

[This is an accepted manuscript of an article forthcoming in *Philosophical Explorations*. Please quote from the published version, available online: <http://www.tandfonline.com/>]

**Thinking of oneself as the thinker:
The concept of self and the phenomenology of intellection**

Marie Guillot*

Department of Philosophy, University College London, London, UK

The indexical word “I” has traditionally been assumed to be an overt analogue to the concept of self, and the best model for understanding it. This approach, I argue, overlooks the essential role of cognitive phenomenology in the mastery of the concept of self. I suggest that a better model is to be found in a different kind of representation: phenomenal concepts, or more generally phenomenally-grounded concepts.

I start with what I take to be the defining feature of the concept of self, namely its “super-reflexivity”: to use this concept isn’t just to think of oneself, but to think of oneself as the thinker of the present thought. I call this familiar observation the “Thinker Intuition”. I review some shortcomings of the indexical model of the concept of self, which is the classical account of the Thinker Intuition. I go on to propose a different account, the “phenomenal model”, according to which the concept of self is a phenomenally-grounded concept, anchored in a generic kind of cognitive phenomenology: the phenomenology of intellection.

Keywords: the concept of self; super-reflexivity; indexicality; phenomenal concepts; cognitive phenomenology; the phenomenology of intellection.

1. Introduction

1.1. *The concept of self, super-reflexivity and the Thinker Intuition*

The aim of this essay is to present and defend a new model for understanding the concept of self, construing it as a phenomenally-grounded concept anchored in a generic form of cognitive phenomenology.

By “the concept of self”, I do not mean the highly abstract, general concept philosophers use to discuss, for instance, the nature of selfhood, the determinants of personal identity, or selves as a natural kind. I mean the concept that each of us uses to reflect on and store information about themselves, specifically. It is the concept at play when I engage in introspection, self-criticism, deliberation and action-planning: when I wonder whether *I* should join some friends on a camping trip, when I notice with a tinge of self-reproach that *I*’m feeling lukewarm, when I go over my memories of the last holiday and ask myself whether *I* really had as good a time as *I* claimed, when I admit to myself that *I*’m not so fond of camping after all, when I make the decision to stay at home by *myself* and let the others go without *me*.

* Email: m.guillot@ucl.ac.uk

It has long been noted¹ that this concept of self, or I-concept for short, is the concept one uses not just to think about oneself, but to think about oneself *as oneself*. To appreciate the difference, consider situations where one thinks of someone who happens to be oneself, without realising that that is the case. Examples of this include Oedipus as he thinks of the murderer of King Laius while ignorant of his own guilt, or the winner of a lottery when she thinks, before the results are announced, of *the winner of the lottery* with some envy. Oedipus, and the unwitting lottery-winner, think of the person they *happen to be*, but they fail to think of this person *as themselves*².

What does it mean to think of oneself *as oneself*? Oedipus's thought "The murderer of Laius must be found" is *reflexive*, in a basic sense: the person thinking and the person he thinks of are, as a matter of fact, one and the same. In comparison with this *simple reflexivity*, the thought that *I don't like camping trips* is what we might call '*super-reflexive*'³. In this case, it is not just that the subject of the thought is identical with its object. That very fact is also part of what I grasp in having the thought. The thought is thus reflexive not just in being a relation of the form $R(x,x)$, a representation of person x by x herself, but in that the reflexive aspect itself is, somehow, also reflected by the representation. I am aware, in other words, that the person being thought of is none other than the person on the other side of the relation of representation, *the person thinking the thought*. Put more concisely, when I think of myself under the I-concept, I think of myself as *the thinker* (of this very thought)⁴. Someone like Oedipus is *in* a reflexive relation of representation to themselves. But someone who thinks "I am F" is somehow also aware *of* the reflexive relation, and of the identity between the two *relata*, the subject and object of the representation. Intuitively, this is the defining feature of the I-concept⁵. No way to think of oneself counts as an I-concept if it fails to display this redoubled form of reflexivity, or *super-reflexivity*⁶. Let us call this the "Thinker Intuition".

1.2. Plan

Since the Thinker Intuition captures the defining feature of the I-concept, we should expect any account of the way the I-concept presents us with its referent to do justice to this intuition. The intuition can thus be used as a test to compare the merits of competing approaches, based on how good they are at explaining it.

The classical way of capturing the Thinker Intuition is by construing the I-concept as an *indexical concept*, a mental representation analogous to linguistic expressions such as "I", "now", "here" or "today". The conventional meaning of the word "I", as used in a particular utterance u , is something like "the person who utters u ". Likewise, it is commonly argued, someone who uses the I-concept as part of a particular thought t must be thinking of herself under a description such as "the person who thinks t ". This approach, however, raises some well-known difficulties, briefly discussed in Section 2.

The goal of this essay is to present an alternative way of fleshing out the Thinker Intuition, which I will call the *phenomenal model*. My approach, described in Section 3, rests on two claims. The first claim is that *phenomenal concepts*, and phenomenally-grounded concepts more generally, provide a good model for understanding the I-concept. Phenomenal concepts are partly individuated by what I will call their *phenomenal basis*, namely the phenomenal quality in which they are grounded. The second claim is that the phenomenal basis distinctive of the I-concept is a generic form of cognitive phenomenology: the phenomenology of intellection. Section 4 presents some of the advantages of the view, and Section 5 discusses some potential objections.

2. The indexical model⁷

2.1. Token-reflexive representations

The Thinker Intuition is the likely motivation behind a range of views according to which the I-concept is an analogue, in the mental realm, to the word “I” in our linguistic practice.⁸ The word “I” belongs to the class of indexical expressions, like “you”, “this”, “here”, “now”, “henceforth”, etc. Indexicals determine their referent, on each occasion of utterance, by exploiting contextual facts about that very utterance. Whenever an utterance *u* of the word “I” is made, for instance, the conventional meaning conveyed is roughly “*The speaker of this utterance u of I*”. In this description, the value of *u* is supplied by the context, as the very utterance which also expresses the meaning under consideration. More generally, the standing meaning of indexicals is captured by descriptions of a special kind, like the italicised one above, which make reference to the utterances (tokens) of the expressions themselves. For this reason, those expressions are sometimes called “token-reflexives”⁹.

On the indexical model of the I-concept, the same kind of token-reflexivity is made use of in the mind as in speech. Where the word “I” exploits linguistic tokens, the concept of self exploits *mental* tokens: occurrent thoughts. An utterance *u* of the word “I” determines its referent as the object which happens to bear a certain contextual relation to *u* (i.e., the relation “*x is the speaker of u*”). In a similar fashion, a tokening *t* of the I-concept in a thought determines as its referent the individual bearing an analogous relation to *t*, namely “*x is the thinker of t*”. Furthermore, this token-reflexive reference-rule, according to the indexical model, is what we entertain in using the concept of self. The cognitive significance of the concept of self is identified with a special kind of description: “being the thinker of this very thought”, to simplify somewhat. This does justice to the Thinker Intuition because it has the consequence that no one can use the I-concept without thereby thinking of oneself, descriptively, *as* the person who is thinking the present thought.

There is a complication, though. Early proponents of the token-reflexive approach thought that what an indexical *meant* just was the token-reflexive description associated with it. This had the unwanted consequence¹⁰ that a sentence such as (1) “I am saying this” became synonymous with (1’) “the person saying this is saying this”. But (1’) is both *a priori* and necessarily true, while (1), although *a priori* true, is only contingently so. Intuitively, the contrast is due to the fact that (1) is about a specific *individual* (who might have said something else, or kept quiet), while (1’) is about a *condition* (“being the one who says this”), trivially fulfilled by whoever satisfies the predicate of the sentence.

This modal dissonance led to a refinement of the theory, of which Kaplan (1989) is the standard version. According to Kaplan, the token-reflexive description “being the speaker of *u*” that is conventionally attached to “I” is only one aspect of its meaning, which he calls its “character” and distinguishes from its “content”. The character is only a *rule* for determining the content of the word in each context (namely the actual person answering the description¹¹); the rule is not itself the content. Thus, a particular use *u* of the word “I” stands, not for a condition (“being the speaker of *u*”) but for a concrete individual (who satisfies the condition). And it stands for that same individual in every possible world, so that the word refers *rigidly*, contrary to ordinary descriptions.

The same should be said for the concept of self on an indexical approach. To *grasp* the concept is to grasp the description “being the thinker of this very thought”, and thus to be possessed of a rule capable of determining a referent in any given context of thought; but to *apply* it on a particular occasion is to refer to the *individual* answering to the description in this instance, rather than to the condition expressed by the description.

2.2. *Some difficulties with the indexical model*

The refined indexical model blocks the modal objection, and some other issues raised by the simple indexical model. However, the two variants equally invite a few further questions. These have to do with the central (if different) role played by the token-reflexive description “the thinker of this very thought” on either variant. Both versions, it should be noted, are *descriptivist* approaches in some sense. Only the simple indexical model is descriptivist in assuming that indexical representations have the same non-rigid semantic behaviour as ordinary descriptions. However, both versions converge in identifying the *cognitive significance* of an indexical representation with a description. What the subject grasps, in deploying the concept of self in her mind, is the token-reflexive description. On the refined indexical model, the description may be denied any truth-conditional relevance, but the subject still needs to apprehend it as a route to the referent, a way to direct her mind to the right object.

Identifying the cognitive import of the I-concept with the description “the thinker of this very thought” raises several issues. A first issue has to do with the “this” in “this very thought”. To think of “the thinker” is to think of *oneself* only if one thinks of the thinker of a specific thought, namely *this* thought, the very thought through which one is representing the relevant individual. This raises the question of how the thought in question is made reference to in the token-reflexive description. How does one latch onto the correct thought, the *present* thought?¹²

Two main responses are likely to be made. One option is to say that the description “the thinker of this very thought” contains another indexical representation, “this”. This opens a further question as to how this new layer of indexicality, in turn, is analysed. If this involves a new token-reflexive description, the model needs to make clear how the risk of an infinite regress is to be avoided.

A second option is to claim that our present thoughts are just “given” to us in a special way. It is plausible that *conscious* thoughts, at any rate, are thoughts we are aware of *as* presently occurring. A natural way to account for this is to say that conscious thoughts are *phenomenally* conscious¹³: there is something it is like to be having them, and it is thanks to this phenomenal ‘mark’ that we can focus on the right thought as it unfolds. I do think that the distinctive phenomenology of conscious thoughts has a key role to play in accounting for how the I-concept works. But, as will become clear in the next section, I believe that this role can be captured in a simpler way than by making it an ingredient of the indexical model. In any case, resorting to such a strategy would make the indexical model a *hybrid* one. It would mean accepting that token-reflexive descriptions, by themselves, are not *sufficient* to secure the super-reflexivity¹⁴ that characterises the concept of self. I do not want to suggest that there is anything wrong with a hybrid model, but only to point out an important, open question on which it is legitimate to expect an indexical approach to say something substantive.

A distinct but related issue to be addressed has to do with a feature of the I-concept which we might call its *exclusivity*. Although anyone can think about me, no one can think of me in the way that I do when I use the I-concept. I am an object open to the public gaze, but my self-conception is private. This is probably part of what Frege (1918-9) has in mind when he claims that “Every one is presented to himself in a special and primitive way, in which he is presented to no one else.” Now, in drawing its inspiration from a linguistic model, the indexical approach makes it harder to account for exclusivity. The conventional meaning of the word “I” – “the speaker of this utterance” – is a description that can be grasped by anyone, speaker and hearers alike, and one that all interlocutors can equally use as their route to the

referent. In other words, when I say “I don’t like camping” and you understand what I say, we both think of me in the same way. Transposing this model into the mental realm, as the indexical model of the I-concept does, creates a tension with the exclusivity requirement¹⁵. Suppose that to think of oneself as oneself is to think of “the thinker of this thought”. Why couldn’t someone else think of my present thought, and thus of the thinker of this particular thought, and in so doing think of me in the same way that I do when I use the concept of self?

Of course, one might want to respond by insisting on an important disanalogy between the linguistic and the mental cases. An utterance is a public object, produced for all to experience; a thought token, by contrast, is an event that only the thinker can experience. Others may *think* about my current thought, but it is not *given* to them as it is given to me, at least when it is a conscious thought. And this might be grounds for arguing that in thinking of “the thinker of this thought” I think of myself in a way that no one else can. But this brings us back to the indexical model’s need, evidenced by the earlier point, to rely on the phenomenology of conscious thoughts¹⁶ and to say something substantial about this.

The indexical model raises a third, more worrying issue. The token-reflexive description it identifies as the cognitive significance of the I-concept is a fairly complex conceptual representation. One might object that this commits the approach to an implausible form of intellectualism. A grasp of the I-concept, on this view, presupposes the prior grasp of further, fairly abstract concepts, at least the concept of a *thinker* and the concept of a *thought*; or, if one wants to paraphrase the token-reflexive description differently, concepts of an equivalent level of generality. But it is unclear whether a small child who has just come to master the I-concept, for instance, is also able to think of thoughts, or thinkers, as such. While the previous two points suggest that the grasp of a token-reflexive description, *per se*, isn’t a sufficient condition to grasp the I-concept, the present point suggests that it may not be a necessary condition either.

3. The phenomenal model

3.1. Back to the Thinker Intuition

The indexical model may seem like a straightforward account of the Thinker Intuition – the intuition that in using the concept of self, one thinks of oneself as *the thinker of this very thought*. However, because of its descriptivist nature, the model raises some unresolved issues, the most problematic of which might be its strong intellectualism.

I want to suggest an alternative new way to explain the Thinker Intuition. Mental indexicality, if it does indeed exist, isn’t the only mental feature that connects us to our present thinking, as such. I have alluded earlier to another, direct way in which we become aware that we are currently thinking: by *experiencing* it. Our most basic grasp of ourselves as thinkers is grounded in the phenomenology attached to conscious episodes of thought, or what is often referred to as “cognitive phenomenology”. The hypothesis I want to put forward is that this phenomenology of thinking plays a more central role in explaining how the I-concept works than has hitherto been acknowledged.

In the rest of this essay, I present the view that the concept of self is a *phenomenal concept*, or at least a phenomenally-grounded concept; and that the distinctive phenomenal quality it is anchored in is a generic form of cognitive phenomenology, which I will call the *phenomenology of intellection*.

Let me go in more detail over the main ingredients of this approach, which I will call the “phenomenal model”, before explaining how it accounts for the Thinker Intuition.

3.2. *Phenomenal concepts*

The word “I” has traditionally been assumed to be an overt analogue to the concept of self, and the best model for understanding it. I would like to suggest that a better model is to be found in a very different kind of representation: *phenomenal concepts*.

Phenomenal concepts – e.g. the concepts “greyish”, “sweet-smelling”, or “painful” – are those we use to reflect on, classify and compare our experiences as such, regardless of the objects they are experiences of. For example, when I concentrate on the sensation of pain in my foot, *as a sensation*, rather than on the information about damage in my foot’s tissues that this sensation may convey, the concept I am using to categorise this pain – e.g. as an itch, a throbbing or a burning pain, or a cramp – typically is a phenomenal concept. Likewise when I introspect and reflect on the bright hypnagogic lights dancing before my eyes after I’ve rubbed them, the unfamiliar taste of cactus, or my feeling of gratefulness at the end of an illness.

Following Gertler (2012), I think of an experience as a special type of event: an instantiation, in a subject, at a time, of a specific phenomenal property (which can sometimes be a complex of simpler phenomenal properties, and can also have a multimodal origin, for instance when we have a unified experience of the taste, texture and temperature of a piece of buttered toast). Phenomenal concepts are, then, those concepts we use to reflect on the phenomenal properties whose instantiations constitute our experiences.

Please note that none of this, or of what follows, entails any particular commitment regarding the metaphysics of phenomenal properties. In particular, I wish to remain neutral as to the relationship between the phenomenal properties instantiated by experiences, and the physical properties of the objects of those experiences. Views on this matter vary greatly, but it is widely accepted, across those views, that there is at any rate “something it is like” to have an experience, a way it qualitatively affects our state of awareness, a subjective impression marking its occurrence. It is also a matter of wide agreement that one can focus on the manifest properties of the subjective impression itself, the “what-it’s-like-ness”, as *prima facie* distinct from the properties of the object that the experience is an experience of, whatever the relation between the two domains might turn out to be. In what follows, the expression “phenomenal properties”, then, is simply a way to latch on to the first set of properties. Phenomenal properties in this non-committal sense are just aspects of our subjective impressions. In keeping with established practice (see e.g. Levine 2001), I will talk (for instance) of “reddishness” to refer to that aspect of a visual impression I typically enjoy when I look at a red rose, as opposed to a white one. To repeat, this manner of speaking does not reflect any particular view on the nature of the “reddishness” of the experience, or on its relationship with the redness of the rose.

While the general function of phenomenal concepts – to enable us to reflect on our experiences as such – is clear, the details of how they work are disputed. As Papineau (2008) notes, existing positions range from taking them to be a *sui generis* type of concept (e.g. Chalmers 2003), to attempts at reductions to other kinds of concepts, including recognitional concepts (Loar 1990), demonstrative concepts (Perry 2001) or quotational concepts (Papineau 2008). However, one key feature is uncontroversial. Phenomenal concepts are in some sense anchored in phenomenally conscious states of the subject. This intuition can be made more precise from different angles.

First, one cannot *acquire* a given phenomenal concept without first *having* the relevant experience. Anosmics, who don’t have sense of smell, simply cannot think of what it is like to

smell roses under the phenomenal concept others typically use (although they can, of course, refer to that experience in other ways, for instance through a description, e.g. as the experience produced in normal subjects by a certain chemical compound).

Second, the relevant phenomenal experience appears to play a crucial role in the way the concept, once acquired, *works*. The experience – the olfactory impression as of roses or the itchy sensation, for instance – is used as a “label” or “template” to categorise the experience itself, and further encounters with experiences of the same kind. We fix our mind’s eye, so to speak, on a piece of phenomenology – a reddish shade, or a particular itchiness, for instance – and think, it is *this* hue, *this* unpleasant sensation, where this is not a matter of filing them under a particular description, or just mentally “pointing” at them, but of isolating the relevant phenomenal properties through their own instantiation, in something like the way we pin down a particular shade of red paint by looking at a little square of that very paint in the manufacturer’s catalogue. As some¹⁷ put it, the experience serves as its own mode of presentation. Or, as Chalmers (2003, 235) says: “The clearest cases of direct phenomenal concepts arise when a subject attends to the quality of an experience, and forms a concept wholly based on the attention to the quality, ‘taking up’ the quality into the concept.”¹⁸

In other words, the conceptual representation of a particular experience is partly *constituted* by the experience itself. A phenomenal concept, we may say, works by *sampling*. The relation between the concept and the experience it represents involves an exemplification of that very experience.

I propose to call the experience that a given phenomenal concept ‘samples’, and thus enables us to refer to, the *phenomenal basis* of that concept. It is the phenomenal property that one needs to be acquainted with in order to acquire a specific phenomenal concept, and that partly constitutes that concept. The phenomenal concept “itchy”, for example, has as its phenomenal basis a particular somatosensory feeling, the feeling of itchiness. The phenomenal concepts “rose-scented”, “painful” or “elated” have as their respective phenomenal bases a sensory phenomenal property, an algedonic phenomenal property, and an emotional phenomenal property. Phenomenal concepts are individuated in part by their phenomenal bases.

3.3. The phenomenology of intellection

One particular kind of phenomenology which is enjoying increasing attention is the phenomenology of conscious thinking. There is something it is like, not just to have perceptual experiences, feelings of pain and pleasure, or emotions, but also to have thoughts passing through one’s mind. Conscious episodes of using our conceptual capacities, be it in making judgments or forming hypotheses, being struck by a doubt, entertaining a possibility, reaching certainty about some proposition or coming to reject it, all register in a certain way in our stream of consciousness. This is what is often called “cognitive phenomenology” (Bayne and Montague 2011).

Here, I would like to focus on the most generic form of this kind of phenomenology. To explain what I have in mind, it is worth mentioning a couple of terminological ambiguities attached to the expression “cognitive phenomenology”.

The first way in which the kind of cognitive phenomenology I’m interested in is *generic* concerns the attitudinal level. The idea that conscious thoughts are marked by a characteristic phenomenology is often spelled out by way of a contrast with perceptual (and sometimes algedonic or affective) attitudes. This is how Pitt puts it:

(...) conscious *thoughts* can't be identified with any other sort of conscious states (...). Thoughts are states of a different kind from all others. Hence, there must be a *proprietary phenomenology of cognition* - (...) a phenomenology that makes a state cognitive, as opposed to visual, auditory, olfactory, somatic, proprioceptive, etc. (Pitt 2011:145; my emphasis.)

The words “cognition” and “thought”, though, are ambiguous. Cognition can be a matter of getting to know, and hence forming beliefs; or it can simply be a matter of processing information, without treating it as true or false. Similarly, as Kriegel (2015) remarks, we can be said to have a thought when we think *that* something is the case, but also when we merely think *about* it, in what can be a wholly neutral way, as when we just consider *whether* something is the case or not, or remain indifferent to it (Kriegel 2015, 116 sq.) A thought can be an occurrent attitude that has assertoric force, like a judgment that *p* or (in the case of a negative force) a rejection of *p*. But it can also be the mere contemplation of the proposition that *p*, which is non-committal as to its truth or falsity.

That thinking-that, i.e. judging (or denying) that *p*, comes with a characteristic ‘feel’, has long been observed. Hume, for instance, famously claimed that

(...) belief consists merely in a certain feeling or sentiment (...). When we are convinced of any matter of fact, we do nothing but conceive it, along with a certain feeling, different from what attends the mere reveries of the imagination. (Hume 1739, Appendix).

However, thinking in the neutral sense of merely entertaining or grasping a proposition *p* also has a certain phenomenological signature.¹⁹ I can be conscious and alert while my mind is wandering, while I muse and aimlessly toy with ideas; or while I concentrate on a question I haven't yet made up my mind about, pondering whether to answer it in the positive or the negative. And it feels a certain way to be in those non-committal states.

While this experience might be more tenuous and elusive than when we judge or deny (or doubt or hypothesize), it is nonetheless present. I would like to suggest that we view the experience as registering our *understanding* of the meaning of a certain proposition that we are currently holding in our mind: as a feeling of “semantic uptake”, or of pure intellection. This way of considering the phenomenology of thinking-about might help explain why it is harder to characterise, and perhaps to focus on, than the phenomenology of thinking-that. To judge or doubt or hypothesize that *p* requires that we first grasp the meaning of *p*. So the phenomenology of intellection, on this approach, is present whenever we engage in in any kind of conscious intellectual activity, and forms a sort of baseline in the symphony of our cognitive life. It is present on its own when we are merely contemplating a proposition, but it is also an ingredient of the more complex phenomenal states involved in judging, doubting or denying. If this is correct, then the very ubiquity of the phenomenology characteristic of entertaining a proposition is what makes it harder to describe.

To mark the contrast between those two senses of the notion of cognitive phenomenology, assertoric *vs.* neutral, I'll use the expression “phenomenology of intellection”²⁰ to talk about the latter. I'll assume here that the phenomenology of intellection, in this sense, is present in every conscious episode of conceptual mentation, across all the attitudes of judging, doubting, denying, contemplating, etc. It is this attitudinally generic phenomenology of intellection that I want to focus on.

The second sense in which the notion I want to isolate is generic has to do with content. Writers who accept the existence of cognitive phenomenology generally agree that it

comes in different shades, depending on the kind of attitude involved. But some think that it is further specified by the kind of content entertained under a given attitude. To quote Pitt again:

In addition to arguing that there is something it is like to think a conscious thought, I shall also argue that what it is like to think a conscious thought is distinct from what it is like to be in any other kind of conscious mental state, and that what it is like to think the conscious thought that *p* is distinct from what it is like to think any other conscious thought (...). (Pitt 2004:2)

This final claim, however, is not accepted by all advocates of cognitive phenomenology (see e.g. Chudnoff 2015). I will remain neutral as to whether different types of propositional content give rise to different kinds of phenomenal signatures, more fine-grained than the kinds typified by the attitudes. Both sides of this in-house debate are at least committed to a “generic” cognitive phenomenology for each attitude *A*, whether they construe it as all the cognitive phenomenology there is for *A*, or as a phenomenal property that all the more fine-grained experiences associated with grasping different propositions *p*, *q*, etc. under *A* have in common. What interests me is the phenomenology of the basic attitude of intellection, considered at this generic level.

3.4. *The phenomenal model: first pass*

Putting together the notion of a phenomenal concept, and the notion of a generic phenomenology of intellection, I would like to suggest a new way in which we might understand the concept of self.²¹

Phenomenal concepts, as noted earlier, are partly individuated by their distinctive phenomenal basis. Now, suppose that I am right in saying that there is a basic experience of intellection present in any conscious thinking – as part of a phenomenal complex in the case of a judgment that *p* or a suspicion that *p* or a supposition that *p*, or on its own in the case where we merely entertain the proposition that *p*. If this experience exists, then it is available as a potential phenomenal basis for the formation of a phenomenal concept in its own right, just like any other phenomenal quality. It is plausible that this concept does indeed exist, and I would like to contemplate the possibility, at a first pass, that it is the I-concept.

The idea, then, is that the phenomenology of intellection constitutes the phenomenal basis of the concept of self²², just like phenomenal reddishness, sourness or itchiness constitute the respective phenomenal bases of the phenomenal concepts *reddish*, *sour* and *itchy*. I think of phenomenal sourness by experiencing what it’s like to taste the sour; similarly, the suggestion here is that I think of myself, under the I-concept, by experiencing what it’s like to be the thinker. In addition, the I-concept works like a phenomenal concept, by exemplification.

This simple way of generalising the phenomenal-concept model to the I-concept seems to capture quite nicely the Thinker Intuition. To think of oneself *as oneself*, remember, is to think of oneself as the thinker of *this very thought*. An awareness of our present mental activity is thus key to the required form of super-reflexivity. There is no better way to achieve this than by exploiting the way that our present thinking is *manifested* to us as it unfolds: namely, through the phenomenology of intellection. What is important here is that, since we are talking about a *phenomenal* form of awareness, the phenomenology of intellection is a *direct* way that our own thinking is given to us; and a way in which it is given to us as *present*.²³ Moreover, since it is the *proprietary* phenomenology of intellection, i.e. what marks it apart from other forms of conscious mentation (vision, audition, proprioception,

affect, etc.), it is a way in which our own thinking is given to us *as a thinking*. Those features can anchor one's conception of oneself in one's occurrent thinking, as such, in a way that the indexical approach, on its own, does not explain. In addition, the phenomenal-concept approach has the consequence that mastering the I-concept is only a matter of attending to the experience of intellection and "taking it up" into the concept, instead of having to feed it into a complex token-reflexive description, as a hybrid indexical model would require. This is a much simpler alternative view, and one that doesn't presuppose further conceptual capacities or an implausibly high level of abstraction.

One objection immediately springs to mind, though. Doesn't this simple extension of the phenomenal-concept model to the I-concept select the wrong type of entity as its referent? The self-concept refers to an *individual*, namely a self, rather than a property. So how can the model of ordinary phenomenal concepts, which are representations of properties, be appropriate in this case? When I think of myself as myself, the objection might go, I think of myself as the *thinker*, not as the *thinking*. Crucial though it may be to think of myself *through* the experience of thinking, it is not the same thing as identifying myself with that experience alone. But this is what I would do on the 'sampling' model of reference distinctive of phenomenal concepts.

Let us call this the "wrong-kind-of-entity objection". Several kinds of responses to the objection are possible. One route, the semantic route, would consist in viewing the I-concept as something like a metonymic concept, enabling us to refer to a whole (the subject) via reference to a part of it (her experience of thinking).²⁴ Another route, the metaphysical route²⁵, would start from the observation that the metaphysics of selves is very far from being a settled matter. That the self isn't a property isn't something we can just take for granted. Several strong accounts exist that do picture the self as a property-like entity: a set of properties (Parfit 1984), a "dimension of experience" or experiential property (Zahavi 2005), or a capacity or set of capacities (Dainton 2008). One could, alternatively, accept that the self is an individual rather than a property, while noting that several kinds of entities count as individuals: things, but also events and processes. Why rule out that the self, a dynamic entity, might turn out to be an individual process like thinking, rather than a *thing* on the model of rocks and houses? I explore these and other possible responses, metaphysical as well as semantic, elsewhere (Guillot ms[b]). Here, however, for the sake of simplicity and space, I'll present just one possible answer which turns on a revision, or rather a generalisation, of the phenomenal model.

3.5. *The phenomenal model: second pass*

The wrong-kind-of-entity objection suggests that simply extending the phenomenal-concept model to the case of the I-concept might not be a straightforward matter. What we need is a model for analysing the I-concept as the concept of the seat or *source* of the present experience of intellection, rather than as a concept of this experience itself. However, I think that the basic inspiration behind the phenomenal-concept model can and should be retained. The intuition that the concept of self is in some way grounded in the phenomenology of intellection is, for the reasons mentioned above, one that seems worth exploring further. The phenomenology of intellection gives the subject a direct grasp of her occurrent thinking, as *present* and as a *thinking*; these are all desirable features in an account of the Thinker Intuition, as we have seen. Could it be, then, that the concept of self is a phenomenally-grounded concept of a slightly different kind than phenomenal concepts *stricto sensu*?

Phenomenal concepts are not the only kind of *phenomenally-grounded concepts*. What I call phenomenally-grounded concepts are those concepts which are partly individuated by

phenomenal experiences. As a consequence, one cannot grasp nor acquire those concepts without having had the relevant kind of experience. Phenomenal concepts are phenomenally-grounded concepts in this sense, since they refer by ‘sampling’ an experience, and thus presuppose an exposure to the experience in question. And they are, as noted above, partly individuated by that experience, as their phenomenal basis.

Another type of phenomenally-grounded concept, however, could provide a better model for the I-concept. We can use phenomenal experiences to refer to those experiences themselves (as we do with phenomenal concepts); but we can also use them to refer to the *cause* or *source* of the experiences. This is what we do with a particular kind of empirical concepts, which I will call *phenomenal-appearance concepts*, and which have recently attracted (under other names) the attention of a number of philosophers, including Dahlgrün (2010), and especially Papineau (2008) and Gertler (2012). Phenomenal-appearance concepts are concepts of empirical objects classified on the basis of how they phenomenally appear to us. Take the first time I saw a panda in the zoo. I registered its peculiar appearance, the pattern of black and white patches, the round cub-like head, the ponderous gait, etc. In doing this, I found myself in a position to acquire a new phenomenal-appearance concept, the concept *panda*. It is a concept for pandas that categorises them based on the typical “panda-look”, the distinctive phenotype of a panda²⁶. The phenomenal impression I received on this first encounter was stored as a kind of “template”; and whenever this template is reactivated, when I perceive or imagine a panda for instance, the new object is filed under the same concept. It is through this phenomenal-appearance concept that I usually refer to pandas, although I could also use other sorts of concepts, for instance a descriptive taxonomic concept, like that of a black-and-white, bamboo-eating subspecies of the *ursidea* family.

Note that the phenomenal character of my impression of pandas plays a constitutive role in the way my phenomenal-appearance concept *panda* works. A given use of the concept will typically involve a panda-like phenomenal impression, and will refer to whatever happens to be causing this impression. As Gertler (2012) puts it, the referent of (what I call) a phenomenal-appearance concept is “the object whose presence and properties causally contribute (in an appropriate way) to the relevant aspect of how things seem to the subject”, i.e. to the activation of the phenomenal template that was set up on the first encounter with the object. As Gertler notes, the concept thus represents its referent in much the same way that a photograph represents its subject: through the impression left on the photographic paper, we refer to the real object which happens to be the causal source of the impression²⁷ (cf Evans 1982, Ch. 3.4).

This mode of reference means that phenomenal-appearance concepts belong to the family of phenomenally-grounded concepts, because the object targeted by a given use of such a concept is determined *via* a relation to a phenomenal impression (i.e., it is selected as the cause or origin of that impression). As a result, the reference-rule that individuates the concept cannot be spelled out without making reference to this piece of phenomenology. In addition, the subject cannot grasp or acquire such a concept (for instance the concept *panda*) without having *had* the relevant phenomenal impression. So phenomenal-appearance concepts, like phenomenal concepts, are partly individuated by a certain experience (e.g. the panda-like experience for the concept *panda*): they, too, have a distinctive *phenomenal basis*.

This suggests a possible way to understand the I-concept. I propose, at a second pass, that the concept of self is a phenomenal-appearance concept whose distinctive phenomenal basis is the phenomenology of intellection. When I use the concept of self, I use the experience in which my present thinking is manifested to refer, not to that experience itself, but to the *source* of that experience: namely myself, the thinker.²⁸

It is important to note that this generalised phenomenal model is in no way a descriptive model. When I use a phenomenal-appearance concept, I refer to *whatever is the source of this*, where “this” is the experience that constitutes the phenomenal basis of the concept. But I do not represent the italicised relation itself; I represent *through* that relation. The photographic model mentioned earlier makes this clear. A photographic portrait enables me to mentally refer to the person who posed for the portrait and caused the appearance fixed on the paper, without requiring that I think descriptively of them as “the person who caused this”; intuitively, the person is simply *presented* to me through the photographic appearance. The same is true of the phenomenal-appearance concept *panda*: causing the relevant panda-like experience in me is a way for the referent, an actual panda, to be presented to me. And so it is, if I am correct, for the concept of self. I use the phenomenology of intellection attached to my present thought to think of *the person who is the source of this thought*, but this doesn’t involve grasping the italicised relation in the form of a description; in being receptive to the phenomenology, I am *presented* with the person in question.

This, then, is the new approach to the concept of self I want to put on the table. The I-concept is a kind of phenomenally-grounded concept, more specifically a phenomenal-appearance concept, whose distinctive phenomenal basis is the phenomenology of intellection.

4. Virtues of the phenomenal model

One clear virtue of the revised phenomenal model is that it accounts well for the Thinker Intuition. If my approach is correct, then in using the I-concept, I can’t fail to be thinking of *the thinker of this very thought*, since the concept is designed to refer to the source of the present thinking. Furthermore, it does so through the direct phenomenal manifestation of that thinking. This enables the phenomenal model to explain The Thinker Intuition more fully, more economically and more plausibly than the indexical model does.

More fully, because the phenomenal model provides a good answer to a central question left open by the indexical approach. To think of oneself *as oneself*, as I remarked above, it is not enough to think of “the thinker”; one needs to think of the thinker of a very specific thought, namely the *present* thought. To think of thinking in general, or of any random thought, obviously doesn’t suffice to secure the required super-reflexivity. But what is more, to think of the thinker of *what happens to be* the present thought is also insufficient: this would merely provide simple reflexivity if one didn’t realise that the thought at issue is the very thought now unfolding. Super-reflexivity obtains only when one thinks of the thinker of the present thought, apprehended *as the present thought*. Only then is the identity between the object and the subject of the relation of representation reflected in the representation itself.

The phenomenal approach explains how, in using the I-concept, we latch onto the correct thought and recognise it as such. On this approach, a subject’s self-conception is anchored in the phenomenology of intellection, through which her thinking is manifested to her as it unfolds. Phenomenology in general has a “presentational” dimension²⁹: it typically makes us aware of what is present, here and now, and it makes us aware of it *as present*: it *presents* us with it. This is most obvious in the case of perceptual phenomenology. When I notice a sweet smell as of roses, it is usually a way to become aware of the roses, or rose-scented objects, that are here at hand; more importantly, the olfactory impression is, so to speak, “present before my consciousness” in a way that also makes me aware of those objects *as present*. The same goes for any kind of phenomenology, and the phenomenology of intellection in particular is a way for us to be aware of our current thinking as the *present* thinking. This gives a full account of super-reflexivity where the indexical model, *per se*,

arguably falls short. As we have seen in Section 2, to say that the meaning of the I-concept is captured by the token-reflexive description “the thinker of this very thought” is not yet to explain in what way reference is made to the thought in question in the description, such that the subject would recognise it as the present thought.

The phenomenal-model account of the Thinker Intuition is also likely to be more economical than the indexical account. As I argued in Section 2, a proponent of the indexical model may want to appeal to the phenomenology of conscious thoughts to answer the above question (how the present thought is made reference to in the description “the thinker of this very thought”). This means adopting a hybrid account. The account I propose is simpler: it starts from the phenomenology that the indexical model might ultimately need to invoke in any case, and builds a model for the I-concept from there, tapping the intuition that the concept’s anchoring into the characteristic experience of conscious thinking is essential to the way it works, and fleshing out the intuition by using the model of phenomenally-grounded concepts. Once the importance of the phenomenology is recognised, it is not clear whether the descriptive apparatus added by the indexical approach casts any further light. Why not, then, adopt the simpler approach I favour?

A further reason to do so is that it provides an account of the Thinker Intuition that has greater *prima facie* psychological plausibility than the indexical model. This is because the phenomenal model is a much less intellectualist option. In order to master the I-concept, the indexical model requires that the subject be able to grasp a complex *description*; on the phenomenal model, she only needs to be receptive to an *experience*. Phenomenal-appearance concepts, to which my view likens the I-concepts, are a relatively basic piece of cognitive equipment, and require little capacity for abstraction. Categorising objects on the basis of the way they phenomenally appear to us is plausibly one of the simpler forms of conceptualisation, one we can practice while still immersed in experience. To refer to herself through a concept of this kind, the subject must be responsive to a certain impression; but she does not need to possess further concepts, *a fortiori* highly abstract concepts such as that of a “thought” or of a “thinker”, as she does on the indexical model. Of course, it might turn out that the mastery of the I-concept really is a competence of the level of sophistication implied by the indexical model. But the fact that there is at least greater *prima facie* psychological plausibility in the competing approach is a reason to prefer it, unless and until proponents of the indexical approach give arguments supporting a more cognitively demanding mechanism.

A further virtue of the phenomenal model is that it accounts for the epistemic exclusivity attaching to the concept of self³⁰. If my capacity to think of myself *as myself* is rooted in a particular aspect of my phenomenology, it will be accessible to me alone, as are phenomenal states in general. Someone else, in order to think about me, will need to employ a different kind of representation. If they use the phenomenology of intellection that they have access to – their own – as the basis for deploying a phenomenal-appearance concept, this will result in their thinking about the source of *this* experience of intellection, namely themselves, rather than me. Thus, anyone (who has the required conceptual abilities) can use the I-concept to think of themselves, and anyone can think of me, NN; but only I can use the I-concept to think of NN. These predictions of the phenomenal model coincide with the intuition of exclusivity.

The phenomenal model may additionally cast some light on the characteristic referential behaviour of the I-concept. This is something I explore in detail elsewhere (Guillot ms[a]), so I will only mention a few points very briefly here, to indicate the way in which the account proceeds.

First, construing the I-concept as a phenomenally-grounded concept correctly predicts the phenomenon of “guaranteed reference” Anscombe (1975) first described. If, in using the I-concept, I refer to the source of the present thinking through the phenomenal manifestation of this thinking, then it is impossible for a use of the concept not to secure a reference: where there is thinking, there is a source to that thinking.

Second, it has long been observed that, at least in its core uses³¹, the I-concept cannot fail to fasten onto the correct object: it is, in those cases, immune not just to empty reference but also to mis-reference. On my approach, this is a consequence of the way the referent is selected. Phenomenal-appearance concepts are designed to target whatever object happens to be the causal origin of a certain phenomenal appearance. As pointed out earlier in the discussion of the photograph analogy, there is no need for an independent epistemic route to that object, other than the phenomenal appearance itself, since this appearance is a way for the object to be manifested. This removes a possible source of error, and explains why the I-concept cannot latch onto something that is not in fact the origin of the current thinking. And whatever system produces my thinking, intuitively, is *me*. (See Dainton 2008, Peacocke 2014). So, at least in its core uses, the I-concept is guaranteed to refer to the right object.³²

This last point gives me an opportunity to highlight both a connection and a degree of contrast between my proposal and some neo-Fregean approaches, in particular the influential view of Evans (1982, Chapter 7).³³ Evans points out that a subject has at her disposal dedicated “information channels” through which she can learn about herself specifically: proprioception, kinaesthesia, pain detection, and all the other ways of gaining information that form part of inner perception. To think of oneself *as oneself*, according to Evans, is to refer to the object one accesses via those dedicated information channels, and to do so in such a way that the thought is *informed* by those channels, sensitive to the information they yield. One point to note is that the only person I can get to know about by exploiting those dedicated epistemic routes happens to be *me* (kinaesthesia, for instance, only informs me about *my* movements). Thus, on Evans’s approach, there is no risk of mis-reference when one uses the I-concept so analysed.

Now, my view belongs to a broadly Evansian tradition insofar as it also treats the I-concept as grounded in an experience of the self. However, the central role that my proposal ascribes to the phenomenological dimension arguably gives it an advantage when it comes to accounting for immunity to mis-reference. Evans’s explanation is couched in functionalist terms: the I-concept tracks the inputs entering a specific circuitry (the mechanisms of proprioception, kinesthesia, etc.). As a matter of cognitive design, these inputs always come from the same person: me. Always, that is, if things are working properly. But deviant causal chains are not in principle ruled out: it *could* be that the systems through which I usually detect my own movement, posture, balance, or temperature (etc.) gets rerouted so as to take inputs from another person’s body.³⁴ If that happened, my self-conception would track a person other than me, giving rise to a form of mis-reference. This means that the form of immunity to mis-reference secured by Evans’s view is a conditional one, dependent on an assumption of normality.³⁵ Intuitively, however, the I-concept (in its core uses) is *absolutely* immune to mis-reference: it doesn’t make sense (other than metaphorically) to say that I could be thinking of another person *as myself*, as long, at least, as the object is only referred to through the I-concept³⁶. My approach may be better equipped to account for his stronger guarantee. On the phenomenal model, the I-concept refers to the causal source of the experience of intellection. But whatever that source is, *that* must be me: it seems something close to an *a priori* truth that only I can have my experiences, so whatever is the seat of my experiences is me, without any room left for a substitution of identities.

5. Some objections

The discussion of the last section, of course, does not cover all the features that make the concept of self puzzling and interesting – very far from it. Such is not the aim here, however; the goal of this essay is only to show that the phenomenal model offers a viable and attractive new approach to the I-concept, which offers a convincing account of its central feature, the super-reflexivity captured by the Thinker Intuition. This, I think, is sufficient reason to believe that the model is worth exploring further.

Before concluding, I would like to consider a few objections. Let me first put aside some potential misunderstandings. It will perhaps be objected that my approach is still fairly intellectualist, in contending that we think about ourselves, *as ourselves*, essentially as *thinking* subjects. Is it not more essential to the sort of entity we are that we enjoy free agency, say, or that we are capable of emotions and of recognising values, or that we belong to a certain animal species (etc.)?

This sort of objection, I think, would be misguided. Please note that I make no metaphysical claim about the nature of selves here; the view only concerns the I-concept as a *representation*, and the mechanism through which it enables us to self-refer. I stressed at the very start the difference between the philosophical concept of self that is used by theorists to reflect on the nature of selfhood, on the one hand, and the ordinary, non-theoretical self-concept you and I use in our daily lives to pick out a specific *individual* – ourselves – among all others. There is no particular reason to expect that this device used for pinpointing someone reveals very much about the defining metaphysical traits of selves in general.

Moreover, that we use the experience attached to our present thinking to self-refer does not prevent us from thinking that many things are true of us, other than that we think. The concept of self is just a way to mark oneself out, and to have a repository to collate information about oneself; but what this information consists in is an independent matter. The I-concept, however it manages to refer, can be used to *predicate* a lot of things of oneself – to think of oneself, for instance, as an agent, a subject of emotions, a living animal, a faithful friend, a worried father, someone who is late filing their tax-report, etc. To view the I-concept as grounded in the phenomenology of intellection is not to say that our quality as thinkers should necessarily take up much of the scope of our self-conception.

Finally, the concept of self is not the only way in which we represent ourselves. There are many non-conceptual ways in which we do so, for instance through basic bodily awareness, *via* the distinctive experiences of proprioception, kinaesthesia, pain sensation, etc. What this essay attempts to do is not to capture all the ways, or maybe even the most fundamental ways, in which we apprehend ourselves as ourselves. It has the much more limited aim of capturing the concept that enables us to represent ourselves *conceptually*.

Still, one might not be content with my observation above that, however my I-concept refers, nothing stops me from using it to predicate any number of things of myself, including physical, biological or social attributes, in addition to intellectual ones. Exactly *how*, it might be asked,³⁷ do subjects manage to make such rich and heterogeneous predications with a concept that corresponds to such a narrow way to think of oneself? If the I-concept builds on a way of being presented to oneself that is as minimal as I claim – if it is built, that is, on one's manifestation to oneself as the thinker of the present thought – does it give us the necessary resources to judge, e.g., “I am left-handed” or “I have a sore throat” or “I have just won the lottery” or even “I am NN”?

It is important to note that the question arises for other accounts of the I-concept, including the indexical model. Even the Evansian approach, which grounds the I-concept in a

broad range of dedicated ways of knowing about oneself, needs to be supplemented to explain how we come to self-attribute features of which we can learn only through “third-person”, as opposed to self-specific, ways of acquiring information (such as reading in the newspaper that NN has won the lottery).

Now, here is one possible way in which this kind of predication might work on the phenomenal model. It is helpful here to follow Papineau (2008) in thinking of phenomenal concepts (and, I would add, phenomenally-grounded concepts more generally) not just as representations whose typical uses exploit a certain phenomenal template. One can also consider the body of information which comes to be stored under a phenomenal concept over the course of its use. As a repository for this body of data, the concept functions as a sort of “file”, or dossier.³⁸ For instance, the file attached with my phenomenal-appearance concept “panda” contains various attributes such as “is black and white”, “eats bamboo”, “sleeps for most of the day”, “climbs trees”, etc. So let us suppose that the I-concept, too, is associated with a file of its own³⁹.

But, as noted above, other types of representations share their referent with the I-concept: various cognitive systems, such as kinaesthesia, proprioception, etc., have the function to track information about the same individual, namely me. It is plausible that the representations of the subject used by each of those systems are also associated with specific repositories of information, or files. Recanati (2012, ch. 4) argues that different files about the same object can come to be “linked” when the subject realises that they co-refer. According to him, this is just what happens when we make a judgment of identity, such as “Cicero is Tully”. As a result of the operation of “linking”, information can flow freely between the two original files. This may be just what happens between the various specialised files in which a subject collects information about herself. In the early stages of a subject’s cognitive development, she comes to learn that the individual named NN is none other than the individual who suffers growth pains, who has bad dreams, who is hungry, who runs a temperature, who is presently thinking, etc. In this way, physical, biological and social information can get filed under the I-concept, even if it originates from a different file. The I-concept, while grounded in the phenomenology of intellection, is thus apt to serve as a repository of data from heterogeneous sources. The details of how we learn to coordinate the relevant files – how we become aware, for instance, that our hearing self and our moving self and our thinking self are one and the same individual – is an empirical question, not unlike that of how multimodal binding happens in perception, and it falls to developmental psychologists to address it. I hope that the suggestion of the general form such an explanation might take in a mental-file framework suffices, at any rate, to alleviate the worry regarding the use of the I-concept in predications of non-intellectual attributes.

A different objection could be raised on the grounds, not that the phenomenal model is too intellectualist, but that it makes essential use of intellectual *phenomenology*. If thinking of oneself under the I-concept involves an experience attaching to one’s conscious thoughts, does that imply that one cannot think of oneself as oneself *unconsciously*? Although several kinds of replies are possible here, the one I find most convincing consists in distinguishing between different uses that phenomenally-grounded concepts can be put to.

Let me note first that the objection echoes one that is often raised against accounts of phenomenal concepts as grounded in conscious experiences. If the phenomenal concept “reddish”, say, is partly individuated by the experience of reddishness, does it mean that I cannot use it when unconscious? A standard response consists in saying that, while the acquisition of a phenomenal concept requires an actual experience of reddishness, the concept is retained beyond the duration of the experience, and apt to evolve into a stable successor concept that can be re-used even in the absence of a similar experience. This is what

Chalmers (2003) calls “standing phenomenal concepts” (to distinguish them from his “direct phenomenal concepts”), or what Papineau (2008, Section 3.5) calls “phenomenally-derived concepts”. According to Papineau, as noted above, the phenomenal template with which the concept originates (what I call its phenomenal basis) comes to be associated with a “file” in which the information falling under the concept is stored. Once in existence, the file itself can be used to form further attitudes, whether or not the associated experience is presently active. Even in its “derivative” uses, though, the concept still counts as a phenomenally-grounded concept, since the file is partly individuated by the originating experience, insofar as its historical-causal conditions of emergence plays a decisive role in making it the concept that it is. A similar reasoning can be applied, beyond the case of pure phenomenal concepts, to the case of phenomenally-grounded concepts in general. The details of how “derivative” or “standing” uses of such concepts are related to their “core” or “direct” uses can be fleshed out in different ways, and much more needs to be said on the issue; this would, however, be a task for a separate essay. What matters for now is the general idea that a concept being phenomenally-grounded, and in particular requiring for its acquisition an exposure to the experience that constitutes its phenomenal basis, doesn’t imply that each and every subsequent use of the concept depends on a fresh instantiation of the experience. Unconscious as well as conscious tokenings of the concept of self are thus in principle compatible with the phenomenal model.

A related issue concerns those systems that are unconscious not just some of the time, as we are, but all the time. Shouldn’t it be possible for information-processing structures which are devoid of any phenomenology, such as machines and zombies, to self-refer? The reply made to the previous point cannot be extended to this one, since those systems have no conscious experiences that a concept could get grounded in to start with, thus excluding *a fortiori* any possibility of “derivative” uses of such concepts.

My response here is to point out the pluralism in the domain of reflexive representations. While a machine can certainly self-refer, there is no compelling reason to believe that the representation mobilised to that effect is identical to our concept of self. Many things self-refer – for instance the bottle of potion in *Alice in Wonderland*, with its label “Drink me” – without warranting the attribution of the same I-concept that we possess. It seems quite plausible that a device of self-reference in a machine, for instance, could secure self-reference, and even necessary self-reference, while still falling short of super-reflexivity. A symbol in a computer can certainly stand for the computer itself (the little hard-drive icon on the computer’s desktop is one example), thus equipping it with simple reflexivity; further, the machine can be set up so that the symbol only tracks changes in the machine itself, thus making the reflexivity necessary rather than (as in the case of Oedipus) accidental. Yet does this amount to the system being aware of the relation of reflexivity itself? Much more would have to be said about the case to evidence a true likeness to what a subject does with her concept of self.

6. Conclusion

I have argued that the super-reflexivity that distinguishes the concept of self is captured by the “Thinker Intuition” – the intuition that in thinking of herself as herself, a subject identifies both *relata* of the relation of representation as one and the same object, thus thinking of herself as *the thinker of this very thought*. As I hope to have shown, this intuition can be successfully accounted for by construing the concept of self as a phenomenally-grounded concept, more specifically a phenomenal-appearance concept, anchored in the phenomenology that makes one’s present thinking manifest as such: the phenomenology of intellection. The account is economical and has psychological plausibility, avoiding in

particular the strong intellectualism of the indexical approach. It also opens up promising new ways of explaining additional features of the I-concept, including guaranteed reference and the immunity to mis-reference. Much more will need be said about those points in future work, and it is also possible, of course, that the details of the approach are not quite right. But I hope to have made at least a convincing case for the need to understand the concept of self as much more deeply rooted in the phenomenology of intellection, and phenomenology generally, than has been acknowledged so far. No doubt there is much room for improvement of the model and further exploration of its implications. The aim of this essay, though, will have been achieved if the reader agrees that this is indeed a direction worth exploring.

Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement [622127]. It was also supported by DGI, Spanish Government, research project [FFI2013-47948-P]. The article arose from a talk at the conference *Cognitive Experiences* organised by the two editors of this volume at University College Dublin in November 2014. I thank the audience of this talk, especially Marta Jorba and Miguel Ángel Sebastián, for a helpful discussion. I am also grateful to the two editors, and to two anonymous reviewers for insightful comments.

¹ By Castañeda (1966, 1968), Perry (1979), Lewis (1979), Kaplan (1989), Evans (1982), O'Brien (2007), among many others.

² The distinction is due to Perry (2002, p. 209).

³ The distinction between “simply reflexive” and “super-reflexive” thoughts echoes the distinction made by Castañeda between “external” and “internal” reflexivity (1999, p. 256).

⁴ Please note that in talking about a subject who thinks of herself *as* the thinker, I am in no way implying that she thinks of herself under the *description* “*x* is the thinker”. To think of something *as* this or that is to think of it, as Kaplan (2012) puts it, in a “rich” rather than “thin” way; namely, to think of it under a certain mode of presentation rather than to target it as a bare individual. But not all modes of presentation are descriptive, a point first emphasized by Evans (1982) and McDowell (1984). It is now standard to accept what have been called, among others, “*de re* senses” (McDowell 1984), “*de re* modes of presentation” (Bach 1987, Recanati 2012) or “non-descriptive ways of having in mind” (Kaplan 2012) alongside descriptive ones. Section 2 below makes the case that a descriptive reading of the Thinker Intuition is unsatisfactory.

⁵ The intuition is endorsed explicitly by many writers, including Russell (1914), Higginbotham (2003) and Peacocke (2014). It has a long philosophical history, in which Descartes, Locke and Fichte, in particular, loom large. The intuition, however, might be disputed. An alternative suggestion might be that, when one thinks of oneself as oneself, one thinks of oneself as *agent*. This idea is at the core of a distinct tradition, which includes Anscombe (1975), Moran (2001) and O'Brien (2007), among others. Arguably, however, in the act of reflecting on oneself, one's *mental* agency, as the source of the current thinking, is one of the central forms of agency one is made aware of. This suggests some common ground with the kind of view I propose here.

⁶ There may be other, lower-level representations that are essentially *egocentric*, in having the function to stand for the subject, where this doesn't necessarily involve her explicit recognition of that fact. Such may be, for instance, the “body map” a subject uses for monitoring her own body in proprioception. Representations such as these play an important role in our psychological lives. They are, however, distinct from the *concept* of self, which is used in thought as opposed to mere experience, and is defined by super-reflexivity.

⁷ Proponents of different indexical approaches to the I-concept include Zemach (1985), Higginbotham (2003), Howell (2006) and García-Carpintero (2016). It may well be that not all of those authors would accept all the features I attribute to the indexical model in this section. In describing what I call the “indexical model”, I am not aiming for exegetical accuracy, but trying instead to characterise a possible view which accounts for the Thinker Intuition along the model suggested by the word “I” – a view whose core ideas are in fact accepted, explicitly or more often implicitly, by many.

⁸ I won't discuss other existing views of the concept of self, such as the demonstrative models of Evans (1982) and Morgan (2015), the no-reference stance of Anscombe (1975), the relativist approaches of Lewis (1979)

-
- and others, or the mental-file models of Perry (2002) and Recanati (2012, Part VIII). Such a survey would take me too far beyond the goal of this essay, which is to propose the phenomenal model as a possible new way to flesh out to the Thinker Intuition. The approaches listed above are not primarily designed to explain the intuition; the indexical model arguably is, which makes it a natural contrast case here.
- ⁹ The first to have fully articulated a token-reflexive theory of this kind are probably Benveniste (1958) and Reichenbach (1947), who also mentions Russell (1940) as an inspiration. Russell (1914, p. 163 sq.) defends a version of the token-reflexive approach for both the word “I” and the I-concept.
- ¹⁰ This modal objection, first presented by Castañeda (1966, 1968), was further developed by Kaplan (1989) and more recently Predelli (2005).
- ¹¹ For Kaplan, the content of “I” in a given context is in fact a constant function from possible worlds to extensions (the same individual in each world), rather than the individual itself. My omission of this further refinement in Kaplan’s view, however, has no bearing on what follows.
- ¹² An acute discussion of this point is to be found in Nozick (1981, Chap. 1.2).
- ¹³ A closely related but distinct option would consist in claiming that we can latch on to the present thought merely by virtue of the fact that we have a conscious access to it, without committing to the thought also being distinguished by a specific phenomenology. (See Block 1995 on the difference between phenomenal consciousness and access-consciousness). In what follows, I ignore the access-consciousness variant for simplicity’s sake, and because I believe it to be less promising than the phenomenal-consciousness variant. One reason is that the access-consciousness variant does not obviously have the resources to explain the epistemic exclusivity of the I-concept (see footnote 15 below). Whatever the chances that an appeal to access-consciousness as a complement to the indexicalist view might be viable, however, it would in any case also result in a hybrid view. My point in Section 4 below that the kind of approach I favour is more economical would thus equally apply to this variant. I am grateful to an anonymous reviewer for inviting me to clarify this.
- ¹⁴ A token-reflexive description, as such, might be able to secure *simple reflexivity* if we suppose that the thought represents itself thanks to a syntactic item whose function just is to stand for the larger representation it is a part of (see García-Carpintero 2016 for a sophisticated account of this kind). Arguably, however, this still wouldn’t amount to *super-reflexivity*. The person one thinks about would extensionally be guaranteed to be identical to the person thinking; but this would arguably leave unexplained why she is aware of this fact.
- ¹⁵ But see García-Carpintero (2016, 194), who preserves exclusivity within a sophisticated framework which shares some features (including token-reflexivity) with the indexical model.
- ¹⁶ Why think that an appeal to *phenomenal* consciousness, rather than access-consciousness, is needed to secure exclusivity? Access-consciousness, after all, is sufficient to make a thought available to introspection, thus enabling the subject to refer demonstratively to it as *this* thought. However, this on its own may fall short of the required asymmetry with a third-person access to the relevant thought. It is not obvious that others can’t equally refer demonstratively to my occurrent thoughts, for instance by monitoring my mental activity in real time with a sophisticated brain scanner. A detailed discussion of this point would take me too far afield, but see Recanati (2007), Chapter 25 on related issues. In any case, an appeal to access-consciousness, whatever its chances of success, would also amount to an acknowledgment that a “pure” indexicalist view is insufficient, and we would be owed some details as to how exactly the explanation might go.
- ¹⁷ For instance Loar (1990/7) and Horgan and Kriegel (2007).
- ¹⁸ Similar ideas are defended in Gertler (2001, 2012), Levine (2003), Papineau (2008), among others.
- ¹⁹ For a thorough case for this claim, as well as some historical context, see Kriegel (2015), Chapter 3.
- ²⁰ Kriegel (2015), Chapter 3, refers to it as “the phenomenology of entertaining”.
- ²¹ The specific view I am presenting here is new (to my knowledge), but it belongs to a wider family of approaches according to which the I-concept is rooted in the phenomenal. See e.g. Castañeda (1966, 1968) and Kapitan (2015).
- ²² In Guillot (ms[b]), I explore a variant of the phenomenal model which uses the subjective character of conscious mental states as the phenomenal basis of the I-concept.
- ²³ More on this below in Section 4.
- ²⁴ Kapitan (2015), while not using the framework of phenomenal concepts, develops a sophisticated account which offers the right sort of semantic flexibility.
- ²⁵ I am grateful to Uriah Kriegel for help in distinguishing the two kinds of strategies.
- ²⁶ I do not mean to challenge in any way the standard approach to natural-kind concepts like “gold” or “water”, developed by Kripke and Putnam. What I suggest here is simply that these are not the only kind of concepts we use to categorise the objects we encounter in perception. Plausibly, we rely heavily on the way perceptible objects *appear* to us for some forms of categorisation. I thank Jonathan Way for inviting me to clarify this.
- ²⁷ One might worry that a phenomenal impression is not generally caused by a *unique* object: the causal chain, in a given case, will typically involve any number of objects. Still, there is an intuitive sense in which a photograph of John stands for *John*, rather than for any of the other things that contributed to causing the impression on the photosensitive film (the silver salts, the camera lens, the light beams bouncing off John,

etc.). Other cases give more latitude for deciding what the “appropriate” cause might be. Papineau (2008) thus notes that a given phenomenal appearance (as of a greenfinch now landing on my apple-tree, for instance) can often equally be used to refer to an individual (this particular bird) or to a kind (greenfinches). This flexibility, however, would only be a source of concern if we expected phenomenal-appearance concepts to be fully individuated by their phenomenal bases. There is no reason why this should be the case. Those concepts, as other kinds of concepts, can be individuated by their distinctive reference-rule, i.e. the condition an object has to satisfy to be their referent (cf Peacocke 1992). As I argue in the next paragraph, in the case of phenomenal-appearance concepts, a mention of their phenomenal basis is *necessary* to spell out the relevant reference-rule (this is why they count as phenomenally-grounded concepts); but it is not sufficient. What else is needed is a question for another occasion. The remarks of this subsection hopefully provide an intuitive handle on the notion of a phenomenal-appearance concept; my purpose is merely to argue that the I-concept has features which warrant classifying it as a concept of this kind, not to give a comprehensive description of those concepts, in general.

²⁸ An anonymous referee objected that we might want to identify the source of the experience of intellection as the thinking, rather than the thinker. It is certainly true in this case, as it typically is with experiences, that a number of different entities play a role in the relevant causal chain. As stressed in the previous footnote, however, this just means that a given experience can serve as the phenomenal basis of different phenomenal-appearance concepts, referring to different entities that can count as a cause or origin. There is no reason why the phenomenology of intellection couldn't be used to form a concept of the thinking process, but it is equally available to form a concept of the system that ultimately causes the experience of intellection, namely the subject.

²⁹ See Chudnoff (2011) on what he calls the “presentational” aspect of phenomenology.

³⁰ See Section 2 above.

³¹ See Coliva (2003) for a strong argument that some complex uses are not immune to mis-reference.

³² As I argue in Guillot (ms[a]), the phenomenal model has the potential to explain why the concept of self, when used in self-attributions of phenomenal properties, is also immune to what is known as “error through misidentification” (Shoemaker 1968); but showing this would lead me too far afield. The previous points, however, should suffice to make the case that the phenomenal model is worth exploring further.

³³ I am grateful to an anonymous reviewer for inviting me to touch upon this filiation, which would deserve a much more detailed discussion, left for future work.

³⁴ See Dennett (1978), Chapter 17 for a detailed scenario of this kind.

³⁵ See Shoemaker (1968) on the related distinction between “*de facto*” and “absolute” forms of another kind of immunity, namely immunity to error through misidentification.

³⁶ Coliva (2003) makes a convincing case that mis-reference is in fact possible with the I-concept. Her examples, however, concern more complex cases, where the individual referred to is simultaneously targeted under the I-concept *and* under further representations (for instance a perception of the person's reflection in a mirror). Thinking of someone simply as “I”, by contrast, absolutely rules out the possibility of mis-reference.

³⁷ I thank an anonymous referee for raising this issue, and for helpful suggestions as to how to tackle it.

³⁸ The idea of mental files or dossiers has been extensively used since the 1960's, by Grice, Strawson, Kahneman, Heim, Perry, and Recanati among many others, to model various other kinds of concepts, including those involved in anaphoric reference, memory, perceptual judgments, and most recently singular or “*de re*” thoughts.

³⁹ Recanati (2012), Ch. 5 and 7, includes a detailed discussion of the concept of self within a mental-file framework.

Notes on contributor

Marie Guillot is a Marie Curie Fellow in the Department of Philosophy at University College London (UK). She has published articles on self-knowledge, *de se* attitudes, semantic relativism, Evans, and Frege, and is co-editing with Manuel García-Carpintero a volume of essays on the subjective character of experience.

References

Anscombe, Gertrude E. M. 1975. “The First Person.” In *Mind and*

-
- Language*, edited by S.D. Guttenplan, 45-65. Oxford: Clarendon Press.
- Bach, Kent. 1987. *Thought and Reference*. Oxford University Press.
- Bayne, Tim and Michelle Montague, eds. 2011. *Cognitive Phenomenology*. New York: Oxford University Press.
- Benveniste, Emile. 1958. "De la subjectivité dans le langage." In *Journal de psychologie normale et pathologique*, July-September 1958, 55(3): 257-265.
- Block, Ned. 1995. "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences* 18(2): 227-287.
- Castañeda, Hector-Neri. (1968) 1999. "On the Phenomeno-Logic of the I." Reprint in H.-N. Castañeda, *The Phenomeno-Logic of the I*, edited by James G. Hart and Tomis Kapitan. Bloomington: Indiana University Press.
- Castañeda, Hector-Neri. (1966) 1999. "'He': A Study in the Logic of Self-Consciousness". Reprinted in Castañeda (1999).
- Chalmers, David. 2003. "The Content and Epistemology of Phenomenal Belief". In Q. Smith, and A. Jokic, eds., *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Chudnoff, E. 2015. *Cognitive Phenomenology*. New York: Routledge.
- Chudnoff, E. 2011. "What Intuitions Are Like." *Philosophy and Phenomenological Research*. 82: 625–654.
- Coliva, Annalisa. 2003. "The First Person: Error Through Misidentification, the Split between Speaker's and Semantic Reference, and the Real Guarantee." *The Journal of Philosophy*, Vol. 100, No. 8: 416-31.
- Dahlgrün, M. 2010. "The Notion of a Recognitional Concept and Other Confusions." *Philosophical Studies*, 150 (1):139-160.
- Dainton, Barry. 2008. *The Phenomenal Self*. Oxford: Oxford University Press.
- Evans, G. 1982. *The Varieties of Reference*. Posthumous edition by J. McDowell. Oxford: Clarendon Press.
- Dennett, Daniel. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.
- Frege, Gottlob. 1918. "Der Gedanke. Eine logische Untersuchung." In *Beiträge zur Philosophie des Deutschen Idealismus* (1): 58-77.
- García-Carpintero, Manuel. 2016. "Token-Reflexive Presuppositions and the *De Se*." In *About Oneself*, edited by M. García-Carpintero and S. Torre, 179-199. Oxford: Oxford University Press.
- Gertler, Brie. 2012. "Renewed Acquaintance." In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar, 89-123. Oxford: Oxford University Press.
- Gertler, Brie. 2001. "Introspecting Phenomenal States". *Philosophy and Phenomenological Research* 63: 305-28.
- Guillot, Marie. ms(a). "On Coliva on Three Properties of 'I'".
- Guillot, Marie. ms(b). "Is the I-Concept a Phenomenal Concept?"
- Higginbotham, Jim. 2003. "Remembering, Imagining, and the First Person." In *Epistemology of Language*, edited by A. Barber, 496-533. Oxford: Oxford University Press.
- Horgan, T. and U. Kriegel. 2007. "Phenomenal Epistemology: What is Consciousness that We May Know It So Well?" *Philosophical Issues*, 17, *The Metaphysics of Epistemology*.
- Howell, R.J. 2006. "Self-Knowledge and Self-Reference." *Philosophy and Phenomenological Research*. Vol. 72, Issue 1: 44–70.
- Hume, David. 1739. *A Treatise of Human Nature*.

-
- Kapitan, Tomis. 2015. "Is Subjectivity First-Personal?" In *Pre-reflective Consciousness: Sartre and Contemporary Philosophy of Mind*, edited by S. Miguens, G. Preyer, and C. Bravo. New York: Routledge.
- Kaplan, David. 2012. "An Idea of Donnellan". In *Having In Mind: The Philosophy of Keith Donnellan*, edited by Joseph Almog and Paolo Leonardi, 122-175. Oxford: Oxford University Press.
- Kaplan, David. 1989. "Demonstratives", in *Themes from Kaplan*, edited by J. Almog, J. Perry and H. Wettstein, 481–563. Oxford: Oxford University Press.
- Kriegel, Uriah. 2015. *The Varieties of Consciousness*. Oxford: Oxford University Press.
- Levine, Joseph. 2003. "Knowing What It's Like." In *Privileged Access*, edited by Brie Gertler. Aldershot: Ashgate Publishing Company.
- Levine, Joseph. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *The Philosophical Review* 88: 513-43.
- Loar, Brian. 1990. "Phenomenal States". In *Philosophical Perspectives: Action Theory and Philosophy of Mind*, edited by J. Tomberlin, J., (4): 81-108.
- McDowell, J. 1984. "*De Re* Senses". In *Frege: Tradition and Influence*, edited by C. Wright. Oxford: Basil Blackwell.
- Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton NJ: Princeton University Press.
- Morgan, Daniel. 2015. 'The Demonstrative Model of First-Person Thought', *Philosophical Studies* 172 (2015): 1795-1811.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- O'Brien, Lucy. 2007. *Self-Knowing Agents*. Oxford: Clarendon Press.
- Papineau, David. 2008. "Phenomenal and Perceptual Concepts." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by T. Alter and S. Walter. Oxford: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Peacocke, Christopher. 2014. *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
- Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- Perry, John. 2002. *Identity, Personal Identity, and the Self*. Indianapolis: Hackett.
- Perry, John. 2001. *Reference and Reflexivity*. Stanford: CSLI Publications.
- Perry, John. 1979. "The Problem of the Essential Indexical." *Noûs* 13: 3-21.
- Pitt, D. 2011. "Introspection, Phenomenality, and the Availability of Intentional Content." In Bayne and Montague, eds. 2011.
- Pitt, D. 2004. "The Phenomenology of Cognition, or, What Is It Like to Think that P?" *Philosophy and Phenomenological Research*, 69 (1):1-36.
- Predelli, Stefano. 2005. *Contexts: Meaning, Truth, and the Use of Language*. Oxford: Oxford University Press.
- Recanati, François. 2012. *Mental Files*. Oxford: Oxford University Press.
- Recanati, François. 2007. *Perspectival Thought. A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: The Free Press, Collier-Macmillan.
- Russell, Bertrand. 1914. "On the Nature of Acquaintance." *The Monist*, 24 (April 1914): 1–16, 161–87, 435–53.
- Russell, Bertrand. 1940. "Egocentric Particulars." In *An Inquiry into Meaning and Truth*. New York: Norton.

-
- Shoemaker, Sydney. 1968. "Self-Reference and Self-Awareness." *Journal of Philosophy* 65 (19), Oct. 1968: 555-67.
- Tye, Michael. 2003. "A Theory of Phenomenal Concepts". In *Minds and Persons*, edited by A. O'Hear. Cambridge: Cambridge University Press.
- Zahavi, Dan. 2005. *Subjectivity and Selfhood*. Cambridge, MA: MIT Press.
- Zemach, Edward. 1985. "De Se and Descartes: A New Semantics for Indexicals." *Noûs*, Vol. 19, n° 2 (June 1985): 181-204.