EuJAP | Vol. 9 | No. 2 | 2013
UDK 130.1
165.1/.2

EXPRESSION AND TRANSPARENCY IN CONTEMPORARY WORK ON SELF-KNOWLEDGE
Hatzimoysis, A., ed. 2011. *Self-Knowledge.* Oxford: Oxford University Press.

ÁNGEL GARCÍA RODRÍGUEZ
University of Murcia

ABSTRACT

A central feature in contemporary discussions of self-knowledge concerns the epistemic status of mental self-ascriptions, such as "I have toothache" or "I believe that p". The overall project of such discussions is to provide an account of the special status of mental self-ascriptions vis-à-vis other knowledge-claims, including ascriptions of mental states to others. In this respect, two approaches have gained currency in contemporary philosophy. Some authors have focused on the notion of expression, stressing that self-ascriptions are expressions of one's mental life; whereas others have singled out the notion of transparency as crucial, since self-knowledge involves self-beliefs, and the latter are transparent to one's available evidence. This critical notice explores one instance of each approach, in order to assess their prospects of success. It concludes that both approaches fail to articulate an account that satisfactorily accommodates the first-person component in the notion of self-knowledge.

Keywords: self-knowledge, avowal, expression, transparency, belief

1.

It is an aspect of everyday life that we claim to know a great deal of facts about the world at large, including facts about other people and ourselves. Thus, we claim to know other people's beliefs or intentions, as well as their desires or emotions; and we often make similar claims about our own mental life. *Self-knowledge* (hereafter, *SK*) is a book about the philosophy of this aspect of everyday life – namely, the nature of our relationship to our own beliefs, desires and other mental states, in virtue of which we claim to know our own minds.

The nature of self-knowledge is a classic topic of philosophy, discussed by many of the great philosophers of the past. But closer to the contents of *SK*, it is also a topic that has received widespread attention in contemporary philosophy, standing as it does at the cross-roads of crucial issues in epistemology, philosophy of mind and philosophy of language. Against this background, *SK* provides a snapshot of the work of philosophers today, covering a vast amount of ground in the process. As a result, it contains papers ranging from the semantics of self-reference and the first-person pronoun (J.L. Bermúdez' "Self-knowledge and the sense of 'I'", and R.M. Sainsbury's "English speakers should use 'I' to refer to themselves") to the relationship between externalism and self-knowledge (Gary Ebbs' "Anti-individualism, self-knowledge and epistemic possibility" and Crispin Wright's "McKinsey one more time"); or from scholarly accounts of Frege's views of the mind (Charles Travis' "Viewing the inner") to the nature of the rational self-control of

beliefs and intentions (David Owens' "Deliberation and the first person"). But above all it contains papers on the epistemology of mental states (such as Ram Neta's "The nature and reach of privileged access", Sven Bernecker's "Representationalism, first-person authority, and second-order knowledge", or André Gallois' "Deflationary self-knowledge"), including papers on two of the most discussed conceptions of self-knowledge nowadays – namely, neo-expressivism (Anthony Brueckner's critical account in "Neo-expressivism", and Dorit Bar-On's reply in "Neo-expressivism: avowals' security and privileged self-knowledge"), and transparency-based accounts (in particular, Alex Byrne's pro-transparency "Knowing that I am thinking", and Brie Gertler's critique in "Self-knowledge and the transparency of belief").

This breadth of scope meets the editor's self-confessed objectives for the book, as stated in the introduction – that is, "to deepen our understanding of self-knowledge", to further "philosophical progress", and "to assess the value of some classic moves in the debate over the epistemic status of self-ascriptions" (2).[1] The aim of this critical notice is more modest, as it focuses exclusively on one part of the ground covered by the book; namely, the epistemic status of self-ascriptions. To this effect, only the most directly relevant contributions made by some, but not all, of the papers included in *SK* will be reviewed and critically examined.

2.

As suggested previously, an important aspect in contemporary work on self-knowledge concerns the nature of self-ascriptions; but what are the self-ascriptions that philosophers focus on? One may ascribe to oneself bodily conditions, such as having gained some weight recently or having high cholesterol levels, but these are not the type of self-ascriptions that give rise to the philosophical problem of self-knowledge. One obtains knowledge of one's weight by looking at the pointer on the bathroom scales, just as one learns about one's son's weight by checking the scales in a similar way; so there is nothing special about this kind of knowledge of oneself. Things are different, though, regarding mental states. One may learn about one's son's beliefs, desires or intentions by talking to him (if he is forthcoming), or by observing and adequately "reading" his facial and bodily gestures (in general, his non-linguistic behaviour), in appropriate circumstances. However, one does not usually (need to) check one's own gestures or one's own linguistic and non-linguistic behaviour in order to know what is on one's mind. It is this double asymmetry that lies at the heart of contemporary philosophers' interest in self-knowledge: one, the asymmetry between knowledge of one's own mind versus someone else's; and two, the asymmetry between knowledge of one's mind versus knowledge of one's body.[2]

This admits of further clarification. I may on occasion learn about my own mental states much in the same way as I learn about someone else's. For instance, I may watch an old home video and, through observation of my bodily gestures and behaviour, understand what was on my mind at the time (something I might have forgotten since, or perhaps something I was unclear about then). Alternatively, I might discover what is on my mind now by

[1] Unless otherwise stated, page numbers in brackets refer to *SK*.

[2] Knowledge of one's body through proprioception or kinaesthesia differs from knowledge of one's weight obtained by looking at some scales. To begin with, the former, unlike the latter, is knowledge available only to oneself. For this reason, materialists like David Armstrong (1968) have used it as the model for introspective knowledge of one's own mind (more on this below). Gareth Evans (1982) has also found important similarities between proprioception and knowledge of one's own mind: neither is based on the identification of one subject amongst others. Therefore, the asymmetry between knowledge of one's mind versus one's body must not be construed as involving, on the side of the body, proprioceptive or kinaesthetic knowledge.

trusting the opinion of my worthy friends, or an expert therapist, much as I might rely on others to tell me what is on my teenage son's mind, when I cannot fathom it for myself or persuade him to tell me. So, the asymmetry that lies at the heart of philosophers' interest in self-knowledge does not concern knowledge of one's own mind versus someone else's, or knowledge of one's mind versus one's body, *tout court*. Rather, what philosophers are interested in is knowledge of one's mind not obtained through testimony, the observation of one's own behaviour, or other routes usually available to other people. For even if on occasion I rely on such means as testimony or observation, undoubtedly on many other occasions I do not. Thus, at any given time, each of us can give a pretty good account of what is going on in his or her mind without reverting to such means. Following widespread contemporary usage, this subset of mental self-ascriptions will be termed *avowals*. Thus, contemporary philosophical interest in the issue of self-knowledge can be captured in the following question: what accounts for the special status of avowals?

In fact, in order to understand contemporary debates on self-knowledge fully, this question needs elucidation, both regarding the boundaries of what is and is not avowable, and regarding what it is about avowals that stands in need of explanation. As to the first point, the debate focuses on which aspects of one's mind can be avowed: for instance, whether or not one can avow *all* the contents of one's mind? Thus, in *SK* Anthony Brueckner (170) follows Dorit Bar-On (2004, 400) in excluding such perceptual self-reports as "I see a tiger" from the set of avowals; but there is no room here to evaluate this claim. Instead, section 4 below will address a related claim – namely, that one can only avow the current contents of one's mind. As to the second point, the debate focuses on whether the special status of avowals is in the end an epistemic (or cognitive) phenomenon. This is the matter to which our attention turns now.

3.

As mentioned earlier, although it is generally assumed that avowals enjoy a special status, by comparison with bodily self-ascriptions and ascriptions of mental states to others, there is no unanimous opinion as to what this special status amounts to. Thus, it is agreed that we treat avowals in a special way, in so far as normally they are regarded as true, are neither questioned nor doubted, and people are not expected to provide grounds for them. Nonetheless, there is disagreement as to whether these features of avowals are to be explained in different epistemic (or cognitive) terms from those involved in bodily self-ascriptions or ascriptions of mental states to others; or alternatively, whether they are to be explained in non-epistemic (or non-cognitive) terms. Ultimately, the discrepancy concerns whether or not the special status of avowals is itself an epistemic phenomenon; or rather, whether it is *only*, or perhaps *primarily*, an epistemic phenomenon.

On the face of it, the special status of avowals looks like an epistemic phenomenon, for it is glossed in terms of such epistemic, or epistemic-related, notions as truth, lack of doubt or lack of grounds. It is safe to add that this is probably the orthodox view, for part of the Cartesian legacy bequeathed to us includes the idea that the asymmetry between avowals, on the one hand, and bodily self-ascriptions and the ascription of mental states to other people, on the other, is a matter of the different epistemic mechanisms involved. Descartes thought that avowals were the result of the exercise of a *sui generis* epistemic faculty to access the contents of one's own mind infallibly and incorrigibly. It is quite common for contemporary philosophers to let go of infallibility and incorrigibility, but many still accept the idea of an underlying epistemic (or cognitive) mechanism.

Accordingly, those who endorse this modified orthodox view use the label "privileged access" to characterize the special status of avowals. In fact, in his contribution to *SK*, Alex Byrne usefully distinguishes between privileged and peculiar access, as follows. On the one hand, privileged access is the idea that "beliefs about one's mental states acquired through the usual routes are more likely to amount to knowledge than beliefs about others' mental states" (106); or alternatively, are "less prone to error" (109). On the other hand, peculiar access is the idea that one knows about one's own mind "in a way that is available to no one else" (107). Both claims are connected, for in so far as the usual routes to one's own mind involve neither testimony, observation, nor any of the ways available to others, then the ensuing beliefs are less prone to error; hence, the special epistemic status of avowals.[3]

Of course, the question why these two claims should be connected in that way is outstanding. More precisely, why is the fact (if it is one) that access to one's own mind is peculiar (in the sense above) more likely to lead one to knowledge, rather than error, regarding the contents of one's own mind? But regardless of the possible answers (and some will be considered below), there is a clear project here as to what a philosophical account of the special status of avowals must aim at – namely, providing the epistemic basis for the special status of avowals, and the ensuing asymmetries already mentioned. The contributions by Byrne and Gertler in *SK* are directly relevant to this project, and will be further reviewed below.

However, not all the papers in *SK* endorse this project; most notably, Dorit Bar-On's does not. What is characteristic about her approach to the problem of self-knowledge is that, although she starts from the special status of avowals,[4] she establishes a clear distinction between the question about why avowals are given such a specially secured status, and the question about whether avowals are instances of privileged self-knowledge. Therefore, according to Bar-On's approach, the philosophical problem of self-knowledge has two separate parts. The second part, which lines up nicely with the project outlined previously in this section, is an epistemic problem, regarding the basis for a knowledge-claim that is specially secured. The first part, though, is not an epistemic problem, and has no counterpart in the project outlined above. For what interests Bar-On when she considers the reasons for the special security of avowals is not the fact that they are knowledge-claims, together with their epistemic basis. Instead, she brackets the very idea that avowals are knowledge-claims (notwithstanding the fact that epistemic, or epistemic-related, notions are used to characterize the special security of avowals), to address directly the question why we treat avowals in the way we do. In other words, her concern is not an alleged knowledge-claim, but our attitudes towards avowals. Hence, this first part of the problem can be captured in the following question: why do we respond to avowals conferring upon them a distinctive and specially secured status, in contrast to what we do with bodily self-ascriptions and ascriptions of mental states to others?

It is a welcome achievement of *SK* that it allows the reader to appreciate the contrast

[3] Other labels used for the idea of privileged-cum-peculiar access include "introspective", "armchair" or "a priori" knowledge. Often, these labels are intended to convey the idea that, in some sense, access to one's own mind differs from access to the empirical world at large. These labels have featured prominently in contemporary philosophy, both in debates about the compatibility between ordinary self-knowledge and content externalism, and in discussions about a related puzzle – namely, whether one could have non-empirical knowledge of the world, once ordinary self-knowledge and content externalism are assumed. (See the papers by Gary Ebbs and Crispin Wright in *SK*, for more on this aspect of contemporary discussions of self-knowledge, which has been left out of this critical notice.) It is important to note, though, that what is common to all of these labels can be captured in the terms used previously in the main text: namely, access to one's own mind is not usually a matter of testimony, observation, or any of the routes usually available to others.

[4] In fact, she uses the terms "special security", for in Anthony Brueckner's words, avowals are "protected from ordinary epistemic assessment" (171).

between these two different ways of characterizing what it is about avowals that raises contemporary philosophical interest; and to consider their prospects of success will undoubtedly prove enlightening. But before doing so, let us pause to consider how far the boundaries of avowals extend.

4.

When the notion of an avowal was introduced in section 2, avowals were distinguished from evidence-based mental self-ascriptions. Accordingly, I might learn about my past mental states by watching myself in an old home video, or I might ascertain the present contents of my mind through the testimony of a close friend or an expert therapist; but none of the subsequent self-ascriptions will have the distinctive and special status of avowals mentioned above. For, in common with other evidence-based claims about the world at large, in so far as such self-ascriptions rely on observational or other kinds of evidence, queries and doubts about one's grounds are perfectly in order, and truth is not granted by default.

This much is uncontroversial in contemporary philosophy. So is the fact that utterances of "My tooth hurts", "I'd love a cup of tea", or "I'll come at six", in ordinary circumstances, are clear-cut instances of avowals. But how are the boundaries of the set of avowals fixed? What is the criterion to determine whether a mental self-ascription qualifies as an avowal? Not being evidence-based might be one such criterion; certainly, our clear-cut instances above satisfy it. However, here is a different proposal, explicitly endorsed by Anthony Brueckner in *SK*: "in making an avowal, one self-ascribes a current mental state" (170). Further conditions pending, what this gives us is a necessary condition, in so far as one can only avow the current contents of one's mind. We may dub this *presentism about avowals*.[5] But why are only the current states of one's mind avowable?

Bar-On has proposed the following answer: avowals such as "My tooth hurts" or "I'll come at six" are expressions of one's mind, but one can only express the current contents of one's mind. In more depth, similar to the way in which I may express my fear when my teeth chatter uncontrollably or when I scream at the top of my voice, and similar to the way in which I may express my gratitude to you by heartily shaking your hand, I may also express my pain or my intention by avowing "My tooth hurts" or "I'll come at six", respectively. In all of these cases, one either acts (linguistically or non-linguistically) or is caused to behave in a way that is expressive of one's mental states. Furthermore, what makes one's actions and unintentional behaviour expressive is the presence of the relevant mental state, be it fear, gratitude or pain, in one's mind. Therefore, presentists conclude that only one's current mental states can be expressed through one's actions or unintentional behaviour; never the mental states of others, or one's own future or past mental states. (Cp. Bar-On 2004, 269-70)

It is a consequence of Bar-On's defence of presentism that memory self-ascriptions do not qualify as avowals. Although her reasons for this claim turn on the expressive nature of avowals, the exclusion of memory self-ascriptions from the set of avowals matches her intuitions regarding the lack of special security of memory self-ascriptions, for as she explicitly states, "memory reports have traditionally been contrasted with avowals in point of their security." (2004, 121) A couple of examples of the memory reports she gives are

[5] This must not be confused with the metaphysical view, also labeled "presentism", according to which only present facts or events exist. Presentists about avowals do not deny that one has a past, and therefore a mental history; rather, their claim is simply that only the current contents of one's mind are avowable.

"'I was very tired then', said by me at a later time" (2004, 16), or what I say "when I report to you how sad I felt yesterday" (2004, 269). However, as Bar-On also makes clear, the special security of avowals does not entail "that they are absolutely infallible or incorrigible [for] we do indeed sometimes take avowals to be false, and we do sometimes challenge or correct them." What matters here is that "we give considerable weight to the very fact that [somebody] avowed being in the condition ... [so]... we would eventually defer to a subject who persisted in avowing ... (assuming we did not question her sincerity)." (2004, 97) Now, it bears pointing out that the latter seems to apply not only to the current contents of the avower's mind, but also to his or her memory self-ascriptions. If I were to insist that yesterday I felt very sad, or that you had disappointed me sometime in the past, then assuming that you did not question my sincerity, you would defer to me. Therefore, in terms of special security, memory self-ascriptions do not differ from clear-cut instances of avowals. So, why do presentists contrast memory self-ascriptions with avowals?

An interesting clue is provided by the fact that avowals are contrasted not only with memory self-ascriptions, but also with ascriptions of mental states to others. I may realize, on the basis of my observation of your behaviour, that you are disappointed about a cancelled trip. In that case, my utterance of "You are disappointed" is an expression of my belief that you are disappointed, something I might have made explicit by avowing "I believe you are disappointed". As Bar-On rightly notices, I cannot avow your mental states, for I cannot express them. So, when I avow "You are disappointed", I express my current belief, but not your disappointment. So far, so good; but Bar-On goes further, applying the same model to memory self-ascriptions. Thus, she states that "my assertions about these matters [i.e., that I felt sad yesterday, or that somebody else feels disappointed about a cancelled trip] may, of course, serve to express my beliefs about the presence of the states, but they will not express those states themselves." (2004, 269fn) In other words, when I say about some past event "You disappointed me then" or "I remember how disappointed you made me feel", I am self-ascribing and therefore avowing my current belief (obtained through personal recollection) about my past disappointment, but not my past disappointment itself. Here, memory self-ascriptions are made to fit the presentist mould by being turned into self-ascriptions of current beliefs.

There is reason to think that this is not a satisfactory account of memory self-ascriptions. When I say "You disappoint me!", what I am doing is bringing my current disappointment into the open. Similarly, if we are going over some past incident, and I say "I remember how disappointed you made me feel" (or simply, "You disappointed me then"), what I am doing is bringing my past disappointment into the open, not my current memory-based belief. So, unless one is misled by surface-grammatical differences between present- and past-tense self-ascriptions, there is no reason to deny that one can bring into the open, express, and therefore avow, both one's current and one's past disappointments.

To sum up, then, presentism about avowals is wrong. Presentists claim that one *cannot* avow one's past mental states, but one's memory self-ascriptions can (at least sometimes) serve to avow one's past mental states. There is nothing in this criticism to discredit Bar-On's conception of avowals as expressions; only her *presentist* view of avowals. But as will be seen next, the very idea that expressivism provides the right account of the status of avowals is a debated issue in contemporary philosophy.

5.

To recap some ideas already mentioned in section 3, contemporary work on self-knowledge

seeks to account for the special status of avowals, but the subsequent task gives rise to two different approaches. According to the first, the task at hand is simply, or primarily, epistemic – namely, to clarify the privileged-cum-peculiar nature of the epistemic mechanism that gives rise to avowals. According to the second, the task at hand is twofold: first, to explain our ordinary practice of *treating* avowals as specially secured; and second, to elucidate how it is that avowals amount to privileged self-knowledge. On the one hand, Dorit Bar-On's neo-expressivism is an instance of the latter approach, according to which avowals are expressions of one's mental life; and in *SK* she responds to objections by Anthony Brueckner. On the other hand, Alex Byrne's transparency-based account is an instance of the former approach, according to which avowals embody self-beliefs obtained on the basis of evidence available in the world; and in *SK* there is a very useful discussion between Alex Byrne and Brie Gertler about the prospects of the project. In this and the following sections, the pros and cons of both approaches will be examined in some detail. But first of all, it will be useful to expand our vista, by considering the place they occupy among other conceptions of self-knowledge.

Both neo-expressivism and transparency-based accounts are alternatives to detectivism, the view according to which self-knowledge involves some form of tracking, or detection, of the contents of one's mind. Cartesian introspectionism is a form of detectivism, where one's access to one's mind involves the exercise of a privileged epistemic faculty producing infallible and incorrigible beliefs. In the Cartesian picture, this epistemic faculty tracks the states of affairs of a *sui generis* non-material realm of reality, in accordance with ontological dualism. But detectivism and ontological dualism are two separate theses. Thus, some contemporary materialists have eschewed all references to a *sui generis* non-material realm of reality, but have endorsed the idea of a (highly reliable) tracking mechanism (albeit a fallible one, as mentioned in section 3). Thus, according to materialist introspectionists (e.g., Armstrong 1968; Churchland 1988), self-knowledge involves a mechanism to track the contents of one's material mind – i.e., the mind/brain studied by contemporary cognitive science.

Against this, a growing consensus has emerged in the late 20th and early 21st century that self-knowledge is not a matter of reliably tracking the truth about one's mind by looking inside, either into an immaterial realm of reality or into the material mind/brain. However, this reaction against detectivism has taken more than one form; therefore, different aspects of the detectivist picture have been called into question. Thus, classical expressivists (Carnap 1935) have claimed that it is wrong to think of ordinary self-knowledge in terms of truth, and therefore in terms of a mechanism to track the truth about one's mind.[6] In contrast, constitutivists have argued that the problem is not truth, adding that the truth about one's mind is not detected, but rather constituted by one's avowals; hence, their special status (Wright 1989, 1991, 2001). Neo-expressivism and transparency-based accounts are two alternative positions within this growing anti-detectivist consensus, as will be seen next.

To recap, the key claim made by neo-expressivists (Bar-On & Long 2001; Finkelstein 2003; Bar-On 2004) is that avowals are expressions of the contents of one's mind. This is not a novel claim, as it forms the kernel of classical expressivism about self-knowledge. According to classical expressivism, the asymmetry between specially secured (non-evidential) mental self-ascriptions and the ascriptions of mental states to others is in fact a semantic asymmetry, in so far as the former are expressions of one's own mind, whereas the latter are assertions (descriptions) instead. Neo-expressivists agree with this, but understand

[6] Classical expressivism has sometimes been attributed to Wittgenstein (1953, 1958), by some recent commentators (Bar-On & Long 2001), but this is a highly contentious attribution (cp. García Rodríguez 2012).

it somewhat differently. Thus, classical expressivists argue that, in so far as avowals are expressions of one's mind, rather than descriptions of a worldly state of affairs, they are neither true nor false; whereas neo-expressivists claim that avowals are both expressions and truth-apt. As Brueckner reminds us (172-3), this is generally regarded as an advantage of neo-expressivism over its classical cousin, for it allows for semantic continuity between first- and third-person mental ascriptions, thereby explaining two phenomena which are puzzling from a classical expressivist point of view: one, that my expressive utterance of "I am in pain" is true in the same circumstances as (and therefore, semantically continuous with) your assertoric utterance of "AG is in pain"; and two, that first-person utterances like "I have toothache" appear in certain truth-involving contexts, like inferences (cp. Geach 1965).[7]

On the basis of this neo-expressivist account of avowals, Bar-On sets out to clarify the nature of self-knowledge in two steps, accounting first for the special status of avowals, and then for the idea of privileged self-knowledge.[8] According to Bar-On, the special security of avowals is a matter of the fact that "avowals serve directly to *express* the avowed state" (193; italics in the original). For when one attributes a given mental state with a certain content to oneself, one is not exercising any epistemic or recognitional capacity, but rather "an *expressive* capacity, the capacity to *use* content *c* (rather than some other content *c'*) to articulate, or give voice to [one's] present state." (194; italics in the original) Furthermore, Bar-On claims that avowals express both a certain mental state (with a certain content), and also a self-belief about that first-order mental state (with its particular content), something she refers to as "the dual expression thesis". This is the basis for her claim that "avowals serve to articulate *privileged self-knowledge*." (190; italics in the original) The crux here is that such self-beliefs are both true and warranted (in line with the traditional JTB-account of knowledge), in so far as they are "warranted by the same thing that serves as the rational cause of the act of avowing – namely, the self-ascribed state itself." (199) This, according to Bar-On, is a form of warrant that "accrues to subjects only when issuing present-tense self-ascriptions of occurrent mental states, and only when they do so in the avowing mode" (2004, 405), which therefore can be seen to involve some form of privilege, in contrast to the warrant for ascriptions of mental states to others, or for bodily self-ascriptions.

Brueckner objects to this account of privileged self-knowledge, because in "this conception of the justification of second-order self-beliefs ... the truth-maker for such belief turns out to be identical to the justifier [...] in contrast to all other sorts of epistemic justification." (183-4) He uses perceptual beliefs as an example, and argues that in perceptual beliefs a distinction must be made between the truth-maker (a certain state of affairs) and the justifier (the believer's visual experience). But Bar-On is not impressed by this objection, and in her reply she points out that epistemological disjunctivists about perceptual belief claim that in veridical perception, it is the perceived state of affairs, rather than the perceiver's mental state, that justifies the belief in question. Therefore, she concludes that her account of privileged self-knowledge relies on a notion of epistemic warrant that is already available;

[7] To give an example, the inferential move from the premisses "People with toothache are more irritable than normal" and "I have toothache" to the conclusion "I am more irritable than normal" involves, on a common view of logical relations, truth-apt propositional contents. However, this creates a problem for a view like classical expressivism, according to which avowals like "I have toothache" are not truth-apt – namely, how can a truth-apt conclusion be drawn from a couple of premisses, one of which lacks, whereas the other has, a truth-value? Neo-expressivists avoid this problem, because the expressive meaning of avowals does not undermine their truth-aptness.

[8] Although this is not the focus of her chapter in *SK*, Bar-On reminds us (190-1) why she objects to materialist introspectionism: if self-knowledge were a matter of tracking the states of one's mind/brain, systematic global error would be an open possibility, despite the reliability of the mechanism involved, in which case it would make no sense for us to accept from the outset the asymmetries generally assumed to give rise to the problem of self-knowledge (cp. 2004, 95-104).

so, the use of such a notion should not be an obstacle *per se*, contrary to Brueckner's objection.

However, Brueckner has a stronger objection against Bar-On's defence of privileged self-knowledge – namely, that "it is not clear in the end what role is being played here by the idea that avowals express mental states" (186), since in her explanation of the positive epistemic status of a given self-belief all the work is done by the disjunctivist point mentioned above, rather than by the idea of expression. Bar-On's reply is less convincing here, and it will only seem plausible to those who are already persuaded of her distinction between two different questions: one about the special security of avowals, and the other about privileged self-knowledge. For what she says is that, once the separate nature of these two questions is allowed, all that remains to be done is to offer an account of privileged self-knowledge that is "compatible" (200) with the expressivist view of the security of avowals. However, this reply has all the trappings of an avoidance strategy, and arguably neo-expressivists need to do more here to clarify the relations between the notion of expression and that of knowledge, if indeed they wish to propose an expressivist account of self-knowledge including the two neat parts distinguished by Bar-On.

6.

Turning to the notion of transparency now, the main idea behind its use in contemporary work on self-knowledge is that, whereas neo-expressivists attack detectivism for taking self-knowledge to be a matter of a tracking mechanism, rather than the expression of one's mind; defenders of transparency-based accounts reject the idea of an *inner* tracking mechanism.[9] Taking the lead from Gareth Evans' (Wittgenstein-inspired) view that "in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world" (1982, 225), some contemporary philosophers have proposed an account where self-knowledge is a matter of "looking out", rather than "looking in". Abstracting from the visual metaphor contained in Evans' claim, the general point is that there is a constitutive link between one's beliefs and one's evidence; hence, in order to determine the content of one's own beliefs, one attends to the evidence for them, as it is available out there in the world. As Evans put it, "I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*." (1982, 225)

In this vein, Jordi Fernández (2003) has argued that obtaining knowledge about one's mind involves a transparent method, in so far as the second-order belief that I believe that *p* is justified in the same way as the first-order belief that *p* – namely, by a reliable mechanism of belief-formation. Among such reliable methods, Fernández includes perception, memory, testimony or inference; and his main point is that the deliverances of such mechanisms serve to justify both first- and second-order beliefs (i.e., beliefs about the particular states of one's mind). For instance, the deliverances of my perceptual mechanisms are those perceptual appearances (e.g., as to *p*) that both justify my perceptual belief that *p*, and my second-order belief that I perceptually believe that *p*. In general, then, knowledge of one's own mind conforms to the method of transparency.

Alex Byrne makes a similar point when he claims that self-knowledge is a matter of

[9] In *SK*, Alex Byrne further explains that Armstrong's version of materialist introspectionism explains why one's access to one's mind is different from other people's (i.e., the mind/brain can only track its own states in this way), but leaves the idea of privileged access unexplained (108-9). Instead, the transparency-based account he favours is proposed to account for both peculiar and privileged access, as will be seen later in this section.

reasoning according to the following epistemic rule "BEL: if *p*, believe that you believe that *p*." (112) In so far as one can try to follow such a rule only by recognizing the fact that *p*, then given the constitutive link between (one's first-order) belief and evidence (one's epistemic reasons), such recognition will automatically bring the second-order belief that one believes that *p*. Therefore, if knowledge about the contents of one's mind is indeed a matter of adhering to such an epistemic rule, then self-knowledge will be privileged (i.e., prone to lead to knowledge and avoid error), as intuition claims. Furthermore, nobody can obtain knowledge about someone else's mind by following similar epistemic rules. For instance, trying to get access to Kylie's mind by adhering to the rule "BEL-3: If *p,* believe that Kylie believes that *p*", would not get one nearer the truth about Kylie's mind: for all one knows, Kylie simply might have missed the fact that *p*. Therefore, by adhering to BEL, one obtains knowledge about the contents of one's mind that is not only privileged, but also peculiar, as intuition also claims.

Now, critics of transparency-based accounts of self-knowledge have objected that, however plausible the accounts may be as regards knowledge of one's beliefs, things are quite different with regard to knowledge of other mental states, like desire or intention (Gertler 2011), and especially sensations (Finkelstein 2003). The critics' point is that one does not seem to look out at the world, or attend to one's epistemic reasons, in order to determine whether one has a particular desire, intention or sensation. Against this objection, defenders of transparency have sought to illuminate how desire (Fernández 2007) or intention (Byrne 2011) might, after all, conform to the same transparency-related mould as belief. But it is worth pointing out here that the critics' objection carries weight only if one assumes that there must be a unifying account of self-knowledge. Without this assumption, the existence of these differences between beliefs and other mental states cuts no ice, and a transparency-based account of knowledge of one's beliefs by itself might be sufficiently interesting to illuminate one aspect of the phenomenon of self-knowledge.

Regardless of this, though, some critics have also objected to transparency-based accounts about the knowledge of one's own beliefs. Thus, in *SK* Brie Gertler has put forward two powerful objections. First, that transparency can explain how one forms new beliefs (namely, by considering the available evidence), but "does not *reveal* [one's] pre-existing belief[s]." (143) Second, that knowledge of one's pre-existing beliefs (e.g., dispositional beliefs) is often a matter of "looking in", which goes against the method of transparency. For instance, one sometimes believes that *p*, on the face of contrary evidence. In these circumstances, transparency would predict that, if one considers whether one believes that *p*, one will attend to the available evidence and form the corresponding second-order belief – in this case, that one believes that not-*p*. But assuming that this is *ex hypothesi* false (i.e., one does believe that *p*), the method of transparency "provides no access to beliefs that withstand counter-evidence." (143) Here is an example. Asked if one holds prejudicial attitudes towards a certain group of people, one may well answer in the negative, although one's behaviour includes clear signs, for anyone to see, that one indeed holds such attitudes. In so far as the evidence is available out there, and this clashes with the content of one's belief (to the effect that one holds no such prejudice), one's belief was not formed on the basis of the available evidence. Hence, the method of transparency does not explain why the subject holds such beliefs.[10]

[10] Two clarifications are in order. First, the objection does not trade on an underlying assumption of infallibility, for neither defenders nor detractors of transparency are committed to infallible self-knowledge. Indeed, the subject in the example holds a false belief, but this is relevant only as an indication that the subject's belief was not formed in accordance with the method of transparency. It is the latter that is the key to the objection, not the fact that the belief is false. Second, the objection relies on the availability of plain, rather than obscure and difficult to obtain, evidence. So, the fact that the subjects "misreads", or simply

To the charge that the method of transparency creates, but does not reveal, one's beliefs, and therefore provides no account of self-knowledge, defenders of transparency can reply that ordinary self-knowledge is a complex phenomenon; accordingly, knowing one's mind does not always involve access to what one already thought about a particular topic, but is sometimes a matter of making up one's mind (cp. Moran 2001), which in turn involves attending to the available evidence, given the constitutive link between one's beliefs and one's evidence. As for the point that the method of transparency does not provide access to beliefs that withstand counter-evidence, and that one must rely on some sort of looking-in mechanism instead, a possible reply is that such phenomena fall outside the boundaries of ordinary self-knowledge. The presence of beliefs that withstand counter-evidence is not something one can self-ascribe in the way that one ordinarily self-ascribes other mental states (including beliefs) – namely, other than through the use of such evidence-based routes as observation or testimony that are typically employed in the ascription of mental states to others. No doubt, one may learn to recognize the presence in one's mind of beliefs that withstand counter-evidence, but that would involve taking notice of certain tell-tale signs, as might someone else. Eventually, one may even learn to self-ascribe such beliefs, but only in this roundabout way. Now, in so far as beliefs that withstand counter-evidence do not bear the marks associated with ordinary self-knowledge, it is no surprise that the method of transparency, which is proposed as an account of ordinary self-knowledge, leaves them out.

7.

So far, neo-expressivism and transparency-based accounts have been presented as two alternative conceptions of self-knowledge. As there are significant differences between them, it might be natural to think of them as competing accounts. Nonetheless, there are points of contact between both theses. Thus, Bar-On has stated that "the transparency-to-the-world of intentional avowals will fall out as a consequence of the expressive character of *all* avowals" (2004, 318; italics in the original). Her view is that considering the question whether *p* is putting oneself in a position to express the contents of one's mind (e.g., that one believes that *p*); therefore, to express one's belief that *p*, one must attend to one's epistemic reasons for *p*. This suggests that, at least where intentional mental states like belief are concerned, neo-expressivist and transparency-based accounts coalesce, rather than compete.[11]

A different point of contact between neo-expressivism and transparency-based accounts of self-knowledge is a joint commitment to the view that self-knowledge involves second-order beliefs about what one believes, desires, feels and so on. The master thought here is that, given that knowledge involves belief (together with truth and justification), self-knowledge must involve beliefs about the contents of one's mind. At first sight, the thought that self-knowledge involves self-beliefs looks innocuous enough, given a standing commitment to the traditional JTB-account of knowledge. But extra care is needed here not to end up with an account that blurs the distinctive first-person component in ordinary

misses, the evidence provides no plausible way out; on the contrary, it is still the case that the subject's belief was not formed by attending to the evidence.

[11] An author who thinks of the relationship between neo-expressivism and transparency-based accounts in terms of competition is Finkelstein (2003). He thinks that the method of transparency cannot cover the whole range of mental states over which self-knowledge extends; especially, sensations (as already mentioned) and brute likes and dislikes. He has also claimed that even those cases best suited to be accounted for in terms of the method of transparency (namely, cases where one makes up one's mind and therefore avows it, through practical deliberation on the basis of one's own reasons) include non-transparent elements.

self-knowledge. This would happen if self-belief (and self-knowledge) were modelled too closely on belief (and knowledge) about the world at large (including other people), for the end result would be a view where the first-person component in self-knowledge would boil down to a relation to a mind that simply happens to be one's own. As will be argued below, there are strands in both the neo-expressivist and the transparency-based accounts considered above that signal this danger.

As explained earlier, Byrne's transparency-based account of self-knowledge is built around the following epistemic rule: "BEL: if *p*, believe that you believe that *p*." According to Byrne, BEL is a rule that, unlike the rule "BEL-3: If *p,* believe that Kylie believes that *p*", produces, in a way that is not available to anyone else, second-order beliefs about one's mind that are more prone to amount to knowledge than error. Now, let us assume that Byrne's account successfully overcomes the objections it faces (reviewed in the last section), and therefore asserts itself as a plausible account of self-knowledge. What this would amount to is that the second-order beliefs about oneself obtained through adherence to BEL would more likely amount to knowledge than those second-order beliefs about the mind of others obtained by adherence to BEL-3. But what Byrne's account would not have provided us with yet is an answer to the question why adherence to BEL gives us self-knowledge; or in other words, why BEL is a good epistemic rule to follow in order to obtain self-knowledge. Byrne would have provided us with an answer to the question why BEL is a better epistemic rule than BEL-3; and therefore, why the resulting second-order beliefs are better, or privileged (plus peculiar). But no attempt would have been made at explaining why BEL is a good rule for self-knowledge; or why following BEL clarifies the nature of one's relationship with the contents of one's own mind. What Byrne appears to be offering is the following recipe: follow BEL, and you will do well – self-knowledge-wise. But that is very much the end of it, if Byrne's recipe is meant to be taken at face value, without further enquiry into one's relationship with the contents of one's mind that underwrites BEL.

This deficit in Byrne's account appears to be directly related to the fact that the problem of self-knowledge is understood as the problem of clarifying the nature of the privileged-cum-peculiar epistemic mechanism underlying avowals, vis-à-vis ascriptions of mental states to others (and bodily self-ascriptions). Now, in so far as it is an integral part of Bar-On's neo-expressivist account of self-knowledge that one's relation to one's mind is not only, or primarily, epistemic, but rather expressive, it should be better placed than Byrne's transparency-based account to accommodate the distinctive first-person component in self-knowledge. Thus, according to neo-expressivism, the self-knowing subject is able to express his or her own mind by avowing it. Furthermore, as one can only express one's own mind by avowal, neo-expressivism appears to get to the core of the first-person component in self-knowledge.

But this is not all there is to Bar-On's neo-expressivist account of self-knowledge, for she also wishes to defend the claim that avowals represent or articulate genuine self-knowledge. To do so, she strives to make her expressive conception of avowals match the traditional JTB-account of knowledge. The crucial move here is her commitment to the dual expression thesis, according to which one's avowals are not only expressions of the contents of one's mind, but also expressions of one's self-judgement or self-belief about the (first-order) contents of one's mind. Thus, when avowing "My tooth hurts", one expresses both one's pain and the second-order belief that one's tooth hurts; when avowing "I'd like a cup of tea", one expresses one's desire for a cup of tea, and the second-order belief that one desires a cup of tea; and when avowing "I'll come at six", one expresses

one's particular intention, and the second-order belief that one has that intention.

Briefly put, the point of the dual expression thesis is that, in so far as avowals express self-beliefs, an expressivist view of avowals can begin to satisfy the requirements imposed on an account of self-knowledge by the traditional JTB-account of knowledge. Full satisfaction of such requirements would involve showing that avowals express justified and true self-beliefs. As seen in the last section, Brueckner has serious doubts that Bar-On's explanation of the justification component lines up with her expressivist commitments. These doubts target Bar-On's account, without targeting the dual expression thesis. However, as will be shown now, there are problems here, too.

Assuming that the self-knowing subject is able to express his or her own pains, desires or intentions (to name but a few examples of mental states) by avowing them, it must be stated from the outset that there is certainly something right in the idea that the self-knowing subject is aware of the avowed contents of his or her mind. In other words, those circumstances in which one apparently avows one's pains, desires, or whatever, but fails to be aware of the avowed contents of one's mind, are not cases of ordinary self-knowledge, and a supplementary story must be told so as to make those situations intelligible (perhaps it is a case of self-deception, or the words were uttered in a parrot-like manner, or from a detached or indirect viewpoint, but no real avowal was made). Thus, in a similar vein, it could be said that the self-knowing subject has access to the avowed contents of his or her mind. Now, what is critical about the dual expression thesis is that "access" is understood in a very particular way – namely, as having second-order beliefs, over and above the avowal of one's first-order mental states – something which is both unnecessary and contrary to the spirit of expressivism. It is unnecessary, because the very act of avowing (one's pain, desire, or whatever) in ordinary circumstances manifests the avowing subject's awareness of, or access to, the avowed contents of his or her mind, without having to add a new expressive act. It is contrary to the spirit of expressivism, because when access is understood as having second-order beliefs over and above one's first-order mental states, the self-knowing subject is also made to adopt a third-person stance towards the contents of his or her mind. As a result, the distinctive first-person component in self-knowledge captured through the idea of expression gets blurred. The upshot of all this is that the dual expression thesis undermines the potential of the expressivist conception of avowals to overcome the deficit of those contemporary accounts of self-knowledge, like Byrne's transparency-based account, that conceive of self-knowledge in an unduly narrow way, as a puzzle about epistemic mechanisms, warrant and the like.

Summing up, to the extent that neo-expressivism and transparency-based accounts of self-knowledge are symptomatic of contemporary work on self-knowledge, philosophers seem to be pulling in two directions. On the one hand, they aim to offer an account of self-knowledge that is adequately epistemic; hence, in accordance with the postulates common in contemporary epistemology, in particular the traditional JTB-account of knowledge. On the other hand, they seek to preserve a distinctive first-person component in ordinary self-knowledge. These two directions are not necessarily convergent. So, often, the end result will be either a narrow focus on epistemic issues, neglecting the distinctive first-person component in self-knowledge (as in Byrne's transparency-based account); or a distortion of the nature of that first-person component (as in Bar-On's neo-expressivism). If this diagnosis is correct, the task ahead is clear – namely, to illuminate the nature of ordinary self-knowledge, including its epistemic pedigree, without neglecting or blurring the first-person component. Unfortunately, knowing what one must do is not the same as doing it – not quite.

Acknowledgements

Work for this paper has been funded by two research grants from the Spanish government (references FFI2009-13416-C02-01 and FFI2012-38908-C02-02). For helpful comments and criticisms, I am grateful to the referees appointed by this journal. Special thanks are due to Noreen Mabin for her skilled editing of the English text.

REFERENCES

Armstrong, D.M. 1968. *A materialist theory of the mind*. London and New York: Routledge and Kegan Paul. (Revised edition, 1993).
Bar-On, D. 2004. *Speaking my mind*. Oxford: Clarendon Press.
Bar-On, D. & Long, D.C. 2001. Avowals and first-person privilege. *Philosophy and Phenomenological Research* LXII: 311-335.
Byrne, A. 2011. Transparency, belief, intention. *Proceedings of the Aristotelian Society Supplementary Volume* LXXXV: 201-221.
Carnap, R. 1935. *Philosophy and logical syntax*. Reprinted by Thoemmes Press in 1996.
Churchland, P.M. 1988. *Matter and consciousness*. Cambridge, Mass. and London: MIT Press.
Evans, G. 1982. *The varieties of reference*. Oxford: Clarendon Press.
Fernández, J. 2003. Privileged access naturalized. *Philosophical Quarterly* 53: 352-372.
Fernández, J. 2007. Desire and Self-Knowledge. *Australasian Journal of Philosophy* 85: 517-536.
Finkelstein, D. 2003. *Expression and the inner*. Cambridge, Mass. and London: Harvard UP.
García Rodríguez, Á. 2012. How to be an expressivist about avowals today. *Nordic Wittgenstein Review* 1: 81-101.
Geach, P. 1965. Assertion. *Philosophical Review* 74: 449-465.
Gertler, B. 2011. Self-Knowledge. *The Stanford Encyclopedia of Philosophy.* Edited by Edward N. Zalta. Accessed January 20th, 2012. http://plato.stanford.edu/archives/spr2011/entries/self-knowledge/.
Hatzimoysis, A., ed. 2011. *Self-Knowledge*. Oxford: Oxford UP.
Moran, R. 2001. *Authority and estrangement*. Princeton and Oxford: Princeton UP.
Wittgenstein, L. 1953. *Philosophical investigations*. Oxford, Blackwell. Translation by G.E.M. Anscombe. (Revised English translation, 2001)
Wittgenstein, L. 1958. *The blue and brown books*. Oxford: Blackwell.
Wright, C. 1989. Wittgenstein's rule-following considerations and the central project of theoretical linguistics. In: *Rails to infinity,* 170-213. Cambridge, MA: Harvard University Press.
Wright, C. 1991. Wittgenstein's later philosophy of mind: sensation, privacy and intention. In: *Rails to infinity,* 291-318. Cambridge, MA: Harvard University Press.
Wright, C. 2001. The problem of self-knowledge (II). In: *Rails to infinity,* 345-373. Cambridge, MA: Harvard University Press.

Received: September 13, 2012
Accepted: August 18, 2013

Department of Philosophy
University of Murcia
Campus de Espinardo
E-30071, Murcia, Spain
agarcia@um.es