THE JOURNAL OF PHILOSOPHY
VOLUME CXVIII, NO. 7, JULY 2021

PUTTING *I*-THOUGHTS TO WORK*

An angry teacher once screamed, "Please leave the room!" I looked around. There was Laura, who always had good grades. There was Paul, the teacher's nephew. And there was *me*, who had been struggling to follow the history class. I thought *He is talking to me* and decided to leave the room. Thoughts of this form—*I*-thoughts—feature tokens of the self-concept, a concept we typically express with different forms of the first-person pronoun ("I," "me," "mine," and "myself") and the inflexion of the verb, as in Descartes's *cogito*.[1]

A traditional view analyzes the self-concept as an indexical concept.[2] Unfortunately, this traditional view has been attacked from various

*This paper has been long in the writing. I am grateful to Alfonso Anaya, Brendan Balcerak Jackson, Magdalena Balcerak Jackson, Axel Barceló, Eric Bayruns García, Reinaldo Bernal, Adam Bradley, Christopher Brown, Simon Brown, Roberto Casati, David Chalmers, Elijah Chudnoff, Adrian Cussins, Jérôme Dokic, Juan Camilo Espejo, Simon Evnine, Chad Kidd, Uriah Kriegel, Béatrice Longuenesse, Juliana Lima, Ruth Millikan, Hedda Hassel Mørch, Bence Nanay, Lucia Oliveri, David Papineau, Claudia Passos, Nicolas Porot, Jesse Prinz, François Recanati, Katherine Ritchie, David Rosenthal, Ezra Rubenstein, Bradford Saad, Susanna Schellenberg, Miguel Ángel Sebastián, Laura Silva, Pedro Stepanenko, Thomas Teufel, Gerardo Viera, and two anonymous referees from another journal for their insightful comments, criticisms, and questions on previous drafts of this paper. Work on this project was funded by grants from the Swiss National Science Foundation (FNS P300P1_161061/1) and DGAPA (PAPIIT IG400219).

[1] See René Descartes, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, vol. 2, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge, UK: Cambridge University Press, 1984), pp. 1–62.

[2] See John Campbell, *Past, Space, and Self* (Cambridge, MA: MIT Press, 1994); John Campbell, "What Is It to Know What 'I' Refers To?," *The Monist*, LXXXVII, 2 (April 2004): 206–18; Héctor-Neri Castañeda, "'He': A Study in the Logic of Self-Consciousness," *Ratio*, VIII (December 1966): 130–57; Héctor-Neri Castañeda, "Indicators and Quasi-indicators," *American Philosophical Quarterly*, IV, 2 (April 1967): 85–100; Andy Hamilton, *The Self in Question: Memory, the Body and Self-Consciousness* (Basingstoke: Palgrave Macmillan, 2013), p. 22; David Kaplan, "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in

0022-362X/21/1807/345–372

fronts. In a highly influential article, Ruth Millikan famously held that there are no mental indexicals. Focusing on the first person, she suggested that the self-concept is a Millian name. On her view, what distinguishes the self-name from ordinary names has nothing to do with indexicality; what makes the self-name special is its functional role in our mental life.[3]

To work as an indexical, the self-concept should have a semantics sufficiently similar to the semantics of linguistic indexicals like "I," "this," "here," and "now." However, a number of reasons seem to speak against this approach. To illustrate, linguistic indexicals shift their reference from context to context. By contrast, the self-concept does not seem to shift its reference from context to context. Plausibly enough, all my tokens of the self-concept refer to me and nobody else.

I shall come back to the no-shift objection in section VI. Furthermore, I shall concede that we might need to posit a self-name to do some explanatory work. I shall argue, however, that we have good reasons to construe the self-concept as an indexical concept. On the view I advocate, we have a self-concept whose reference is fixed by a reflexive rule (*RR*): A token of *I* in a thinking stands for the subject of that thinking. The conceptual tools I will deploy will not *demonstrate* that it is *RR*—and not another indexical rule—that provides the right semantics for the self-concept. Still, these conceptual tools will put pressure on the Millian view and lend substantial support to the indexical view.

The article proceeds as follows. In sections I–II, I offer a reconstruction of the contrast between the Millian view and the indexical view. Section III provides a partial characterization of the functional role of the self-concept. The key idea is that the self-concept functions as a device of information integration. Sections IV–V present

---

Joseph Almog, John Perry, and Howard Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), pp. 481–563, at pp. 533–34; Lucy O'Brien, *Self-Knowing Agents* (Oxford: Oxford University Press, 2007); Christopher Peacocke, *Truly Understood* (New York: Oxford University Press, 2008); Christopher Peacocke, *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness* (New York: Oxford University Press, 2014); Christopher Peacocke, *The Primacy of Metaphysics* (New York: Oxford University Press, 2019); John Perry, "Frege on Demonstratives," in *The Problem of the Essential Indexical and Other Essays* (Stanford, CA: CSLI, 2000), pp. 1–26; John Perry, "The Problem of the Essential Indexical," in *op. cit.*, pp. 27–44; and R. M. Sainsbury, "English Speakers Should Use 'I' to Refer to Themselves," in Anthony Hatzimoysis, ed., *Self-Knowledge* (New York: Oxford University Press, 2011), pp. 246–60.

[3] See Ruth Garrett Millikan, "The Myth of the Essential Indexical," in *White Queen Psychology and Other Essays for Alice* (Cambridge, MA: MIT Press, 1993), pp. 265–77. Millikan's article has been widely cited in the recent literature on *de se* skepticism. Nevertheless, her view has received little attention. In section VII, I draw some general morals from our discussion for the recent debate on the essential indexical.

an explanatory argument against the Millian view and in favor of the indexical view. This explanatory argument differs from previous arguments in a key respect. Instead of focusing on intuitions about the incompleteness of some non-indexical explanations, it applies some explanatory tools from the literature on mental representation to the self-concept. These tools spell out the idea that the semantic rules of a representation-type are the rules that best explain why that representation-type successfully fulfills its (broad) functional role. I will suggest that an indexical rule like *RR* yields a better explanation of some forms of information integration than the Millian rules that determine the reference of Millian names. Section VI responds to some objections, and section VII draws some morals for the debate on the essential indexical.

Before diving in, let me flag some important assumptions and terminological decisions. I shall use "concept" to denote a type of mental representation that can be shared by different subjects and can participate in inference. One could construe thoughts either as representations in a Fodorian language of thought[4] or as internalized linguistic utterances.[5] I will remain neutral on these two options. I shall assume that thoughts are complexes of concepts and that the contents of thoughts are Russellian propositions, that is, structured entities constituted by individuals and properties.[6] Some philosophers characterize concepts as Fregean senses and thoughts as complexes of Fregean senses.[7] Despite this crucial difference, my arguments are consistent with a broadly Fregean approach. Other philosophers prefer to speak of mental files.[8] What I shall say here can be easily translated into the mental-file framework too. I shall use quotation marks to mention linguistic expressions ("I") and italics to mention mental representations (*I*). Whenever I speak of the self-concept, it can be understood

[4] Jerry Alan Fodor, *The Language of Thought* (Cambridge, MA: Harvard University Press, 1975).

[5] Daniel Clement Dennett, *Consciousness Explained* (London: Penguin Books, 1991).

[6] Kaplan, "Demonstratives," *op. cit.*; Millikan, *Language, Thought and Other Biological Categories* (Cambridge, MA: MIT Press, 1984); Millikan, *White Queen Psychology*, *op. cit.*; Perry, "Frege on Demonstratives," *op. cit.*; Perry, "The Problem of the Essential Indexical," *op. cit.*; and Sainsbury, "English Speakers Should Use 'I' to Refer to Themselves," *op. cit.*

[7] José Luis Bermúdez, *The Paradox of Self-Consciousness* (Cambridge, MA: MIT Press, 1998); José Luis Bermúdez, *Understanding "I": Language and Thought* (New York: Oxford University Press, 2016); Gareth Evans, *The Varieties of Reference* (Oxford: Clarendon Press, 1982); Christopher Peacocke, *A Study of Concepts* (Cambridge, MA: MIT Press, 1992); and Peacocke, *Truly Understood*, *op. cit.*

[8] John Perry, *Reference and Reflexivity* (Stanford, CA: CSLI, 2001), §7.3; John Perry, *Identity, Personal Identity, and the Self* (Indianapolis, IN: Hackett, 2002), pp. 193–97; and François Recanati, *Mental Files* (New York: Oxford University Press, 2012).

either as a name or an indexical. I shall use "self-thoughts" to denote thoughts that feature tokens of the self-concept understood in either of these two ways.

I. THE MILLIAN VIEW

Philosophers have often suggested that a mental name could play the functional role of the self-concept. To illustrate, Evans denies that "self-conscious thought would be beyond the reach of those who. . .had to refer to themselves with their own names."[9] It is unclear, however, how Evans would have developed the name view. By contrast, Ruth Millikan is explicit on this count. In a highly influential article, she writes, "So-called 'essential indexicals' in thought are indeed essential, but they are not indexical. It is not their semantics that distinguishes them but their function, their psychological role."[10] Focusing on the self-concept, Millikan suggests that it works like a name:

> Now it is trivial that if I am to react in a special and different way to the knowledge that I, RM, am positioned *so* in the world, a way quite unlike how I would react knowing anyone else was positioned so in the world, then my inner term for RM must bear a very special and unique relation to my dispositions to act. *But what does that have to do with indexicality?* My inner name "RM" obviously is not like other names in my mental vocabulary. It is a name that hooks up with my knowhows, with my abilities and dispositions to act, in a rather special way. . . .My inner "RM" is indeed special. Let us call it @"RM", or RM's "active self name". It names the person whom I know, under that name, *how* to manipulate directly; I *know how* to effect her behavior.[11]

A few pages later, Millikan suggests that the self-concept is a Millian name whose content is exhausted by its referent.[12] We can elicit a "Millian view" from the previous remarks:

> THE MILLIAN VIEW
>
> MILLIAN CLAIM. The self-concept is a mental representation that lacks an indexical semantics. Instead, the self-concept should be construed as a Millian name.
>
> FUNCTIONAL ROLE CLAIM. What distinguishes the self-concept from other names is its functional role.

[9] Gareth Evans, "Understanding Demonstratives," in *Collected Papers* (Oxford: Clarendon Press, 1985), pp. 291–321, at p. 321.
[10] Millikan, "The Myth of the Essential Indexical," p. 265.
[11] *Ibid.*, pp. 273–74.
[12] *Ibid.*, p. 276.

I shall criticize the Millian view. I do think, however, that my arguments could be generalized to other semantic accounts of names. Nevertheless, I leave for another occasion the discussion of these views.[13]

II. THE INDEXICAL VIEW

The most natural way of opposing the Millian view is to reject the Millian claim and argue that the self-concept has an indexical semantics.[14] So, one might argue that the self-concept has a *de se* content.[15] This strategy is so intuitive that recent *de se* skepticism has focused on *de se* contents.[16] Nevertheless, I have suggested that the contents of thoughts are Russellian propositions. Therefore, I must draw the distinction between the *I*-concept and the self-name in a different way.

There are many asymmetries between linguistic and mental representations. These asymmetries will become important when we discuss some objections to the indexical view (section VI). Despite these asymmetries, however, we can try to identify a constitutive mark of *any* indexical representation. To this end, it will be useful to consider an indexical rule for the *I*-concept:

RR REFLEXIVE RULE. A token of *I* in a thinking stands for the subject of that thinking.[17]

[13] Other authors have suggested that the self-concept could be construed as a mental name. They include Dennett, *Consciousness Explained*, *op. cit.*, p. 429; Perry, "On Knowing Your Self," in Shaun Gallagher, ed., *The Oxford Handbook of the Self* (Oxford: Oxford University Press, 2011), pp. 372–93, at p. 385; and Herman Cappelen and Josh Dever, *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person* (New York: Oxford University Press, 2013), p. 44. Recanati (*Mental Files*, *op. cit.*, p. 60) thinks that the self-concept works like an indexical because we can draw the type/token distinction for the self-concept. Nevertheless, he does not provide an indexical semantics for the self-concept. Therefore, his account bears similarities to accounts that construe the self-concept as a mental name.

[14] Cappelen and Dever, *The Inessential Indexical*, *op. cit.*, p. 148.

[15] Bermúdez, *The Paradox of Self-Consciousness*, *op. cit.*; Bermúdez, *Understanding "I"*, *op. cit.*; James Higginbotham, "Remembering, Imagining, and the First Person," in Alex Barber, ed., *Epistemology of Language* (Oxford: Oxford University Press, 2003), pp. 496–533; Peacocke, *Truly Understood*, *op. cit.*; Peacocke, *The Mirror of the World*, *op. cit.*; and Stephan Torre, "In Defense of *De Se* Content," *Philosophy and Phenomenological Research*, XCVII, 1 (July 2018): 172–89.

[16] Cappelen and Dever, *The Inessential Indexical*, *op. cit.*; and Ofra Magidor, "The Myth of the *De Se*," *Philosophical Perspectives*, XXIX, 1 (December 2015): 249–83.

[17] For similar rules, see Bermúdez, *The Paradox of Self-Consciousness*, *op. cit.*; Bermúdez, *Understanding "I"*, *op. cit.*; Campbell, *Past, Space, and Self*, *op. cit.*; Campbell, "What Is It to Know What 'I' Refers To?," *op. cit.*; Manuel García-Carpintero, "*De Se* Thoughts and Immunity to Error through Misidentification," *Synthese*, CXCV, 8 (August 2018): 3311–33; Kaplan, "Demonstratives," *op. cit.*, pp. 533–34; O'Brien, *Self-Knowing Agents*, *op. cit.*; Michele Palmira, "Immunity, Thought Insertion, and the First-Person Concept," *Philosophical Studies*, CLXXVII (December 2020): 3833–60; Perry, "Frege on Demonstratives,"

We want *RR* to be consistent with different metaphysical accounts of subjects. So, we will only assume that a subject is a possessor of mental states and events.[18] We can think of *RR* as an informal statement of the relation that a subject must bear to a token of *I* in order to be the referent of that token.[19] Consider my thought, *He is talking to me*. Why did my token of *me* refer to me and not to the angry teacher, Laura, or Paul? *RR* offers a straightforward answer: I was the only subject who was in a thinking relation to my token of *me*.

What makes *RR* a paradigmatic example of an indexical rule? My suggestion is that *RR* spells out a relation between an entity that instantiates a contextual property and a representation-token. Call this an "indexical relation." The first relatum is a thinking subject. Presumably, subjects engage in thinking in *some* but *not all* contexts. The second relatum is a representation-token.[20]

Consider now a Millian rule for the mental name *Santiago*:

MR MILLIAN RULE. *Santiago* stands for Santiago.

MR codifies a relation that links a referent and a mental name. The first relatum is Santiago, independently of whether Santiago instantiates a contextual property. The second relatum is a concept-type. Therefore, *MR* does not codify an indexical relation. For ease of exposition, we can say that *MR* codifies a "Millian relation."[21]

---

op. cit.; Perry, "The Problem of the Essential Indexical," *op. cit.*; Peacocke, *Truly Understood*, *op. cit.*; Peacocke, *The Mirror of the World*, *op. cit.*; Peacocke, *The Primacy of Metaphysics*, *op. cit.*; and Sainsbury, "English Speakers Should Use 'I' to Refer to Themselves," *op. cit.* There is some controversy as to what the best formulation of the indexical rule for the *I*-concept is. Moreover, some authors supplement the indexical rule with various forms of non-conceptual experience. My arguments are consistent with other formulations of the indexical rule for the *I*-concept, and I will not examine the role of non-conceptual experience in self-reference.

[18] Peacocke, *The Mirror of the World*, *op. cit.*, p. 38.

[19] Peacocke, *Truly Understood*, *op. cit.*, p. 57.

[20] Philosophers of logic often formulate indexical rules in relation to abstract *occurrences* rather than *tokens* (Kaplan, "Demonstratives," *op. cit.*; and Stefano Predelli, *Contexts: Meaning, Truth, and the Use of Language* (Oxford: Clarendon Press, 2005). If we are dealing with thoughts, however, it seems necessary to formulate indexical rules in relation to representation-tokens. There are different ways of individuating representation-tokens. For the purposes of this article, I do not need to take a stance on this issue. I will work with an intuitive understanding of a representation-token as a particular instance of a representation-type.

[21] Millian accounts of names typically come with causal-historical explanations of reference. These causal-historical explanations are consistent with *MR*. It suffices to think of Millian rules as codifying the semantics of Millian names and of causal-historical explanations as providing a meta-semantics for Millian rules. See Kaplan, "Afterthoughts," in Almog, Perry, and Wettstein, eds., *Themes from Kaplan*, *op. cit.*, pp. 565–614, at pp. 573–74.

We can now see what a successful argument against the Millian view would look like: it should show that the reference of the self-concept is determined by a rule that codifies an indexical relation. In the case of *RR*, that indexical relation connects a thinking subject with a token of the self-concept. We have seen, however, that the Millian view distinguishes the self-name from other names by its functional role (section I). So, we should offer a partial characterization of the functional role of the self-concept.[22]

III. THE FUNCTIONAL ROLE OF THE SELF-CONCEPT

Many philosophers think that we can integrate various pieces of self-concerning information without deploying a self-concept. Here is John Perry:

> For many purposes we don't need notions of ourselves at all. Consider the simple act of seeing a glass of water in front of one and drinking from it. The perceptual state corresponds to a relation between an agent and a glass of water. It is the state an agent is typically in when there is a glass of water in front of that agent. . . .The identity between the perceiver and the agent is (normally) guaranteed outside of thought, by the "architectural" relations between the eyes and arms. One need not keep track of it in thought.[23]

Perry's example is not uncontroversial. Nevertheless, it illustrates a common worry. It might turn out that some forms of perception-action coordination involve *no representations* at all (think of simple forms of conditioned behavior). Others might involve representations but *no self-representation*. And still others might involve only *non-conceptual self-representations*.[24]

A central philosophical task is to spell out the conditions under which it becomes necessary for a subject to deploy non-conceptual

[22] Given the current analysis, a successful argument against the Millian view does not need to reject the traditional picture of propositions as having absolute truth values. After all, indexical rules can connect representation-tokens with constituents of Russellian propositions. I leave open the question of whether a *full* development of the indexical view would lead us to reject the traditional picture of propositions. For further discussion of this topic, see Manuel García-Carpintero, "The Philosophical Significance of the *De Se*," *Inquiry*, LX, 3 (Spring 2017): 253–76; Manuel García-Carpintero and Stephan Torre, eds., *About Oneself* (Oxford: Oxford University Press, 2016); David Lewis, "Attitudes *De Dicto* and *De Se*," *The Philosophical Review*, LXXXVIII, 4 (October 1979): 513–43; Perry, "The Problem of the Essential Indexical," *op. cit.*; and Torre, "In Defense of *De Se* Content," *op. cit.*

[23] Perry, *Identity, Personal Identity, and the Self*, *op. cit.*, p. 208.

[24] For arguments similar to Perry's, see Bermúdez, *The Paradox of Self-Consciousness*, *op. cit.*, pp. 35–39; Campbell, *Past, Space, and Self*, *op. cit.*, pp. 119–20; Cappelen and Dever, *The Inessential Indexical*, *op. cit.*, p. 43; Jenann Ismael, *The Situated Self* (New York: Oxford University Press, 2007), p. 49; Magidor, "The Myth of the *De Se*," *op. cit.*,

and conceptual self-representations. It would take us too far afield to address this issue here.[25] Instead of tackling this question, I will highlight some aspects of the functional role of the self-concept that are not prey to Perry's line of thought. On the view I shall recommend, the self-concept makes self-concerning information from various channels available for use in inference. Given that concepts are mental representations that can participate in inference, it is not unreasonable to hold that this device—the self-concept—works as a concept.[26]

I will use the first-person pronoun to describe the functional role of the self-concept. However, my use of the first person should be understood as an expository convenience; it will leave open the possibility that the self-concept is a Millian name. I shall proceed in two steps. First, I will introduce a rough distinction between two types of information channels. Second, I will use that distinction to offer a partial characterization of the functional role of the self-concept.

III.1. Reflexive and Non-Reflexive Channels. Self-thoughts are typically responses to input conditions. In the angry teacher scenario, my thought *He is talking to me* was a response to the teacher's utterance, "Please leave the room!" We will need a rough characterization of these input conditions. To this end, I will introduce a distinction between reflexive and non-reflexive channels:

> REFLEXIVE CHANNELS. If an information channel, C_r, *always* delivers information of a subject, *S*, then C_r is a *reflexive channel* relative to *S*.

---

p. 262; Millikan, "The Myth of Mental Indexicals," in Andrew Brook and Richard DeVidi, eds., *Self-Reference and Self-Awareness* (Amsterdam: John Benjamins, 2001), pp. 163–77; Kristina Musholt, *Thinking about Oneself: From Nonconceptual Content to the Concept of a Self* (Cambridge, MA: MIT Press, 2015), p. 56ff.; O'Brien, *Self-Knowing Agents*, *op. cit.*, p. 60; Peacocke, *The Mirror of the World*, *op. cit.*, chapter 2; Perry, "Thought without Representation," *Proceedings of the Aristotelian Society, Supplementary Volumes*, LX, 1 (July 1986): 137–52; Perry, "On Knowing Your Self," section 5; Recanati, *Perspectival Thought* (New York: Oxford University Press, 2007), pp. 256–57; and Susanna Schellenberg, "*De Se* Content and *De Hinc* Content," *Analysis*, LXXVI, 3 (July 2016): 334–45.

[25] But see Peacocke, *The Primacy of Metaphysics*, *op. cit.*, chapter 4; and Santiago Echeverri, "The Generality of First-Person Thought" (manuscript).

[26] This characterization of the functional role of the self-concept does not entail that the self-concept is the *only* concept that can fulfill that functional role. My aim is purely descriptive. I want to identify some ways in which the self-concept functions as a concept. One can say that hearts function as blood-pumping devices without thereby claiming that hearts are the only entities capable of pumping blood.

NON-REFLEXIVE CHANNELS. If an information channel, C_n, *does not always* deliver information of a subject, *S*, then C_n is a *non-reflexive channel* relative to *S*.[27]

Plausible examples of reflexive channels are proprioception and introspection. In normal conditions, my proprioceptive system always delivers information of the position and movement of my body. Therefore, my proprioceptive system is a reflexive channel relative to me. In normal conditions, introspection always delivers information of my own mental states and events. Therefore, introspection is a reflexive channel relative to me.[28]

Examples of non-reflexive channels include mirrors and written and spoken utterances. In normal conditions, mirrors deliver information of my appearance but also of the appearance of other things. That is why I can fail to recognize my own image in a mirror. Therefore, mirrors are non-reflexive channels relative to me. In normal conditions, written and spoken utterances do not always deliver information of me. Indeed, they often deliver information of other people, the weather, or the political situation. Thus, written and spoken utterances are non-reflexive channels relative to me.

There are mixed cases as well. My visual experiences often deliver information of things other than me, such as a cup on the table or Peter approaching. If we focus on those contents, vision counts as a non-reflexive channel. Nevertheless, all visual experiences are also centered on the perceiver.[29] Hence, although visual experiences do deliver information of entities other than the perceiver, there is a sense in which visual experiences always carry information of the perceiver. When I see a cup on the table, the location of the cup is fixed

[27] Our concept of reflexive channels largely overlaps with Perry's theory of normally self-informative ways of gaining information. See Perry, *Identity, Personal Identity, and the Self*, *op. cit.*; and Perry, "On Knowing Your Self," *op. cit.* A full-fledged theory of information channels should spell out their individuation conditions. I shall address this issue in future work. The argument I present below is consistent with different accounts of information channels.

[28] Some philosophers have imagined possible worlds in which a subject's alleged reflexive channel is wired to another subject (David Malet Armstrong, "Consciousness and Causality," in David Malet Armstrong and Norman Malcolm, eds., *Consciousness and Causality: A Debate on the Nature of Mind* (Oxford: Blackwell, 1984), p. 113; Quassim Cassam, *Self and World* (Oxford: Oxford University Press, 1997), p. 62; and Evans, *The Varieties of Reference*, *op. cit.*, p. 235ff.). One might think that those scenarios cast doubt on the concept of a reflexive channel. I have tried to avoid this objection by relativizing reflexive channels to normal conditions. Of course, a theory of reflexive channels should offer an account of normal conditions. We can leave open the question of whether some channels are reflexive in *all* metaphysically possible worlds.

[29] Evans, *The Varieties of Reference*, *op. cit.*, p. 222; and Peacocke, *A Study of Concepts*, *op. cit.*, chapter 3.

in relation to me. When I see Peter approaching, the direction of Peter's movement is also fixed in relation to me. Given its egocentric character, vision counts as a reflexive channel.[30]

To be sure, a lot more could be said on reflexive and non-reflexive channels. However, this rough characterization is all we need to offer a partial description of the functional role of the self-concept.

III.2. Introduction and Consumption Rules. We can use our two types of information channels to describe two "introduction rules" for the self-concept.

ARTICULATION. Self-thoughts can be used to *make explicit* the subject in response to self-concerning information originating from a reflexive channel. If I see a ball flying toward me, I can think *That ball is approaching me*. In this case, the self-concept articulates the subject of my visual experience.

ANCHORING. When the subject is provided with self-concerning information from a non-reflexive channel, self-thoughts enable the subject to *anchor* that information. In anchoring, information from a signal that could potentially carry information of different things gets "translated" into a self-thought, that is, a representation that features at least one token of the self-concept. Recall the angry teacher scenario. I parsed the imperative, "Please leave the room!" That imperative could potentially concern Laura or Paul. However, I excluded those hypotheses and thought *He is talking to me*. Here is a schematic representation of the two introduction rules:

Articulation	*Anchoring*
A signal from a reflexive channel	A signal from a non-reflexive channel
Self-thought	Self-thought

We shall not describe "elimination rules" for self-thoughts. There are two reasons for this decision. First, we shall remain neutral on the nature of the immediate precursors of action. Second, there are good reasons to think that self-thoughts are connected with other thoughts. So, we will rather describe two "consumption rules."

PRACTICAL CONSUMPTION. The angry teacher scenario illustrates a typical effect of self-thoughts. Articulated and/or anchored self-concerning information can be used in a practical inference. When

[30] The reflexive character of visual experience is *prima facie* consistent with the claim that the subject is not represented in visual experience (footnote 24). However, some have held that, at least in some cases, the subject is non-conceptually represented in visual experience (Bermúdez, *The Paradox of Self-Consciousness*, *op. cit.*; Peacocke, *The Mirror of the World*, *op. cit.*; Peacocke, *The Primacy of Metaphysics*, *op. cit.*; and John Schwenkler, "Vision, Self-Location, and the Phenomenology of the 'Point of View'," *Noûs*, XLVIII, 1 (March 2014): 137–55). Our approach is consistent with any of these views.

the teacher shouted "Please leave the room!," I thought *He is talking to me* and formed the intention to leave the room. Yet, that self-thought only led me to form the intention to leave the room because it was embedded in a broader inferential network. That inferential network plausibly included a desire to follow the order and the visually based thought: *The exit is over there.*

THEORETICAL CONSUMPTION. We also use the self-concept to perform theoretical inferences. After the publication of the *Critique of Practical Reason*, Kant presumably indulged in the following train of thought: *I am a philosopher. I am old. I am an old philosopher now!* Transitions of this sort may be integrated into a self-conception. A self-conception is a long-standing representation of ourselves that includes information about our own birthday, our place of origin, our relatively stable physical properties, and our deepest convictions and values. We can think of self-conceptions as inferential networks, crystallizations of previous self-thoughts. If Kant's self-conception included a representation of being the author of the *Critique of Practical Reason*, it was probably because he once thought *I am the author of the Critique of Practical Reason.*

Here is a schematic representation of the two consumption rules:

Practical consumption	*Theoretical consumption*
Self-thought	Self-thought
Practical inference	Theoretical inference

IV. TWO EXPLANATORY TOOLS

I shall argue that an indexical rule like *RR* provides a better explanation of the (broad) functional role of the self-concept than the Millian rules that spell out the semantics of Millian names. To this end, I will introduce two explanatory tools: the relations-explain-success principle (RES PRINCIPLE) and the proportionality principle (PROPORTIONALITY). These explanatory tools are quite popular in the literature on mental representation. As we shall see, Millikan herself employs these tools in her own account of mental representation. However, these tools have been largely ignored in the literature on indexicality.[31]

[31] An exception is Christopher Peacocke ("Externalist Explanation," *Proceedings of the Aristotelian Society*, XCIII, 1 (June 1993): 203–30; *Truly Understood*, *op. cit.*; and *The Primacy of Metaphysics*, *op. cit.*). I am highly indebted to his work.

IV.1. The RES Principle.

> RES PRINCIPLE. The relation a representation stands in to its represented domain explains why that representation successfully plays its broad functional role.[32]

Functional roles can be narrowly or broadly individuated. A functional role is broadly individuated just in case its inputs and outputs include relations that "reach out" into the world.[33] It is narrowly individuated otherwise. The RES PRINCIPLE is restricted to broadly individuated functional roles.

The RES PRINCIPLE is highly intuitive. We can appreciate its intuitive force with an analogy. The broad functional role of a map is to help us find our way in a given environment. Crucially, the relation of a map to the mapped domain explains why a normal map user manages to find her way in a given environment. Imagine that Hannah and Pierre are trying to find a restaurant in New York City (NYC). Hannah is using a map of NYC. Pierre is using a map that a crazy joker gave to him. Unbeknownst to Pierre, he was given a map of Paris marked as a map of NYC. Suppose now that Hannah and Pierre are equally good map readers. Suppose also that *all other things are equal*. If Hannah were to find the restaurant, we could explain Hannah's successful outcome by saying that her map stands in a mapping relation to NYC. By contrast, if Pierre's search was unsuccessful, we could explain Pierre's unsuccessful outcome by saying that Pierre's map of Paris does not stand in a mapping relation to NYC. Had Pierre found the restaurant, we could say that Pierre's successful outcome was an accident.

One could think of semantic rules as ways of spelling out the sorts of relations that a representation bears to the represented domain.[34] Some of these relations characterize reference and truth conditions.

[32] Although I coined this principle, it is implicit in the work of theorists with very different views of mental representation. They include Daniel C. Dennett, "Real Patterns," this JOURNAL, LXXXVIII, 1 (January 1991): 27–51; Peter Godfrey-Smith, *Complexity and the Function of Mind in Nature* (Cambridge, UK: Cambridge University Press, 1996); Millikan, *Language, Thought and Other Biological Categories*, *op. cit.*; Millikan, *White Queen Psychology*, *op. cit.*; Ruth Garrett Millikan, *Varieties of Meaning* (Cambridge, MA: MIT Press, 2004); Peacocke, "Externalist Explanation," *op. cit.*; Peacocke, *Truly Understood*, *op. cit.*; Peacocke, *The Mirror of the World*, *op. cit.*; and Nicholas Shea, "Consumers Need Information: Supplementing Teleosemantics with an Input Condition," *Philosophy and Phenomenological Research*, LXXV, 2 (September 2007): 404–35. It goes without saying that these authors develop the RES PRINCIPLE in different ways. In my view, we can safely abstract from these differences in the current context.

[33] Ned Block, "Inverted Earth," *Philosophical Perspectives*, IV (1990): 53–79.

[34] Peacocke, *Truly Understood*, *op. cit.*, p. 57.

Therefore, given the RES PRINCIPLE, reference and truth are explanatorily prior to (broad) functional role.[35]

To be sure, defenders of conceptual role semantics have denied that reference and truth are explanatorily prior to (broad) functional role.[36] My initial reaction to these views is to insist that the RES PRINCIPLE is highly intuitive. Moreover, I have two responses to these authors. First, those who are sympathetic to conceptual role semantics can read this article as a defense of a conditional claim: If the RES PRINCIPLE is true, the self-concept is plausibly understood as a mental indexical. Second, Millikan herself rejects conceptual role semantics and relies on the RES PRINCIPLE in her own account of mental representation.[37] Therefore, our reliance on the RES PRINCIPLE does not beg the question against Millikan.[38]

Millikan's reliance on the RES PRINCIPLE is apparent in her frequent comparison of mental representations to maps. Moreover, the same commitment is apparent in her influential example of the waggle dance of honeybees. Successful foraging honeybees often perform a dance involving a small "waggle" pattern. This dance enables them to share information about the distance and direction of a source of nectar. This dance may cause other bees to fly off to the location of flowers bearing nectar. According to Millikan, different elements of the dance map onto different elements of the represented domain.[39] Roughly, the duration of the dance stands for the distance of nectar from the hive and the angle of the dance to the vertical stands for the direction of nectar from the hive. Based on these observations, Millikan contends that the waggle dance maps onto the state of affairs required to guide the bees when they search for nectar in response to the waggle dance. Crucially, searching for nectar in response to the waggle dance is part of the broad functional role of the waggle dance.

[35] On some views, the relation of a dot on a map to a restaurant is not a referential relation and maps are more aptly characterized in terms of accuracy conditions. These differences are certainly important. However, they will not affect our discussion.

[36] Ned Block, "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy*, X, 1 (September 1987): 615–78; Robert Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge, MA: Harvard University Press, 1994); Gilbert Harman, "Conceptual Role Semantics," *Notre Dame Journal of Formal Logic*, XXIII, 2 (April 1982): 242–56; and Wilfrid Sellars, "Meaning as Functional Classification," *Synthese*, XXVII, 3–4 (September 1974): 417–37.

[37] Millikan, *White Queen Psychology*, *op. cit*, p. 90ff., pp. 130–31; and Ruth Garrett Millikan, "The Father, the Son, and the Daughter: Sellars, Brandom, and Millikan," *Pragmatics and Cognition*, XIII, 1 (January 2005): 59–71.

[38] For a critique of Brandom's conceptual role account of indexical concepts, see Peacocke, *Truly Understood*, *op. cit.*, p. 110ff.

[39] Millikan, *Language, Thought and Other Biological Categories*, *op. cit.*, pp. 98–99.

Therefore, Millikan's analysis of the waggle dance illustrates her commitment to the RES PRINCIPLE.

To be sure, the waggle dance is not a mental representation. However, Millikan thinks that mental representations display a similar structure. Instead of being produced and consumed by different organisms, mental representations are produced and consumed by different parts or aspects of the same subject.

It might be objected that Millikan's account is embedded within a teleosemantic theory of representation. So, our account of the waggle dance is at best incomplete. Millikan would insist that the waggle dance has the etiological function of mapping the nectar-sun-hive domain in a specific way. This point is well taken. However, the RES PRINCIPLE is logically independent from teleosemantics. Therefore, we can safely ignore teleosemantics and focus on the RES PRINCIPLE. This approach has an advantage: even if teleosemantics is false, we can still rely on the RES PRINCIPLE to elucidate the self-concept.[40]

IV.2. Proportionality. It is not easy to tell in the abstract when an explanation is good or bad. Nevertheless, once we agree on an explanandum and are confronted with various competing explanations, we are quite good at telling whether a given explanation is *better* or *worse* than other explanations. Consider the question, "Why did Johnny die?" Suppose that we are offered three rival explanations:

E1. Poor Johnny was hit by an object.
E2. Poor Johnny was hit by a bus.
E3. Poor Johnny was hit by a bus traveling at the speed of light.

Let us suppose now that some buses could travel at the speed of light. However, we live in a world where it is very rare for a bus to travel that fast. In this explanatory context, E2 provides a better explanation of Johnny's death than E1 and E3. The intuition behind this assessment is that E2 provides the right degree of specificity. E1 is not specific enough. Indeed, E1 is compatible with a situation in which Johnny was hit by a very light object (such as a pillow), so Johnny did not die. By contrast, E3 is overly specific. Had Johnny been alerted by his mother that a bus was approaching, he could have escaped his tragic death. But E3 wrongly predicts that Johnny could have not escaped his tragic death. Therefore, E3 is overly specific; it connects the two

[40] Campbell (*Past, Space, and Self, op. cit.*, p. 137) suggests that "the rule of reference and the conceptual role are in reciprocal regulation." The RES PRINCIPLE is consistent with this type of view. Indeed, it provides a way of spelling out the sense in which rules of reference regulate the types of (broad) functions that a representation can fulfill.

events more tightly than they were actually connected in the actual world.

We can therefore say that our explanatory practices are governed by a principle of proportionality:

> PROPORTIONALITY. Explanatory facts or properties must be proportional to their explananda.[41]

We want explanations that are neither unspecific nor overly specific.

PROPORTIONALITY does not give us a magical recipe to identify good explanations. The suggestion is rather that our comparisons of rival explanations are *governed* by the regulative idea of PROPORTIONALITY. To simplify a bit, we can think of the explanandum as a set of phenomena that we are pre-theoretically inclined to group together. Minimally, an explanation seeks to identify a fact or property that correlates with that set of phenomena better than other known facts or properties.

We shall rely on PROPORTIONALITY in our explanatory argument. Before that, we will need to see how PROPORTIONALITY interacts with the RES PRINCIPLE.

IV.3. Relations, Success, and Proportionality. The RES PRINCIPLE says that the relation a representation stands in to its represented domain explains why that representation successfully plays its broad functional role (section IV.1). The RES PRINCIPLE enables us to capture the contrast between cases in which an activity is controlled by a representation that stands in a suitable relation to a given domain (for example, using a map of NYC to find one's way in NYC) with cases in which there is no suitable relation (for example, using a map of Paris to find one's way in NYC). Some maps have the property of mapping a given domain; that is why they are good navigational tools.

PROPORTIONALITY enables us to compare different hypotheses about how a representation relates to a given domain. Suppose that we have been spying on Hannah. We know that she has been using a map of NYC to find her way in the city. But we do not know which map she has been using. Our first hypothesis is that Hannah has been using an up-to-date map of NYC. Our second hypothesis is that she has been using an old map she found in her father's tourist guide from 1997. Although both maps would enable Hannah to find her

[41] Adapted from Stephen Yablo, "Wide Causation," *Philosophical Perspectives*, XI (1997): 251–81, at p. 254. See also Millikan, *White Queen Psychology*, *op. cit.*, p. 221; and Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), p. 65ff.

way in the city, each hypothesis yields different predictions of her navigational success. Were Hannah to be using the 1997 map, she would sometimes fail to find some restaurants that have disappeared since 1997 and overlook the existence of the most recent subway lines.

The RES PRINCIPLE helps us explain Hannah's navigational success in *both* cases. PROPORTIONALITY helps us explain the *difference* in her success rates given the slightly different relation that each map bears to NYC. In the 1997 map, some dots for restaurants have no corresponding restaurant in the city and the most recent subway lines lack corresponding lines on the map. More generally, the different ways in which the two maps relate to NYC are proportional to Hannah's different success rates. A good explanation of these different success rates should identify the differences between the relations that each map bears to the city.

We will employ the same type of argument against Millikan's Millian view. Millian rules do not yield proportional explanations of the broad functional role of the self-concept. Indeed, we have good reasons to think that *RR*—or a similar indexical rule—spells out the sort of relation that we need to offer a good explanation of activities that hinge on articulation and anchoring.

V. THE EXPLANATORY ARGUMENT

Many explanatory arguments have been proposed in the literature on the essential indexical. Some of those arguments point out that indexicals enable us to capture sameness of beliefs and desires among agents who are motivated to act in the same way.[42] Others focus on the transition from no-action to action, as when Perry's messy shopper thinks *I am making a mess*, stops, and rearranges the torn sack.[43] These issues are important. However, the explanatory argument I shall provide is different. It seeks to explain, via a hypothesis about the semantics of the self-concept, a non-accidental pattern. This pattern consists of episodes of anchoring and mental events like forming an intention to act.[44]

[42] Lewis, "Attitudes *De Dicto* and *De Se*," *op. cit.*, p. 525; Dilip Ninan, "What Is the Problem of *De Se* Attitudes?," in García-Carpintero and Torre, *op. cit.*, pp. 86–120, at pp. 100–02; Perry, "Frege on Demonstratives," *op. cit.*; and Torre, "In Defense of *De Se* Content," *op. cit.*, p. 184.

[43] Juliana Faccio Lima, "Indexicality and Action: Why We Need Indexical Beliefs to Motivate Intentional Actions," *Inquiry* (forthcoming), doi: 10.1080/0020174X.2018.1487881; Ninan, "What Is the Problem of *De Se* Attitudes?," *op. cit.*; Perry, "The Problem of the Essential Indexical," *op. cit.*; and Recanati, *Perspectival Thought*, *op. cit.*

[44] Although an intention to act may lead to action, our argument does not focus on the transition from no-action to action. This point will become clear below.

Let us assume that we have a semantic account of predicate concepts, negation, conjunction, and so on. Thus, whenever we have a self-thought, we know what the content of the predicative component is. Our task is to show that, unless tokens of the self-concept conform to an indexical rule, we could not offer a good explanation of the broad functional role of the self-concept. I will proceed in two steps. First, I will argue that the Millian view does not provide a proportional explanation of the broad functional role of the self-concept. Second, I will suggest that the indexical view does provide a proportional explanation of the broad functional role of the self-concept.

V.1. The Millian View. When the angry teacher screamed "Please leave the room!," I engaged in anchoring: I thought *He is talking to me*. Since I wanted to comply with the order, I formed the intention to leave the room. So, the following fact obtained in the classroom:

> LEAVING. The subject who tokened a self-concept (Santiago) then formed the intention to leave the room.[45]

There are reasons to think that LEAVING was not an accident. First, there were neither magical tricks nor angels in the room, only ordinary causal relations. Second, it is easy to cite similar situations where a similar fact obtained. Here is an example. This morning, I went to a coffee shop and ordered a hot latte. After a few minutes, the barista shouted, "Santiago, hot latte!" Being aware that I had ordered a hot latte, I thought *My latte is ready*. I saw a big cup marked "Santiago" waiting on the counter, approached the counter, and grabbed the cup. In this situation, the following fact obtained:

> GRABBING. The subject who tokened a self-concept (Santiago) then formed the intention to grab the cup.

LEAVING and GRABBING have something in common: a token of a self-concept is followed by an intention that seems to be relevant to the thinker of the prior token of the self-concept. It is easy to come up with similar correlations. Those correlations seem to be part of a pattern. That pattern cries out for an explanation.

Here is my argument in a nutshell. These non-accidental correlations are instances of the broad functional role of the self-concept. Given the RES PRINCIPLE, we should explain those correlations with a rule that codifies a suitable relation between the self-concept and

[45] The argument also works if we think of the explanandum as a succession of events: the subject's tokening of a self-concept was followed by the subject's intending to leave the room.

its referent. The Millian relation that connects a Millian name with its referent yields an unspecific explanation. If that Millian relation is supplemented with a direct connection with abilities and dispositions to act, as Millikan suggests, we get an overly specific explanation.

Suppose that the self-concept is a Millian name. This view faces a problem: the Millian relation is not specific enough to explain the target pattern: a token of a self-concept is followed by an intention that seems to be relevant to the thinker of the prior token of the self-concept. We can illustrate this point by comparing LEAVING (C1) with an alternative fact (C2):

Context	Millian Rule	Facts
C1	*Santiago* stands for Santiago	Santiago thought *He is talking to me* & Santiago formed the intention to leave the room.
C2	*Santiago* stands for Santiago	Santiago thought *He is talking to Santiago* & Santiago did not form the intention to leave the room.

Suppose that Santiago masters the Millian name *Santiago* in context C2. Therefore, the Millian relation that links *Santiago* to Santiago holds in C2. However, Santiago ignores the identity I = Santiago. Therefore, Santiago thinks the Millian name *Santiago* but does not form the intention to leave the room.

This case is structurally similar to explanation E1 of Johnny's death. The property of being hit by an object is instantiated when Johnny is hit by a pillow. However, Johnny could be hit by a pillow and not die. Similarly, the Millian relation that connects the Millian name *Santiago* with Santiago is instantiated in context C2. Still, Santiago does not form the intention to leave the room. Therefore, there are contexts in which the Millian relation is instantiated, but Santiago does not form the relevant intention. Assuming that LEAVING is an instance of the broad functional role of the self-concept, it follows that the Millian relation yields an unspecific explanation of the broad functional role of the self-concept.

The Millian theorist needs to exclude context C2. To this end, she can insist that the self-name differs from other Millian names in its narrow functional role. As Millikan puts it, the self-name "hooks up with. . . my abilities and dispositions to act."[46] Given this amendment, there could be no situation in which the *Santiago*-thinker ignores the identity I = Santiago. Upon thinking a *Santiago*-thought, Santiago is immediately disposed to act in various ways. So, the Millian theorist

[46] Millikan, "The Myth of the Essential Indexical," *op. cit.*, p. 273.

can exclude context C2 by distributing the explanatory burden between the Millian relation that connects the self-name with its referent and its narrow functional role. The Millian view adds a *narrow* functional role because it introduces a direct connection between the self-name and some abilities and dispositions to act, and this direct connection does not need to reach out into the world (section IV.1).

This solution can discriminate LEAVING from context C2. In C1, Santiago's self-thought *He is talking to Santiago* was underwritten by a Millian name that was directly connected with Santiago's abilities and dispositions to act. Therefore, it is no surprise that Santiago formed the intention to leave the room. In C2, Santiago's thought *He is talking to Santiago* was underwritten by an ordinary Millian name. Ordinary names are not directly connected with the subject's abilities and dispositions to act. That is why Santiago did not form the intention to leave the room in C2.

Notice, however, that this response imposes an *ad hoc* restriction on the RES PRINCIPLE. This is a problem for someone like Millikan, who relies on the RES PRINCIPLE in her own theory of representation (section IV.1). Her Millian view leads to a disunified account of mental representation. More importantly, the resulting account is unsatisfactory, for it yields an overly specific explanation of the broad functional role of the self-concept. To see why, let us examine two other contexts:

Context	Millian Rule	Facts
C3	*Santiago* stands for Santiago	Santiago thought *He is talking to me* & Santiago formed the intention to stay in the room.
C4	*Santiago* stands for Santiago	Santiago thought *He is talking to me* & Santiago thought *When is he going to stop bullying me?*

The Millian theorist explains the broad functional role of the self-concept by associating it with a Millian rule and introducing a direct connection between the self-concept and the subject's abilities and dispositions to act. However, this strategy cannot easily accommodate contexts C3–C4, where Santiago lacks a disposition to leave the room. In C3, Santiago wanted to challenge the teacher's authority. So, upon thinking *He is talking to me*, he formed the intention to stay in the room. In C4, Santiago's first self-thought led him to retrieve some old memories of past bullying. Those memories motivated Santiago to raise a question: *When is he going to stop bullying me?* Millikan's Millian view is structurally similar to explanation E3 of Johnny's death. The property of being hit by a bus traveling at the speed of light is overly specific. It connects the two events more tightly than they actually

were. This view predicts that Johnny could not have escaped his tragic death. Similarly, the conjunction of the Millian relation and direct connections with dispositions and abilities to act is overly specific. It excludes cases in which a self-thought is not directly connected with dispositions and abilities to act.

The Millian theorist might reply that the disposition to leave the room can be "masked" by background desires and memories. So, contexts C3–C4 are not cases in which the subject lacks the disposition to leave the room. Instead, they are cases in which that disposition was not manifested.

Let us grant that dispositions can be masked. I do think, however, that this line of response misses the point of the second part of the argument. As early critics of behaviorism have pointed out, intelligent action often results from the interaction of different mental states and events.[47] This explains why intelligent responses are highly flexible. To illustrate, Putnam imagined a population of super-Spartans that can suppress all pain behavior. "They do not wince, scream, flinch, sob, grit their teeth, clench their fists, exhibit beads of sweat, or otherwise act like people in pain. . . .However, they do feel pain, and they dislike it (just as we do)."[48] A moral of this example is that intelligent responses result from the complex interaction of various mental states and events. These interactions can result either in no behavior at all or in a rich repertoire of different responses. Something similar occurs with self-thoughts, which interact in complex ways with other mental states and events. Notably, some of these interactions reflect a sensitivity of subsequent mental states and events to the first-personal character of prior self-thoughts. In C3, I thought that the teacher was talking to *me* and *I* wanted to challenge the teacher's authority. That is why I formed the intention to stay in the room. In C4, my initial self-thought triggered an interrogative self-thought. These complex interactions among self-thoughts explain why one and the same self-thought can have different outcomes in different contexts. Crucially, we cannot explain these complex interactions by merely insisting that the disposition to act in a certain way was masked in contexts C3–C4. Instead, we need an account in which tokens of the self-concept can bear semantically relevant relations to other self-thoughts *before they lead to action.*

In sum, direct connections with abilities and dispositions to act can block cases in which Santiago does not act because he ignores

[47] Hilary Putnam, "The Nature of Mental States," in *Mind, Language, and Reality* (Cambridge, UK: Cambridge University Press, 1975), pp. 29–40.

[48] *Ibid.*, p. 332.

the identity I = Santiago. However, this amendment introduces too tight a connection between tokens of the self-concept and action. The Millian theorist overlooks the important fact that self-thoughts can combine with other self-thoughts in inference. Given the inferential promiscuity of the self-concept, its connections with action are flexible.

To be sure, a sufficiently imaginative philosopher could introduce another amendment to the Millian view. I will not speculate about merely possible amendments. I will insist, however, that any amendment to the Millian view will inevitably weaken the force of the RES PRINCIPLE, a key tenet of Millikan's own theory of representation. In what follows, I will argue that the indexical view has a straightforward response to the previous objection. This response has the merit of sticking to the RES PRINCIPLE. Thus, even if I do not manage to convince the skeptic, I will show that the indexical view of the self-concept is not a myth; *pace* Millikan, the indexical view has a lot going for it.

V.2. The Indexical View. The indexical view can steer a middle course between unspecific and overly specific explanations. To see why, I will focus on *RR*. However, the argument also holds for slightly different formulations of the indexical rule that determines the reference of the *I*-concept (footnote 17). The argument goes as follows. The relation codified by *RR* is specific enough to exclude context C2. However, it is not as specific as combinations of a Millian rule with direct connections of the self-concept with abilities and dispositions to act. Therefore, *RR* does not exclude contexts C3–C4.

RR tells us that a token of *I* in a thinking stands for the subject of that thinking. In other words, the referent of a token of *I* is the same as the thinker of that token. Given the RES PRINCIPLE and PROPORTIONALITY, the activities that follow a token of *I* should be proportional to the relation of identity that connects a token of *I* with the thinker of that token. But what would it take for subsequent activities to be proportional to that identity relation? Here is my proposal: all those activities should presuppose the identity of the referent of the token of *I* with the *I*-thinker. In this context, "activities" should be understood in a liberal sense, including bodily and mental acts. By contrast, if a bodily or mental activity does not presuppose the identity of the referent of the token of *I* with the *I*-thinker, that activity does not satisfy the RES PRINCIPLE and PROPORTIONALITY.

Suppose that Perry thinks *I am making a mess* but keeps searching for the messy shopper. In this scenario, Perry has not engaged in an activity that presupposes the identity of the referent of the former token of *I* with the *I*-thinker. Instead, the subsequent activity presupposes that

the referent of the token of *I* is different from the *I*-thinker. By contrast, if Perry then forms the intentions to stop and rearrange the torn sack, he has engaged in activities that presuppose the identity of the referent of the former token of *I* with the *I*-thinker. This approach can discriminate context C1 from context C2. In C1, Santiago was the thinker of *He is talking to me*. Subsequently, he formed the intention to leave the room. That intention presupposed the identity of the referent of the former token of *I* with the *I*-thinker, Santiago himself. Suppose now that *Santiago* is an ordinary, Millian name. *MR* tells us that *Santiago* stands for Santiago. Suppose now that Santiago ignores that I = Santiago. So, upon thinking a *Santiago*-thought, Santiago is not disposed to engage in activities that presuppose the identity of the referent of *Santiago* with the *Santiago*-thinker. Why? The Millian rule for *Santiago* does not codify any identity relation between the *Santiago*-thinker and Santiago. Therefore, many activities that do not presuppose the identity of the token of *Santiago* and Santiago are proportional to the Millian relation. Of course, were the *Santiago*-thinker to know the identity I = Santiago, he would be disposed to engage in activities that presuppose the identity of the referent of the token of *Santiago* with the *Santiago*-thinker. But this could be explained by the RES PRINCIPLE. Santiago knows the identity I = Santiago, and *I* is governed by a rule that codifies an identity relation between tokens of *I* and the thinker of those tokens.

When Santiago forms the intention to leave the room, he engages in an activity that presupposes the identity of the referent of a token of *I* with the *I*-thinker. However, this is not the only type of activity that presupposes that identity. Consider C3. If Santiago wants to challenge the teacher's authority, forming the intention to stay in the room also presupposes the identity of the referent of the first token of *I* with the *I*-thinker, Santiago himself. Or consider C4. The initial self-thought motivates Santiago to wonder: *When is he going to stop bullying me?* Santiago's raising this question is not a disposition to leave the room. However, Santiago's raising this question also presupposes the identity of the referent of the first token of *I* with the *I*-thinker, Santiago himself.

Tokens of the self-concept are broadly relevant to the subject's current projects. However, their relevance is not restricted to their connection with specific abilities or dispositions to act. The self-concept enables us to engage in a broad spectrum of bodily and mental activities. *RR* enables us to offer a proportional explanation of this broad spectrum of activities, for it explains the reference of the self-concept via a thinking relation that connects tokens of *I* with the *I*-thinker. All it takes for a downstream consequence of an *I*-thought to rely on

the relation of a token of *I* with its referent is to presuppose the identity of the referent of the token of *I* with the *I*-thinker. That is why *RR* can, while *MR* cannot, explain the flexible functional role of the self-concept.

VI. OBJECTIONS AND REPLIES

Some philosophers might remain unconvinced. Isn't it plainly obvious that someone—let us say a small child—could "co-opt" her proper name to do what others do with the *I*-concept? Why insist that we *should* analyze her self-concept as a mental indexical?[49]

This common line of argument rests upon a "syntactic fallacy." The objector assumes that, because a subject can co-opt an expression that has the syntactic properties of a proper name, the underlying mental representation belongs to the same semantic type as a proper name. Alas, this line of argument does not address the question at hand: whether the mental counterpart of a proper name in a subject's thinking has the semantics of a proper name.

It might be objected that there is a key difference between linguistic indexicals and the self-concept. Linguistic indexicals systematically *shift* their reference from context to context.[50] By contrast, the self-concept does not seem to shift its reference from context to context. All my tokens of the self-concept refer to me and nobody else. So, the self-concept seems to be more similar to a proper name than to a linguistic indexical.[51]

This argument has two main problems. First, it is one thing to say that the tokens of a representation-type exploit indexical relations to get their reference (indexicality); it is another thing to say that the reference of that representation-type shifts from context to context (shiftability). Shiftability provides good evidence for indexicality. However, there is no good reason to think that shiftability is necessary for indexicality. All we need to reject the Millian view is the weaker claim that an indexical rule determines the reference of tokens of the self-concept. Second, shiftability is a property of representation-types. Suppose now that the *I*-concept is individuated by its reference rule.[52] This view predicts that different thinkers can master the same type of

[49] Perry, "On Knowing Your Self," *op. cit.*, p. 385.

[50] Perry, *The Problem of the Essential Indexical and Other Essays*, *op. cit.*, p. 313.

[51] Cappelen and Dever, *The Inessential Indexical*, *op. cit.*, pp. 16, 24, and 39; Millikan, "The Myth of the Essential Indexical," *op. cit.*, p. 275; and Simon Prosser, "Why Are Indexicals Essential?," *Proceedings of the Aristotelian Society*, CXV, 3 (December 2015): 211–33, at p. 228.

[52] Peacocke, *Truly Understood*, *op. cit.*; Peacocke, *The Mirror of the World*, *op. cit.*; and Peacocke, *The Primacy of Metaphysics*, *op. cit.*

I-concept. If we consider various subjects who master the same type of *I*-concept, it follows that the self-concept does shift its reference. After all, this view entails that different tokens of the same concept-type (the *I*-concept) can refer to different subjects. Its seeming lack of shiftability is an illusion that arises from the fact that each subject has her own capacity to produce *I*-thoughts. But shiftabilty was never meant to be restricted to tokens of one individual capacity. It was meant to range over tokens of the same representation-type.[53]

Millikan thinks that the self-concept cannot be an indexical because "what *defines* the indexical use of a sign is that its context is used by the interpreter to *determine* the content, to determine the referent, not talked about in the content."[54]

Millikan develops this idea as follows: the interpretation of an indexical token always conjoins two independent sources of information: information about which token was produced and information about the context. Unfortunately, self-thinkers do not seem to conjoin two independent sources of information when they produce a self-thought. Consider Perry's shopper case.[55] When Perry thinks *I am making a mess*, he does not need to conjoin information about which token he produced with information about who thought *I*. Perry simply thought a self-thought that disposed him to stop and rearrange the torn sack.[56]

Unfortunately, Millikan's characterization of indexicality is more demanding than it needs to be. We have granted that there are differences between linguistic and mental representations (section II). Yet, this does not undermine a theoretically relevant similarity: both linguistic and mental indexicals get their reference via an indexical relation to their referent. This point generalizes to non-linguistic representations. Cognitive psychologists would not reject the existence of

[53] The Millian view conflates the plausible claim that all my tokens of *I* refer to me with the controversial claim that each self-thinker has a private representation-type (for example, *Santiago*) that only refers to her herself. If concept-types must be shareable (Jerry A. Fodor, *LOT 2: The Language of Thought Revisited* (New York: Oxford University Press, 2008)), the Millian view has the additional problem of preventing different subjects from sharing the same type of self-concept. This is a serious problem for those of us who want to formulate intersubjective psychological generalizations.

[54] Millikan, "The Myth of the Essential Indexical," *op. cit.*, p. 272. See also Tomis Kapitan, "Indexical Identification: A Perspectival Account," *Philosophical Psychology*, XIV, 3 (Summer 2001): 293–312, at p. 294.

[55] Perry, "The Problem of the Essential Indexical," *op. cit.*

[56] *Pace* Millikan, some philosophers have argued that subjects do conjoin different sources of information when they produce an *I*-thought: their tacit mastery of *RR* and the (apparent) action-awareness that accompanies their exercises of mental agency (O'Brien, *Self-Knowing Agents*, *op. cit.*; and Peacocke, *Truly Understood*, *op. cit.*). I will bracket these views for the sake of the argument.

mental maps on the grounds that a normal map user needs to compare marks on a map with landmarks in the environment. When a cognitive psychologist posits the existence of mental maps, she is interested in a more abstract property that some mental structures share with public maps: a mapping relation between a mental vehicle and the environment that preserves some spatial relations. By parity of reasoning, all we need to refute the Millian view is the weaker claim that the reference of the self-concept is determined by an indexical relation.

To be sure, we do have another self-representation that seems to work like a proper name. I called this representation a "self-conception" (section III). To repeat, a self-conception is a long-standing representation of ourselves that includes information about our own birthday, our place of origin, our relatively stable physical properties, and our deepest convictions and values. We can see our self-conception as the repository of information incorporated via articulation and anchoring. Given their stability over time, self-conceptions might have a Millian semantics.[57] Yet, insofar as we acknowledge the existence of episodic self-thoughts that we use to articulate and anchor self-concerning information, we still need to posit a concept-type that is governed by an indexical rule like *RR*.

VII. THE ESSENTIAL INDEXICAL

Many philosophers have been interested in the question of whether indexicals are essential.[58] Many of these authors cite Millikan's seminal article as a precursor to current *de se* skepticism. Let me conclude with some ramifications of my discussion for the current debate.

There are many intertwined issues in the debate on the essential indexical. The Millian view targets the claim that the self-concept has an indexical semantics. However, many philosophers are interested in other types of questions. Consider two representative examples: Are indexical representations *necessary* in our mental life? Is the *I*-concept *necessary* in our mental life?

In my view, meeting Millikan's challenge is philosophically prior to answering the latter questions. If it makes little sense to hold that

[57] *Pace* Dennett, *Consciousness Explained*, *op. cit.*

[58] Bermúdez, *Understanding "I"*, *op. cit.*; Cappelen and Dever, *The Inessential Indexical*, *op. cit.*; Castañeda, "'He'," *op. cit.*; Castañeda, "Indicators and Quasi-indicators," *op. cit.*; García-Carpintero, "The Philosophical Significance of the *De Se*," *op. cit.*; Lewis, "Attitudes *De Dicto* and *De Se*," *op. cit.*; Magidor, "The Myth of the *De Se*," *op. cit.*; Daniel Morgan, "Accidentally about Me," *Mind*, CXXVIII, 512 (October 2019): 1085–115; Ninan, "What Is the Problem of *De Se* Attitudes?," *op. cit.*; Perry, "The Problem of the Essential Indexical," *op. cit.*; and Prosser, "Why Are Indexicals Essential?," *op. cit.*

some mental representations have an indexical semantics, our answer to the latter questions should be "no." That is why Millikan's important paper has a central role in the current debate on the essential indexical. Furthermore, the tools one deploys to defend an affirmative answer to the former question will end up imposing substantial constraints on one's answer to the latter questions. The explanatory argument shows that an indexical rule like *RR* provides a more proportional explanation of some patterns than the explanations provided by a Millian rule, even if that rule is combined with direct connections with abilities and dispositions to act. Nevertheless, this and other explanatory arguments have methodological limitations that will prevent us from *demonstrating* that the *I*-concept (or any other indexical concept) is metaphysically necessary.

First, our explanatory argument relies on empirical observations of the broad functional role of a putative indexical concept. Even if these empirical observations are true, it is hard to show that the alleged broad functional role is *constitutive* of our mental life. Second, even if the alleged broad functional role is constitutive of our mental life, explanatory arguments could not demonstrate that the *I*-concept is *metaphysically necessary* to fulfill that role. Consider our descriptive claim that the self-concept enables us to articulate and anchor information for use in practical and theoretical inference (section III). Even if our explanatory argument is persuasive, some possible subjects could have slightly different types of self-concepts. Think of subjects who possess an *I**-concept that is governed by *RR* plus some weird condition, like raising one's left hand. Nevertheless, these methodological limitations do not make explanatory arguments *irrelevant* to the issues raised by the problem of the essential indexical. We could see explanatory arguments as providing non-demonstrative considerations in favor of a relative claim: indexical concepts are metaphysically necessary to fulfill some functional roles.

Philosophers have often argued that indexical concepts are necessary for action.[59] Many critics have rejected this claim (footnote 24). I have suggested that the self-concept enables us to engage in complex forms of information integration from reflexive and non-reflexive channels for subsequent use in inference. This provides a new perspective on Perry's shopper case. For Perry, the indexical belief *I am making a mess* was meant to explain his stopping and rearranging the

[59] Colin McGinn, *The Subjective View: Secondary Qualities and Indexical Thought* (Oxford: Clarendon Press, 1983), p. 104; and David Owens, "Deliberation and the First Person," in Hatzimoysis, ed., *Self-Knowledge*, *op. cit.*, pp. 261–77.

torn sack.[60] Our treatment suggests a different analysis. Perry had information that *someone* was making a mess. He gained that information through the non-reflexive aspect of vision (section III.1) conjoined with assumptions about supermarkets and sacks of sugar. Perry then noticed that the trail of sugar was becoming increasingly thicker, until it dawned on him that *he himself* was the shopper he was trying to catch. What happened at the epiphany? Perry engaged in anchoring. He formed an *I*-thought in response to information from a channel that also delivers information of other people. However, Perry's *I*-thought did not need to have any direct connections with action. He could have simply formed another thought (for example, *Oh, I'm so clumsy!*). Or he could have stored his *I*-thought into his self-conception (for example, *I once made a mess with a torn sack of sugar*).

Cappelen and Dever attack what they call the "Impersonal Incompleteness Claim" (IIC). Roughly, impersonal action explanations are necessarily incomplete because of a missing indexical component.[61] Cappelen and Dever insist that impersonal action explanations can *also* offer action explanations. Although we have not discussed the explanations Cappelen and Dever are interested in, our explanatory argument enables us to grant Cappelen and Dever's point without endorsing their skepticism about mental indexicals. You can certainly explain Johnny's death by citing the fact that he was hit by an object. However, this explanation is not as good as another explanation that cites the fact that Johnny was hit by a bus. Similarly, you can explain my forming the intention to leave the room by citing the fact that Santiago produced a thought that included a token of a concept that refers to Santiago. This explanation would not discriminate between the Millian view and the indexical view. Unfortunately, this explanation would not be as good as another explanation that introduced an indexical concept governed by an indexical rule like *RR*, for *RR* can, while *MR* cannot, discriminate context C1 from context C2. Moreover, *RR* predicts that the self-concept bears a flexible relation to action, as shown by contexts C3–C4. The issue is not whether impersonal explanations are available. The issue is whether they are as good as indexical explanations. If our explanatory argument is convincing, indexical explanations are better than non-indexical explanations.[62]

[60] Perry, "The Problem of the Essential Indexical," *op. cit.*, p. 36.

[61] Cappelen and Dever, *The Inessential Indexical*, *op. cit.*, p. 37. See also Magidor, "The Myth of the *De Se*," *op. cit.*, p. 259.

[62] García-Carpintero ("The Philosophical Significance of the *De Se*," *op. cit.*, p. 259f.) relies on John Stuart Mill's method of difference (*A System of Logic* (London: John W. Parker, 1843)) and some phenomenal contrasts between *de se* and other mental

VIII. CONCLUSION

The Millian view holds that the self-concept is not essentially indexical, for it is best construed as a Millian name that differs from other names in its narrow functional role. I have proposed an explanatory argument against the Millian view and in favor of the indexical view. The explanatory argument exploits some tools from the literature on mental representation: a principle that connects the relations of a representation to its represented domain with the capacity of that representation to fulfill its broad functional role (RES PRINCIPLE) and another principle that governs our explanatory practices (PROPORTIONALITY). Our explanatory argument does not focus on the usual explanatory facts, like detecting mental commonalities across subjects or explaining the transition from no-action to action. Our explanatory argument focuses on the relevance of semantic rules to explain some correlations between *I*-thoughts and subsequent mental events. These events include forming the intention to act, but also forming other thoughts, and storing self-concerning information in memory. Our explanatory argument does not stress that an indexical element must be added to non-indexical explanations to turn them into complete explanations. It rather identifies some patterns that are best explained by indexical concepts. Our explanatory argument shows the explanatory relevance of the indexical rule that fixes the reference of the self-concept, so it avoids the recent complaint that the problem of the essential indexical can be reduced to the problem of opacity for proper names. To be sure, this explanatory argument will not convince those theorists who reject (one of) our explanatory principles. However, rejecting those principles would incur a serious cost for those theorists who think of reference and truth as explanatorily prior to broad functional role. Those who remain unconvinced can see the arguments presented here as a challenge. At the very least, they show that mental indexicals are not a myth.

SANTIAGO ECHEVERRI

Universidad Nacional Autónoma de México

---

attitudes to defend IIC. My explanatory argument has the advantage of being third-personal. So, it does not make any assumptions about cognitive phenomenology.

Cappelen and Dever (*op. cit.*) and Magidor (*op. cit.*, p. 258, p. 265) complain that many examples of essential indexicals are *mere* Frege cases. Our argument blocks this objection as well. Even if these examples are Frege cases, our explanatory argument shows that indexical rules make a crucial difference to the goodness of some explanations.