); Ofra Magidor, “The Myth of the *De Se*,” *Philosophical Perspectives*, XXIX, 1 (December 2015): 249-83; Ruth Garrett Millikan, “The Myth of the Essential Indexical,” *Noûs*, XXIV, 5 (December 1990): 723-34.

⁴ Santiago Echeverri, “Guarantee and Reflexivity,” this JOURNAL, CXVII, 9 (September 2020): 473-500.

⁵ See Coliva “The First Person: Error through Misidentification, the Split between Speaker’s and Semantic Reference, and the Real Guarantee,” *op. cit.*, p. 429.

person concept obeys a reflexive rule of reference to the effect that any token first-person thought is about the thinker of that thought.⁶

In my opinion, Echeverri makes a convincing case for GUARANTEE and succeeds at dispelling a number of worries about the reflexive rule. Despite this, I shall argue Echeverri is mistaken to think that a fairly flatfooted interpretation of the rule explains GUARANTEE. The correct explanation, I shall contend, appeals to a specific gloss on the notion of a thinker of a thought featuring the reflexive rule which ultimately commits us to an epistemically loaded account of how the rule fixes the reference of 'I'.

I Explaining GUARANTEE: Unsuccessful Rule Accounts

In this section I first spell out GUARANTEE and I then argue that two versions of the reflexive rule for 'I' fail to explain it.

I.1. Guarantee and the Bare Rule Account. Echeverri defines GUARANTEE as follows:⁷

A thought, T, has GUARANTEE if and only if:

⁶ The reflexive rule for 'I' is endorsed, amongst others, by John Campbell, "Schizophrenia, the Space of Reasons and Thinking as a Motor Process," *The Monist*, LXXXII, 4 (October 1999): 609-25, p. 623; Manuel García-Carpintero, "De Se Thoughts and Immunity to Error Through Misidentification," *Synthese*, CXCIV, 8 (August 2018): 3311-33, p. 3324; Lucy O'Brien, *Self-Knowing Agents* (Oxford: Oxford University Press, 2007), p. 78; and Christopher Peacocke, *Truly Understood* (New York: Oxford University Press, 2008), p. 81; *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness* (New York: Oxford University Press, 2014), p. 83.

⁷ *Ibid.*, p. 478.

- (1) T has a token of a concept-type, C, that is *de jure* reflexive.
- (2) A thinker of T cannot (coherently) raise questions of reference relative to C while having T.

To unpack clause (1), Echeverri defines *de jure* reflexivity as follows: “A concept C is *de jure* reflexive just in case, if a thinker S produces a token of C, the referent of that token is C”.⁸ To illustrate this definition, let us use Echeverri’s own example and consider the difference between Oedipus’s thoughts ‘The solver of the Sphinx’s riddle killed Laius’ and ‘I killed Laius’.⁹ While both ‘The solver of the Sphinx’s riddle’ and ‘I’ refer to Oedipus, they do so differently: ‘The solver of the Sphinx’s riddle’ refers to Oedipus in virtue of the fact that Oedipus satisfies the description *being the solver of the Sphinx’s riddle*, whereas ‘I’ refers to Oedipus in virtue of the fact that Oedipus is the thinker of that token of ‘I’. To capture this difference, we say that ‘The solver of the Sphinx’s riddle’ is *de facto* reflexive whereas ‘I’ is *de jure* reflexive. Clause (2) of GUARANTEE receives both a normative and a constitutive interpretation:¹⁰ assuming that Oedipus doesn’t know that he is the solver of the Sphinx’s riddle, he can think ‘The solver of the Sphinx’s riddle killed Laius’ while, at the same time, asking *Does ‘The solver of the Sphinx’s riddle’ refer to me?*, without betraying any irrationality or conceptual deficiency. By contrast, if Oedipus thought ‘I killed Laius’ while, at the same time, wondering *Does ‘I’ refer to me?*, we would immediately question his rationality and conceptual competence with the first-person concept.

⁸ *Ibid.*, p. 476.

⁹ *Ibid.*, p. 473.

¹⁰ *Ibid.*, p. 477.

Echeverri's explanation of GUARANTEE is encapsulated in the following biconditional:¹¹

BARE RULE ACCOUNT: A thought, T, has GUARANTEE if and only if T contains a token of a concept-type that is governed by a bare reflexive rule to the effect that a token of I in a thinking stands for the thinker of that thinking.

The right-to-left side of the biconditional is meant to have explanatory priority: so, I cannot coherently and competently ask what 'I' refers to while also thinking a first-person thought *because*, given the reflexive rule, any token of 'I' refers to the thinker of that token of 'I'.

Before going on, let me clarify that the "bare" qualification is my own, and I use it to stress that the reflexive rule proposed by Echeverri doesn't contain a specific gloss on the notion of a *thinker* (or a *subject*) of a thought.¹² So, BARE RULE ACCOUNT purports to explain GUARANTEE without committing us to a specific conception of subjecthood. *Prima facie*, this is a good feature of BARE RULE ACCOUNT: the less an explanation of a phenomenon requires taking a stand on a controversial matter, the better. Relatedly, a second alleged advantage of BARE RULE ACCOUNT, one that is emphasised by Echeverri,¹³ is that it avoids commitment to epistemically loaded principles of reference determination. By so doing, BARE RULE ACCOUNT promises to offer a less controversial explanation of GUARANTEE than some

¹¹ Echeverri, "Guarantee and Reflexivity," *op. cit.*, p. 480.

¹² While I take the notions of "subject" and "thinker" as synonymous here, at some points it will be more natural to use one term rather than the other.

¹³ *Ibid.*, p. 487 fn. 23.

of its competitors, chief amongst them the SELF-ACQUAINTANCE ACCOUNT to the effect that GUARANTEE holds since whenever we think a first-person thought we are acquainted with ourselves.¹⁴ I shall argue, however, that Echeverri is mistaken on both counts: a reflexive rule account of GUARANTEE does require a specific gloss on the notion of a thinker of a thought, a gloss which ultimately leads to an epistemically loaded account of how the rule fixes the reference of the first-person concept.

I.2. Disowned first-person thoughts and the failure of BARE RULE ACCOUNT.

Echeverri correctly observes that a crucial test case for an explanation of GUARANTEE comes from the possibility of disowned first-person thoughts.¹⁵ This possibility is best illustrated via reports of thought insertion made by schizophrenic subjects. Consider the following:

I look out of the window and I think the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his... He treats my mind like a screen and flashes his thoughts on to it like you flash a picture.¹⁶

To make sense of reports like this, several philosophers have adopted John

¹⁴ Echeverri – *Ibid.*, p. 487 fn. 23 – cites, amongst others, Coliva, “The First Person: Error through Misidentification, the Split between Speaker’s and Semantic Reference, and the Real Guarantee,” *op. cit.*, p. 430; Saul A. Kripke, “The First Person,” in *Philosophical Troubles*, vol. 1 (New York: Oxford University Press, 2011), pp. 292–321, p. 301.

¹⁵ Echeverri, “Guarantee and Reflexivity,” *op. cit.*, §V.

¹⁶ This report is contained in C. S. Mellor, “First Rank Symptoms of Schizophrenia,” *The British Journal of Psychiatry*, CXVII, 536 (July 1970): 15-23, here p. 17.

Campbell's distinction between two notions of a thinker of a thought T: ¹⁷ the thinker *qua* "agent" and the thinker *qua* "recipient" of T. In order to count as the agent-thinker of T, the thought "must have been generated by me", ¹⁸ that is, the existence of T is explainable fully by the activity and cognitive powers of my own mind. In order to count as the recipient-thinker of a thought, by contrast, what matters is "the possibility of self-ascription of it by me", ¹⁹ that is, the fact that I can self-ascribe T in a distinctively direct and non-observational way.²⁰ Campbell uses this distinction to maintain that a deluded S genuinely self-ascribes recipient-thinking T while, at the same time, denying agent-thinking it. Hence, the report I quoted above can be spelled out as follows: 'I am recipient-thinking *that the garden looks nice* but Eamonn Andrews is agent-thinking it'. This kind of paraphrase makes thought insertion reports rationally intelligible and non-contradictory, despite being false.

Thought insertion reports seem to suggest the following possibility: S, despite being deluded, competently uses the first-person indexical to coherently question whether it is them who are thinking a first-person thought. If so, S would be using 'I' in accordance with the bare reflexive rule even though the thought would not have

¹⁷ John Campbell, "Schizophrenia, the Space of Reasons and Thinking as a Motor Process," *op. cit.*; "The Ownership of Thoughts," *Philosophy, Psychiatry and Psychology*, IX, 1 (March 2002): 35-39. Other authors who endorse a broadly Campbellian model are Shaun Gallagher, "Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error Through Misidentification," in Dan Zahavi, ed., *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-Experience* (Philadelphia: John Benjamins, 2000), pp. 203-39; G. Lynn Stephens and George Graham, *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts* (Cambridge, MA: MIT Press, 2000); and Peacocke, *Truly Understood*, *op. cit.*, §7.8.

¹⁸ Campbell, "The Ownership of Thoughts," *op. cit.*, p. 36.

¹⁹ *Ibid.*, p. 35.

²⁰ *Ibid.*

GUARANTEE, *contra* what BARE RULE ACCOUNT says.

Echeverri offers two considerations against this type of counterexamples to BARE RULE ACCOUNT.²¹ First, a deluded S cannot really have genuine disowned *first-person* thoughts; secondly, S can have genuine disowned first-person thoughts but those thoughts still possess GUARANTEE. Dialectically, it would be stronger to hold a reference rule that is not hostage to the outcome of the debate about whether thought insertion cases give us a reason to accept that there can be disowned first-person thoughts.²² Thus, in the following, I will propose a reference rule that can accommodate the possibility of disowned first-person thoughts.²³

²¹ Echeverri, “Guarantee and Reflexivity,” *op. cit.*, §5.

²² See, for instance, the exchange between Campbell and Coliva. Campbell, “Schizophrenia, the Space of Reasons and Thinking as a Motor Process,” *op. cit.*; “The Ownership of Thoughts,” *op. cit.*; Annalisa Coliva, “Thought Insertion and Immunity to Error through Misidentification,” *Philosophy, Psychiatry, and Psychology*, IX, 1 (March 2002): 27–34; “On What There Really Is to Our Notion of Ownership of a Thought,” *Philosophy, Psychology and Psychiatry* IX, 1 (March 2002): 41–6.

²³ Campbell, “Schizophrenia, the Space of Reasons and Thinking as a Motor Process,” *op. cit.*, p. 621; Peacocke, *Truly Understood*, *op. cit.*, pp. 89–90, maintain that a deluded counterpart of Oedipus could report the following: “I am aware of the thought ‘I killed Laius’ passing through my mind but it is not me who is thinking it”. This report would then be made rationally intelligible by spelling it out as follows: “I am recipient-thinking ‘I killed Laius’ but it is not me who is agent-thinking it”. As Echeverri notes – “Guarantee and Reflexivity,” *op. cit.*, p. 496 – Peacocke’s discussion of disowned first-person thoughts is in tension with a rule account of GUARANTEE. Campbell clearly took disowned first-person thoughts to show that ‘I’ doesn’t possess any special epistemic property, but Campbell has in mind here immunity to error through misidentification and not GUARANTEE. This is important, since certain first-person thoughts can be vulnerable to error through misidentification and yet exhibit GUARANTEE, something which was first noted in Coliva, “The First Person: Error through Misidentification, the Split between Speaker’s and Semantic Reference, and the Real Guarantee,” *op. cit.*, p. 429. To illustrate. It is widely acknowledged that one’s judgement ‘a is F’ is immune to error

To vindicate the contention that disowned first-person thoughts possess GUARANTEE, Echeverri reasons as follows:²⁴ a deluded S lacks action-awareness of their first-person thought T and, in this sense, they disown being the agent of T.²⁵ However, this does not bring S to question whether it is them who are thinking T. In reporting T the way they do, S does not question that they're bearing a thinking relation to T; they rather question that they are the agent of the token T they are presently thinking.

The foregoing gives us reason to think that disowned first-person thoughts possess GUARANTEE. However, BARE RULE ACCOUNT cannot explain why this is so. To see why, let us begin with the following idea: the reflexive rule of reference for 'I' not only describes its pattern of reference, but it also explains why 'I' refers to what it

through misidentification relative to the concept 'a' only if one's judgement is made on certain grounds. (See for instance Coliva, "Error Through Misidentification: Some Varieties", *op. cit.*, pp. 410-11; Gareth Evans, *The Varieties of Reference* (Oxford: Clarendon Press, 1982), pp. 218-9; James Pryor, "Immunity to Error through Misidentification," *Philosophical Topics*, XXVI, 1/2 (Spring and Fall 1999): 271-304, pp. 274, 282; Shoemaker, "Self-Reference and Self-Awareness," *op. cit.*, p. 557.) As a consequence, immunity to error through misidentification is a *ground-relative* property, whereas GUARANTEE is not. Suppose now, to use one of Coliva's examples (see Annalisa Coliva, "Stopping Points: 'I', Immunity and the Real Guarantee," *Inquiry*, LX, 3 (Spring 2017): 233-52, p. 238), that I judge 'My hair is blowing in the wind' by looking at a glass mirror: it may well be the case that I am misidentifying someone else's hair with mine, but it would not make sense to ask whether it is *me* who is making that very judgement. So, even if Campbell were right about disowned first-person thoughts and immunity to error through misidentification, this would not *ipso facto* show that disowned first-person thoughts do not exhibit GUARANTEE. I am thankful to an anonymous referee for this JOURNAL for their help here.

²⁴ Echeverri, "Guarantee and Reflexivity," *op. cit.*, pp. 497-8.

²⁵ Echeverri draws here on Peacocke, *Truly Understood*, *op. cit.*, p. 276.

refers to. This is tantamount to saying that the reflexive rule plays a type-individuating role. According to a prominent approach championed by Christopher Peacocke,²⁶ this means that the reflexive rule underwrites the canonical patterns of use of the concept – to be specified in terms of suitable primitive introduction and elimination rules – that one must have an implicit grasp of and be disposed to follow in order to count as possessing the concept and be a competent user of it.

Crucially, if the reflexive rule type-individuates the first-person concept, it cannot be bare. The possibility of disowned first-person thoughts ensures that S can coherently assent to ‘I am recipient-thinking T’ while dissenting to ‘I am agent-thinking T’ even though, as a matter of fact, the agent and the recipient of the thought is the same, that is, S. By Frege’s criterion of distinctness of concepts,²⁷ this shows that the concept AGENT-THINKER and the concept RECIPIENT-THINKER are distinct even if, with respect to given a context c, they pick out the same entity in all possible worlds. So, the distinction between AGENT-THINKER and RECIPIENT-THINKER does make a *hyperintensional* difference, namely the kind of difference that generates different answers to the question of why ‘I’ refers to what it refers to. Thus, the

²⁶ See for instance Christopher Peacocke, *A Study of Concepts* (Cambridge, MA: MIT Press, 1992); *Truly Understood*, *op. cit.*; *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*, *op. cit.*

²⁷ I formulate Frege’s criterion as follows: If a rationally intelligible subject S believes x to be F and also believes x not to be G, where F=G, then there are distinct concepts C and C*, such that S believes x to be F under C and disbelieves x to be G under C*. This slightly revises one of the best-known formulations of the criterion due to Stephen Schiffer, “The Basis of Reference,” *Erkenntnis*, XIII (January 1978): 171-206, p. 180. Let me stress that Frege’s criterion is compatible with Echeverri’s favourite account of concepts in terms of mental representations. See for instance François Recanati, *Mental Files* (Oxford: Oxford University Press 2012).

reflexive rule for ‘I’ cannot be type-individuating unless we settle on a given understanding of the notion of a thinker of a thought.

As far as I can tell, this point is implicitly acknowledged in other discussions of the reflexive rule for ‘I’. For instance, Peacocke’s formulation of the rule looks like this.²⁸

AGENT REFLEXIVE RULE: $\forall x \forall \text{event of thinking } \vartheta$: if T involves the use of ‘I’, that use of ‘I’ refers to x just in case x is the agent of ϑ .

AGENT REFLEXIVE RULE is not bare since it does feature a specific gloss on the notion of a thinker of a thought: the thinker is the agent of the thought. A similar sensitivity is reflected in the works of Campbell and Lucy O’Brien.²⁹

I have argued that the bare reflexive rule is unable to type-individuate the first-person concept. As a consequence, BARE RULE ACCOUNT is untenable. In reply, Echeverri may either reject the contention that the reflexive rule must be type-individuating, or try to fall back and use AGENT REFLEXIVE RULE to explain GUARANTEE.³⁰ The former option strikes me as ineffective, as it is reasonable to

²⁸ I slightly revise from Peacocke, *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*, *op. cit.*, p. 83.

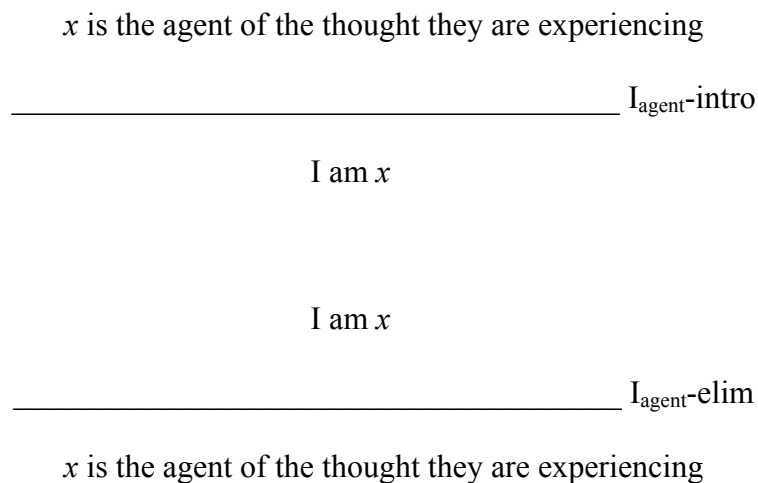
²⁹ Campbell, “Schizophrenia, the Space of Reasons, and Thinking as a Motor Process,” *op. cit.*, p. 623; Lucy O’Brien, *Self-Knowing Agents*, *op. cit.*, p. 78. Both Campbell and O’Brien propose an agent-based gloss on the notion of a thinker of a thought. However, while Campbell accepts that disowned first-person thoughts are genuine first-person thoughts, O’Brien seems to think that they are not – see O’Brien, *Self-Knowing Agents*, *op. cit.*, pp. 70, 85.

³⁰ This fallback position has clear affinity with Echeverri’s final comments on Peacocke’s view – Echeverri, “Guarantee and Reflexivity,” *op. cit.*, pp. 499-500.

expect that a complete reference-fixing story about ‘I’ give an answer to the question of what makes the first-person concept the type of concept it is.³¹ However, closer inspection reveals that the fallback option fails too.

I.3. AGENT REFLEXIVE RULE and GUARANTEE. The canonical patterns of use of ‘I’ underwritten by AGENT REFLEXIVE RULE can be stated like this:

Agent-based canonical patterns of use:



Since we are considering the possibility that AGENT REFLEXIVE RULE could explain GUARANTEE, let us stipulate that S’s disowned first-person T exhibits GUARANTEE. This means that S is a coherent and competent user of ‘I’. However, given that S competently and rationally uses ‘I’ while denying that they are the agent of the thought they are experiencing, they cannot eliminate ‘I’ according to $I_{\text{agent-elim}}$. S is not disposed to introduce ‘I’ via $I_{\text{agent-intro}}$ either, as they refuse to ascribe agent-thinking T to themselves. That is to say, once it is granted that disowned first-person

³¹ See for instance Annalisa Coliva, “Stopping Points: ‘I’, Immunity and the Real Guarantee,” *op. cit.*, p. 244; O’Brien, *Self-Knowing Agents*, *op. cit.*, p. 57; Peacocke, *Truly Understood*, *op. cit.*, p. 78ff.

thoughts possess GUARANTEE, the agent-based patterns of use do not adequately describe the possession conditions of ‘I’. We are thus faced with a dilemma: either AGENT REFLEXIVE RULE type-individuates ‘I’ but GUARANTEE is not a property of all first-person thoughts, or all first-person thoughts have GUARANTEE but they cannot type-individuated by AGENT REFLEXIVE RULE. This dilemma seriously undermines the fallback option I have sketched on behalf of Echeverri.

II Explaining GUARANTEE: The Recipient Rule Account

In this section I propose a way out of the dilemma by articulating a recipient-thinker formulation of the reflexive rule which, I contend, is able to explain GUARANTEE.

II.1. Introducing The Recipient Rule Account. Let us consider the following recipient-thinker formulation of the reflexive rule:³²

RECIPIENT REFLEXIVE RULE: $\forall x \forall \text{event of thinking } \vartheta$: if ϑ involves the use of ‘I’, that use of ‘I’ refers to x just in case x is the recipient of ϑ .

The canonical patterns of use of ‘I’ underwritten by RECIPIENT REFLEXIVE RULE can be stated like this:

Recipient-based canonical patterns of use:

x is the recipient of the thought they are experiencing

_____ I_{recipient-intro}

³² See also Michele Palmira, “Immunity, Thought Insertion, and the First-Person Concept” *Philosophical Studies*, CLXXVII, 5 (December 2021): 3833-60.

I am x

I am x

$I_{\text{recipient-elim}}$

x is the recipient of they are experiencing

On the face of it, a deluded S's uses of 'I' are vindicated by the recipient-based canonical patterns of use of 'I'. S acknowledges that T is passing through their mind and that it is them who are experiencing T. They report this much by using the first-person indexical in a way that can be spelled out schematically as follows: "I am thinking 'I am F'". This is evidence that S is disposed to use 'I' according to $I_{\text{recipient-intro}}$ and $I_{\text{recipient-elim}}$. On these grounds, we can articulate the following explanation of GUARANTEE: S cannot coherently and competently question whether 'I' refers to themselves while also tokening 'I' because, given RECIPIENT REFLEXIVE RULE, any token of 'I' refers to the recipient-thinker of that token of 'I'. More precisely:

RECIPIENT RULE ACCOUNT: A thought, T, has GUARANTEE if and only if T contains a token of a concept-type that is governed by RECIPIENT REFLEXIVE RULE.

To get a grip on RECIPIENT RULE ACCOUNT, we need to offer more details on the notion of a recipient-thinker of a thought. Campbell suggests that if S is the recipient-thinker of T then it is possible for S to self-ascribe T in a direct and non-observational way.³³ I begin to unpack this idea by noting that even when S reports a disowned T, S

³³ Campbell, "The Ownership of Thoughts," *op. cit.*, p. 35.

attends to T and they experience it as something that they are undergoing. So, to experience T is to have information about the phenomenal likeness T has for S. Thus, the possibility of self-ascribing T ensured by being the recipient-thinker of T amounts to the possibility of engaging in a mental activity whereby S attends to T and gains information about what it is like for them to think it. Let us call this attentional mental activity *introspecting* T.³⁴ We thus have:

RECIPIENT-THINKER: x is the recipient-thinker of T if and only if, upon attending to T, x is introspectively aware of the phenomenal character T has for x.

As a consequence, RECIPIENT REFLEXIVE RULE reads as follows:

$\forall x \forall \text{event of thinking } \vartheta$: if ϑ involves the use of 'I', that use of 'I' refers to x just in case x, upon attending to ϑ , is introspectively aware of the phenomenal character ϑ has for x.

On the view on offer, then, the possibility of disowned first-person thoughts does not speak against GUARANTEE,³⁵ nor does it force us to say that such thoughts are first-personal in a derivative sense.³⁶ Disowned first-person thoughts can be legitimately

³⁴ See for instance Charles Siewart, "On the Phenomenology of Introspection," in Declan Smithies and Daniel Stoljar, eds., *Introspection and Consciousness* (Oxford: Oxford University Press, 2012), 129-68; and Anna Giustina, "Introspection Without Judgment," *Erkenntnis*, LXXX, 2 (April 2021): 407-27, for a defence of the idea that introspecting phenomenal properties requires attention.

³⁵ Cf. Campbell, "Schizophrenia, the Space of Reasons and Thinking as a Motor Process," *op. cit.*, p. 621; Peacocke, *Truly Understood*, *op. cit.*, pp. 89-90.

³⁶ Cf. O'Brien, *Self-Knowing Agents*, *op. cit.*, p. 85.

regarded as self-conscious thoughts, in at least the following sense: a subject S of a disowned first-person thought T is aware of what it is like for them to think T as they think it. However, S lacks action-awareness of T, and this is what explains the difference between deluded subjects and normal subjects: while the latter are both action-aware and phenomenally aware of their thoughts, the former are only phenomenally aware of them.

A problem arises though. Suppose that S self-ascribes T by introspecting it, and the phenomenal likeness T has for S is Φ . Suppose further that, upon attending to T and engaging in the introspective activity, S fails to be introspectively aware of Φ . It follows that S is not introspectively aware of what it is like for them to undergo T; hence, S fails to be the recipient of T according to RECIPIENT-THINKER. And yet, it seems clearly possible for S to self-ascribe T on introspective grounds. In this case, then, 'I' would not refer to me in virtue of the fact that I am introspectively aware of the phenomenal character of T, *contra* RECIPIENT REFLEXIVE RULE.

II.2 The epistemic basis of reflexivity. To make RECIPIENT REFLEXIVE RULE – and therefore RECIPIENT RULE ACCOUNT – immune to this type of counterexamples, I submit that we should accept the following thesis: an introspective state targeting the phenomenal character of an occurrent thought is partly constituted by the target phenomenal character. This constitution thesis ensures a certain type of epistemic infallibility to the effect that, whenever S is introspectively aware of the phenomenal likeness T has for S, S is not mistaken about which phenomenal likeness T has for them. If we accept this view of introspection, it follows that the problematic scenario in which S gets the phenomenal character of T wrong but still self-ascribes T on introspective grounds can't simply arise. RECIPIENT REFLEXIVE RULE is saved, so is RECIPIENT RULE ACCOUNT. But, one may reasonably ask, at what price?

The constitution thesis here proposed, together with its epistemic implications, are of course well-known in the self-knowledge literature.³⁷ Let me note, however, that they are often connected to the new acquaintance view of introspection,³⁸ so I will frame my discussion by taking such view as a foil. The discussion will have to be brief, as I cannot offer a full discussion of introspective knowledge by acquaintance in the space of this note. I'll instead try to assuage at least some of the natural worries one might have about harnessing such a view to defend RECIPIENT RULE ACCOUNT.

To begin with, note that the new acquaintance view of introspection does not advocate a naively unrestricted form of infallibility according to which we are infallible about everything we introspect. The infallibility claim takes instead the form of an existential generalisation: there are just *some* phenomenal properties such that we infallibly know them by acquaintance.³⁹ For present purposes, we can restrict the

³⁷ For endorsers, see for instance David Chalmers, "The Content and Epistemology of Phenomenal Belief," in Quentin Smith and Aleksandar Jokic, eds., *Consciousness: New Philosophical Perspectives* (Oxford: Oxford University Press, 2003), 220-72; Manuel García-Carpintero, "De Se Thoughts and Immunity to Error Through Misidentification," *op. cit.*, pp. 3328-9; Brie Gertler "Introspecting Phenomenal States," *Philosophy and Phenomenological Research*, LXIII, 2 (September 2001): 305-28; "Renewed Acquaintance," in Smithies and Stoljar, eds., *Introspection and Consciousness*, *op. cit.*, 93-127; Terence E. Horgan, John L. Tienson and G. Graham "Internal-World Skepticism and Mental Self-Presentation," in Uriah Kriegel and Kenneth Williford, eds., *Self-Representational Approaches to Consciousness* (Cambridge, MA: MIT 2006), pp. 191-207; Terrence E. Horgan and Uriah Kriegel, "Phenomenal Epistemology: What Is Consciousness that We May Know It So Well?," *Philosophical Issues*, XVII, 1 (September 2007): 123-44; Crispin Wright, "Self-Knowledge: The Wittgensteinian Legacy," in Crispin Wright, Barry C. Smith and Cynthia MacDonald, eds., *Knowing Our Own Minds* (Oxford: Oxford University Press 1998), pp. 15-45.

³⁸ See Gertler, "Introspecting Phenomenal States," *op. cit.*, "Renewed Acquaintance," *op. cit.*

³⁹ See Gertler, "Renewed Acquaintance," *op. cit.*

scope of infallibility even further: to redeem RECIPIENT REFLEXIVE RULE it is only required that S have infallible introspective knowledge of the phenomenal texture of their occurrent thoughts only. This makes the infallibility claim I need compatible with the fact that we are *not* infallible in introspecting the complex phenomenal characters of some of our perceptual experiences, as witnessed by the puzzle of the speckled hen. Secondly, to be acquainted with the phenomenal characters of our thoughts in no way commits us to the highly debated thesis that thought comes with a distinctive phenomenal feel that cannot be reduced to sensory phenomenology. All we need for there to be introspective knowledge by acquaintance is the much more humdrum claim that there something it is like for us to have a thought passing through our minds. Finally, let us compare and contrast introspective knowledge by acquaintance of our thoughts' phenomenal characters with self-acquaintance. As has already emerged in §1, Echeverri takes SELF-ACQUAINTANCE ACCOUNT to be more problematic than BARE RULE ACCOUNT. Echeverri offers three considerations in favour of this claim:⁴⁰ First, BARE RULE ACCOUNT, unlike SELF-ACQUAINTANCE ACCOUNT, is neutral about Hume's thesis that the self is not manifested in introspection. Secondly, BARE RULE ACCOUNT, unlike SELF-ACQUAINTANCE ACCOUNT, does not have the implication that when S thinks a first-person thought S has an experience that reveals the essence of things. Thirdly, being the thinker of a thought is an episodic relation we bear to our thoughts as we are thinking them, but self-acquaintance is a relation we always bear to ourselves.

Fortunately, however, RECIPIENT RULE ACCOUNT retains these advantages over SELF-ACQUAINTANCE ACCOUNT despite being committed to an acquaintance view of introspection. The first point to note is that the kind of introspective knowledge by

⁴⁰ Echeverri, "Guarantee and Reflexivity", *op. cit.*, p. 487 fn. 23.

acquaintance I appeal to is compatible with Hume's elusiveness of the self thesis, for the idea is that I'm only acquainted with the phenomenal texture my thoughts have for me and not with myself. Secondly, being acquainted with the phenomenal character of my thoughts does not imply that I'm acquainted with the essence or the nature of those thoughts, unless one also assumes the thesis that a thought's phenomenal character somehow constitutes its essence or nature. This additional assumption, however, is not part of RECIPIENT RULE ACCOUNT. Thirdly, I take introspective knowledge by acquaintance to require attending to the phenomenal character of T. For this reason, we must keep it distinct from the sort of peripheral awareness of our conscious states we (allegedly) have whenever we are conscious.⁴¹ Thus, being acquainted with the phenomenal character T has for me, just like thinking T, is an episodic relation I bear to T.

This shows that RECIPIENT RULE ACCOUNT is not saddled with the problems that affect SELF-ACQUAINTANCE ACCOUNT. I thus conclude that RECIPIENT RULE ACCOUNT is superior to BARE RULE ACCOUNT, AGENTIVE RULE ACCOUNT and SELF-ACQUAINTANCE ACCOUNT.

III Conclusions

Echeverri fails to accomplish his primary aim, namely to explain GUARANTEE via BARE RULE ACCOUNT. GUARANTEE can't be explained by AGENT REFLEXIVE RULE either. If we aim to explain GUARANTEE by appealing to the reflexive rule for 'I', we must understand the notion of a thinker of a thought that features the rule in the

⁴¹ For an articulation and defence of the notion of peripheral awareness, see Uriah Kriegel, *Subjective Consciousness: A Self-Representational Theory* (Oxford: Oxford University Press 2009): chapter v.

recipency sense, where S is the recipient of a thought T just in case, when S attends to T, S obtains infallible introspective knowledge by acquaintance of the phenomenal character T has for them. This makes the rule account of GUARANTEE I end up proposing, RECIPIENT RULE ACCOUNT, epistemically loaded and I think ultimately tenable albeit controversial.

Be that as it may, I do suspect that if the first-person concept really is special it will not be easy to come up with lightweight philosophical explanations of why this is so.⁴²

⁴² **Acknowledgements**

I am grateful to Annalisa Coliva, Manuel García-Carpintero, Anna Giustina, Marta Jorba, Daniel Morgan, Sven Rosenkranz, Léa Salje, Carlota Serrahima, Victor Verdejo and an anonymous referee for this journal, for their helpful comments on earlier drafts. Work on this article has received funding from the Spanish Government's Ministerio de Ciencia, Innovación y Universidades under grant agreement RYC2018-024624-I.