Philosophia (2019) 47:281–291
https://doi.org/10.1007/s11406-018-9995-4

Précis

Annalisa Coliva[1]

Received: 18 June 2018 / Accepted: 18 June 2018 /
Published online: 17 July 2018

The main and novel idea in *The Varieties of Self-Knowledge* is that self-knowledge—that is, our knowledge of our own mental states—comes in many ways. We have first-personal knowledge of our own mental states when, for instance, we are immediately aware of our occurrent sensations. By contrast, we have third-personal knowledge when, for example, we realize that we enjoy a given mental state by reflecting on our behavior and by inferring to its likely cause. Even when distinctively first-personal knowledge is at stake, it must be kept in mind that we have a variety of mental states. For instance, we enjoy sensations, such as pains and tickles, which have a characteristic phenomenology, but also perceptions that have both a phenomenal and a representational content; we have propositional attitudes, such as beliefs, desires and intentions and these come in various fashions—as dispositions and as commitments—that is, as the result of one's own deliberations based on considering evidence for or against a given proposition or course of action. Finally, we enjoy emotions, whose nature still escapes philosophical consensus. Such a variety of mental states invites caution in propounding single, all-encompassing accounts of how we may know each of these types of mental state. In particular, while it is clear that sensations and at least some emotions have a distinctive phenomenology and can be had also by creatures who cannot self-ascribe them, it is more difficult to maintain that propositional attitudes have an intrinsic phenomenology which can distinguish wishes from hopes, say, or beliefs from acceptances, etc. Perceptions too have their typical phenomenology, but they also provide a representation of the environment around the perceiver, or of her body, which is independent of the exercise of concepts, at least when "basic" perceptions are at stake. Hence they can be enjoyed by creatures who are incapable of self-ascribing them. By contrast, for propositional attitudes as commitments it makes sense to hold that they can at least in part be constituted by their very self-ascription, like when one deliberates by judging "I intend to do such and so" and there does not seem to be any room for the suggestion that one would thereby be tracking a pre-existing intention.

✉ Annalisa Coliva
a.coliva@uci.edu

[1] Department of Philosophy, University of California, Irvine, CA 92697-4555, USA

It is fair to say that, despite the fact that by now a lot of philosophers working on self-knowledge—particularly on first-personal self-knowledge—are aware of the limitations in scope of their preferred accounts, and are therefore at least implicitly committed to pluralism about self-knowledge (particularly of methods, but perhaps, in some cases of both methods and states), they have been reluctant to embrace it explicitly.[1] For some reason, which seems mostly to reveal a monistic prejudice, they seem to think that if their preferred theory has only limited application, it is not interesting (or not interesting enough). Subject to a craving for generality, which, as said, is likely due to a deep-seated monistic prejudice, they often attempt to extend their preferred theory of self-knowledge to mental states that are after all resilient to the treatment, thus ending up weakening their own accounts. Sometimes, in this vein, they realize that the attempt to generalise their preferred accounts stumbles, in particular, against the asymmetry between first- and third-personal self-knowledge; and, as a consequence, they are led to denying it, or to making it a difference in degree rather than in kind.[2] Or else, they tend to consider of limited philosophical interest and significance the kinds of self-knowledge they knowledgeably do not account for. Thus, you find theorists mostly interested in first-personal self-knowledge who downplay the importance of an inquiry into third-personal self-knowledge, typically on the grounds that it would not be especially interesting from an epistemological point of view.[3] Conversely, those who offer an account that works mostly for third-personal self-knowledge, and who realize that they cannot fully account for first-personal self-knowledge, insist on the irrelevance of the latter particularly to personal development vis-à-vis the importance of the former.[4] Hence, the bias towards monism can have various effects; going from leading one to the pursuit of generality at the expense of credibility, to the denial of structural differences between first- and third-personal self-knowledge, or, finally, to being chauvinist with respect to those forms of self-knowledge one admittedly cannot account for.

The Varieties of Self-Knowledge unashamedly buys into pluralism about self-knowledge. It does so by first presenting in some detail the plurality of mental states we enjoy and their intrinsic differences. It then defends the existence of a deep asymmetry—that is, an asymmetry in kind and not merely in degree—between first- and third-personal forms of self-knowledge. It then reviews several theories of first-personal self-knowledge, discussing their various pitfalls, but also accepting their kernels of truth, which are put at the service of a pluralistic account of self-knowledge. The latter consists in a plurality of methods, in particular when third-personal self-knowledge is at stake, and of states, since, while in third-personal self-knowledge subjects stand in a genuinely epistemic relation to their own mental states, in cases of first-personal self-knowledge they don't.

[1] A notable exception is Boyle (2009, 2011). His kind of pluralism, however, is more limited than the one defended in *The Varieties of Self-Knowledge*, for he mainly stresses the difference between first-personal self-knowledge of propositional attitudes as commitments and of one's passive mental states, such as sensations and perceptions. Furthermore, he thinks that knowledge of our own beliefs is more fundamental than any other kind of first-personal self-knowledge. The latter is not a claim I endorse.

[2] Gopnik (1983) and Cassam (2014) are a case in point.

[3] Moran (2001) is a clear case in point, but most theorists working on self-knowledge tend to implicitly endorse the same attitude as Moran's.

[4] Cassam (2014), who defends inferentialism, is a case in point. For a critical assessment, see Coliva (2016).

In more detail, the chapter titled "Varieties of mental states" introduces the variety of mental states we enjoy, and proposes a systematization of the complex geography of the mental. It first distinguishes between sensations and perceptions, by reference to the fact that only the latter have correctness conditions, while allowing that their contents, at least in the case of "basic" perceptions, may be entertained also by creatures who do not possess the concepts necessary to their canonical specification. Then propositional attitudes such as beliefs, desires and intentions are distinguished between two kinds—that is, those as dispositions and as commitments. While the former may be independent of judgment and may well be unconscious, the latter depend on judging either that P is the case, or that P would be good to have or do (in light of one's further goals). For such a reason, these mental states may also be called "judgment-dependent" propositional attitudes. Moreover, they constitutively involve the ability to accept criticism or of being self-critical if one does not live up to them. Afterwards, the complex case of emotions is considered. The view of emotions as *sui generis* mental states, not reducible to either sensations or judgements (or to an amalgam of the two), is put forward.

The subsequent chapter, titled "Varieties of self-knowledge", presents and discusses the characteristic traits of first-personal self-knowledge—namely, so-called "transparency", "authority" and "groundlessness". At first approximation, transparency amounts to the idea that subjects who possess the relevant concepts, as well as rational and possessed of normal intelligence, are such that when they enjoy a given mental state they are immediately in a position to self-ascribe it. Authority, in contrast, has that subjects' psychological self-ascriptions are correct, at least in the normal run of cases. Finally, according to "groundlessness", subjects' psychological self-ascriptions are not based on the observation of their own mental states or on inference starting from their own observed behaviour and possibly further aspects of one's own psychology. In fact, each element in this triad admits of different readings and the chapter discusses them in depth and defines their proper domain of application. The key idea defended in this chapter is that transparency, authority and groundlessness are not contingent but necessary and a priori aspects of what goes by the name of *first-personal* self-knowledge. For massive failures at this kind of self-knowledge would either display the lack of the relevant psychological concepts or failures at rationality. Rationality, in this connection, has to be understood in a "thick", rather than in a "thin" sense. The latter amounts to the idea that we are critical reasoners insofar as we revise our propositional attitudes and goals on the basis of countervailing reasons. However, I agree with several philosophers (Peacocke (1999), Bar-On (2004) and Cassam (2014), just to mention a few) who, *contra* Sydney Shoemaker (1996) and Tyler Burge (1996), do not think that self-knowledge is necessary for being critical reasoners. If that is the notion of rationality one has in mind, then lack of self-knowledge will not make one necessarily irrational. Yet, we also have a thick notion of rationality, according to which making certain psychological self-ascriptions and behaving in ways which run systematically against them would impugn the idea that we are confronted with a normal subject, up to the point of rendering her pronouncements onto herself irrelevant, a mere *flatus vocis* devoid of any significance, if not of meaning altogether. These characteristic traits of first-personal self-knowledge are then defended against possible objections stemming from recent findings in cognitive sciences. For instance, several studies in cognitive science tend to show that we do not have knowledge of our own character traits, that we are bad at affective forecasting—that is, at figuring out how we

would actually feel if some relevant change happened to our lives—and, finally, that we are really poor at identifying the causes of our decisions and further behavior. None of this, however, shows that we never have essentially first-personal self-knowledge. Rather, it shows that its scope is limited and does not extend to our deep seated and future dispositions, or to the causal relations among our various mental states, which are known in a third-personal way. Yet, all this is compatible with the fact that we have essentially first-personal knowledge of a wide range of mental states, such as our sensations, perceptions, basic emotions and propositional attitudes as commitments.

The volume then presents and critically discusses various accounts of first-personal self-knowledge that have been proposed, with special emphasis on contemporary versions of each of these theories. Hence, in the chapter titled "Epistemically robust accounts" examines inner-sense and inferentialist accounts of self-knowledge. The former model tends to equate self-knowledge to forms of knowledge based on outer observation, though granting a subject's privileged access to her own mental states. In particular, its contemporary versions, due mostly to David Armstrong (1968) and William Lycan (1996), claim that we have a reliable inner mechanism that "scans" our first-order mental states and produces the corresponding second-order ones. The chief objection against that view is that the model presupposes a crude form of reliabilism that severs the constitutive connection between self-knowledge, rationality and concepts' possession.

The inferentialist model, in contrast, tends to assimilate self-knowledge to knowledge of other people's mental states. Recently, it has been taken up and partially refashioned by Alison Gopnik (1983), who has developed a "theory-theory" account. Within the first 3–4 years of life, children acquire and develop a little theory of the mind, that they apply both to themselves and others, in order to (self-)ascribe mental states starting from the observation of overt behavior (or other "inner promptings"). Her views have given rise to a heated debate, at the interface of philosophy of mind, psychology and neuroscience, between supporters of the theory-theory approach and partisans of so-called "simulation" theories, such as Alvin Goldman (1993) and Robert Gordon (1995, 2007). According to simulation theorists, who are otherwise divided on many issues, knowledge of other people's mental states is not based on the application of a theory, but on the simulation of the other person's point of view, which gives rise to a psychological ascription based on what one oneself would feel and think if one were in the other person's shoes. These views are exposed and critically examined. The main objection against the inferentialist account is that it implausibly assimilates first-personal self-knowledge to knowledge of other people's mental states. Furthermore, it runs the risk of providing a circular account of self-knowledge and it succumbs as soon as one tries, like in Quassim Cassam's (2014) recent version of it, to make it transcend its proper domain of application. The main criticism against simulation theories, in contrast, is that they are in fact unclear about how we would get knowledge of our own minds, on the basis of which we should then gain knowledge of other people's mental states, and risk falling back onto other, problematical models of self-knowledge (such as the inner sense model). Simulation theorists, in particular Gordon, also have interesting but underdeveloped views about the nature and acquisition of psychological concepts, such as the concept of belief and of other propositional attitudes. Still, both inferentialism and simulative accounts have important things to say about some instances of third-personal self-knowledge, such

as knowledge of our deep-seated dispositions and the kind of self-knowledge we can gain through affective forecasting.

In the chapter called "Epistemically weak accounts" various models, which are united in claiming that self-knowledge is indeed a kind of modest, yet genuinely cognitive achievement, are discussed. Some of them can be traced back to some remarks by Gareth Evans in *The Varieties of Reference*. According to Evans (1982), in order to know our own beliefs, we only need to look outward, see if we can answer "yes" to the question as to whether P is the case, and then preface P with "I believe". Lately, Evans' insights have been developed especially by Richard Moran (2001) and Jordi Fernández (2013). The latter is criticized for implausibly claiming that the evidence which justifies one's belief in P is also the one that justifies one's self-ascription of that belief. The former, in contrast, is criticised for not offering any suitable explanation of why self-knowledge of our propositional attitudes should, after all, count as an epistemic achievement and for tending to equate first-personal self-knowledge with making up one's mind. Intuitively, however, we also have first-personal self-knowledge of several mental states, which are not the result of any deliberation on our part, such as sensations, perceptions and (at least basic) emotions.

Significantly different, yet still epistemic accounts have been proposed by Christopher Peacocke (1999, 2003) and Tyler Burge (1996, 2011). Peacocke, in particular, places crucial emphasis on the fact that first-order propositional attitudes have a characteristic phenomenology. Accordingly there is something that it is like to judge that P, for instance. We are therefore aware of our judgment that P, *qua* such a *judgment* and, by tacitly applying the rule that if one judges that P, one believes it, we correctly self-ascribe the corresponding belief. Burge's account, finally, takes self-knowledge to be a requirement of rationality (in a "thin" sense): in order to be rational thinkers we must be prepared to revise our beliefs on the basis of countervailing evidence. Hence, we are entitled—that is, non-discursively justified—to self-ascribe them. Such a second-order belief, in its turn, amounts to knowledge since it is true and justified (albeit non-discursively).

The main objection against those epistemic accounts that devote special attention to inner phenomenology is that such a distinctive phenomenology does not really differentiate between various kinds of propositional attitudes. For instance, it is difficult to say what distinguishes hopes from wishes at the phenomenological level. This will have a direct bearing on Peacocke's position. For, if the phenomenology is not sufficiently fine-grained to license a specific psychological attribution, it cannot be appealed to in order to explain self-knowledge. Furthermore, it is claimed, against Peacocke's position, that it runs the risk of providing a circular account of our knowledge of our propositional attitudes. For, if, in order to avoid the previous problem, it posits a subject's antecedent knowledge of her own beliefs (or of other related propositional attitudes such as judgments *vis-à-vis* beliefs), it would actually presuppose self-knowledge rather than explain it.

In contrast, Burge's account is criticized mainly for either implausibly claiming that "thin" rationality requires knowledge of what kinds of attitude one is enjoying, or else, for resting on an ad hoc notion of rationality, which compromises the interest of his theory. Moreover, claiming that self-knowledge is constitutive of being a reasoner does not provide an epistemic account of it. It merely points out an a priori connection. Indeed, if Burge were to supplement his account by saying that one gets to know one's

attitudes through the operation of some reliable cognitive mechanism, the epistemic aspects of his account of self-knowledge would be dangerously close to crude reliabilist theories of self-knowledge, already presented and criticized in the previous chapter.

The following chapter, titled "Expressivism about self-knowledge", considers expressivist accounts of our knowledge of our own mental states. The basic, underlying idea is that self-ascriptions of mental properties are ways of expressing our own minds other than in natural and instinctive ways, such as, for instance, by means of cries and laughter. After presenting and critically examining Wittgenstein's (1953) approach, which is at the origins of expressivist positions, as well as of some aspects of constitutive ones, I dwell on Dorit Bar-On's (2004) recent and powerful defence of that model. While generally sympathetic to that approach, I claim that it is much better suited to account for our knowledge of sensations, rather than of propositional attitudes and certainly it cannot be generalized across the board to provide an all-encompassing account of our knowledge of our minds. In particular, it does not explain those cases in which our first-order mental states originate in our self-ascriptions, like when, for instance, we deliberate "I intend to φ" or judge "I judge/opine/wish… that P" and there does not seem to be room for the idea that we would thereby be expressing a pre-existing mental state. Nor does it explain how we can actually have knowledge, obtained through a cognitive achievement, of many dispositional mental states we enjoy. Furthermore, difficulties emerge as soon as one tries to combine expressivism with the view that self-knowledge is, after all, the result of some sort of cognitive achievement, like in Bar-On's account. For if the model presupposes the existence of an inner scanning mechanism, it falls prey to the objections raised against inner-sense theories. If, in contrast, it presupposes some other kind of epistemic access to one's own first-order mental states, it succumbs to the difficulties presented against Burge's idea that we are entitled to our psychological self-ascriptions. Bar-On's new "expressive entitlements", moreover, are reviewed and found wanting. Hence, the supposed advantage of expressivism over its rivals, which should allegedly consist in avoiding observationalism, inferentialism and other unpalatable accounts of the epistemology of first-personal self-knowledge, is spoiled. Still, expressivism has something important to say about our "knowledge" of our own sensations; moreover, it can be extended also to our "knowledge" of our own perceptions and can offer interesting insights about the nature and the acquisition of several psychological concepts. These insights are built upon in the final chapter of the volume.

In the following chapter, so-called "constitutive" accounts of self-knowledge are dealt with. At the heart of this kind of approach lie two main ideas. First, that first-personal self-knowledge is not the result of any cognitive achievement, but rather consists in some conceptual truths, corresponding to transparency, authority and groundlessness, which can be variously redeemed. Hence, properly speaking, self-knowledge is not really a form of knowledge. This result is indirectly supported by the failure of the various attempts to account for first-personal self-knowledge as a real cognitive accomplishment examined in previous chapters. Second, proper constitutive positions are characterized by two metaphysical claims. The first one is that, under specifiable conditions, first- and second-order mental states do not have separate existence. The second is that, at least in part and under specifiable conditions, our first-order mental states are constituted by their very self-ascription.

The model has been defended in various ways starting with Sydney Shoemaker's (1996) pioneering work, through Crispin Wright's (2001) and Jane Heal's (2002) linguistic version of constitutivism, up to Akeel Bilgrami's (2006) agential version of constitutivism. A profitable way of presenting their debate is to see them as according different priorities to either side of the following biconditional, known as the Constitutive thesis, and as providing different characterizations of its C-conditions:

> Given C, one believes/desires/intends that P/to φ iff one believes (or judges) that one believes/desires/intends that P/to φ

According to Shoemaker, priority must be given to its left-to-right side and the C-conditions must be characterized by reference to subjects who possess normal intelligence, rationality and are endowed with the relevant psychological concepts. According to Wright, in contrast, the right-to-left side is the fundamental one and the C-conditions must refer to the communal linguistic practice of making psychological avowals, which are usually taken as authoritative. Finally, according to Bilgrami, the two sides of the biconditional are on a par and the C-conditions must make reference to the fact that the mental states at issue are such that it makes sense to regard the subject as responsible for them—that is, to be either blame- or praise-worthy for them. Each of these positions is presented and found wanting for either resting on dubious a priori claims regarding, for instance, the necessity of self-knowledge for being a reasoner, or for failing to vindicate the central metaphysical contentions of constitutivism.

I then introduce a metaphysically robust brand of constitutivism, which is claimed to hold only for a very limited class of mental states. Namely, for those propositional attitudes as commitments we undertake by deliberating what to believe, desire, intend to do, etc., on the basis of evaluating (or at least of being able to evaluate) evidence in favor of P/φ-ing, or of its desirability or advisability. When these propositional attitudes are at stake, and the subject is endowed with the relevant psychological concepts, which are acquired "blindly",[5] both sides of the biconditional hold as a matter of conceptual necessity and, in particular, the right-to-left side actually makes good the second metaphysical commitment characteristic of constitutive accounts. Thus, adult human beings actually have two ways of forming commitments, either by judging their contents, or else by directly self-ascribing them. In the latter case, then, authority is secured in a much stronger way, since the psychological self-ascription is actually self-verifying. Furthermore, the account is supplemented by an explanation of how we acquire and canonically deploy the relevant psychological concepts, which does away with the idea that psychological concepts are either tags for mental states one should already have in view, or a priori rules one should self-consciously apply, often by having in view either other mental states or even the very mental states one would thereby categorize. This account, in its turn, helps to make good the first metaphysical claim at the heart of constitutive positions. Namely, that, when subjects are rational, intelligent and conceptually endowed, first-order mental states and their self-ascriptions do not have separate existence. For the latter are seen as replacements of instinctive and direct forms of expression of one's on-going first-order mental states, which are integral

[5] That is to say, by being drilled to substitute their immediate avowal, "P", "P would be good to have", "I will φ", with the corresponding psychological one—that is, "I believe that P", "I want/desire that P", "I intend to φ".

to those very first-order mental states, rather than judgements about already singled out first-order mental states.

Such a position is then defended against the objection that we may be self-deceived and thus ascribe to ourselves a mental state—particularly a propositional attitude—we in fact lack. The key move consists in denying—following Bilgrami's lead—that self-deception is a case in which one goes wrong about one's first-order mental states. Rather, it consists in having two mutually inconsistent propositional attitudes—one as a commitment and one as a disposition—which give rise to a subject's somewhat irrational behavior. Yet one's self-ascription of the commitment is actually correct, even if one happens to behave in ways that run contrary to it, due to one's counter dispositions.

In the last chapter, called "Pluralism about self-knowledge", a pluralist account of self-knowledge is put forward. As the discussion of the previous chapter makes apparent, constitutive accounts can hold in their full-blooded version only for our (so-called) knowledge of our propositional attitudes as commitments. By contrast, it is argued that knowledge of one's own propositional attitudes as dispositions is usually achieved through inference to the best explanation—in the same way in which we can know of other people's mental states by inferring to them from their owners' overt behavior and by exploiting some general theory of the mind. However, only in one's own case can the inference be based on relevant inner promptings, such as sensations, emotions and further mental states. In some other cases, instead, it can depend on deploying simulative methods, like when we engage in affective forecasting. Moreover, it can be obtained by means of the self-conscious deployment of highly dispositional psychological concepts. In this case, there is inferential reasoning going on, but it is not a kind of inference to the best explanation. Rather, it consists in subsuming some aspects of one's overall behavior and mental states under a concept by self-consciously exploiting its characteristic notes. Finally, at times, third-personal self-knowledge can be obtained through testimony.[6]

Turning to first-personal self-knowledge, it is claimed that strong constitutive accounts have limited purchase also because, for instance, they do not extend to past self-ascriptions of propositional attitudes as commitments, which are known, when they are, based on mnestic evidence. Still, it is true that being able to remember one's past mental actions, or indeed other mental states, as well as one's own past actions, is constitutive of being a cognitively well-functioning human being. Yet, to stress, that does not mean that we can account for our knowledge of these past mental states along constitutivist lines.

Moreover, strong constitutive accounts are not apt to explain self-knowledge of our sensations and of other mental states that have a distinctive phenomenology and that are clearly independent of our ability to self-ascribe them, such as bodily sensations, basic emotions, perceptions and perceptual experiences. Here the most promising account will have to forsake the second metaphysical claim at the heart of strong constitutive explanations, according to which psychological self-ascriptions can at least partially constitute the first-order mental states they ascribe to a subject. What remains are simply the other characteristic claims of constitutive positions, according to which conceptually competent creatures are authoritative, at least in the normal run of cases,

[6] I have explored this theme further in Coliva 2018.

with respect to their own mental states and are immediately in a position to self-ascribe them without either observing their own mental states or their overt behavior. These first-order mental states, however, can exist independently of their self-ascription. Hence, the allegedly epistemic problem of self-knowledge becomes the problem of explaining how the relevant concepts are acquired and canonically applied without falling back into observational or inferential models. Expressivism becomes crucial in this connection because it allows one to avoid these pitfalls. In particular, the idea is put forward that when we deal with self-ascriptions of sensations and occurrent basic emotions, which have a distinctive (often bodily) phenomenology, possessing the relevant concepts is the result of having been drilled to substitute their more immediate expressions with verbal behavior. This conceptual drilling is what gives rise to their characteristic first-personal "knowledge". Yet, the latter is crucially not the result of any, however modest, cognitive achievement. Hence, the use of term "knowledge" in this connection is more the—"grammatical", as Wittgenstein would have it—signal of the absence of room for sensible doubt and ignorance (at least in the normal run of cases), rather than the mark of a genuinely epistemic relationship between a subject and her own sensations and basic emotions. Furthermore, seeing the avowal as a replacement of more instinctive forms of behaviour helps vindicate the claim that the first-order mental state and its self-ascription are not separate existences.

Similarly, I propose an expressivist account of our "knowledge" of our own perceptions, which is held to originate in blind drilling. The idea, once more, is that we first learn to voice their contents and, on that basis, we are drilled to express ourselves by prefacing such contents with "I see that" or "I hear that", etc. Therefore, our knowledge of our perceptions does not usually require us to attend to our experiences and to identify them as seeings (or hearings, etc.) either directly or through the application of a little psychological theory.

The case of non-basic emotions is different. For we usually know them by attending to a complexity of events, such as their characteristic phenomenological aspects (if and when they have them) as well as our own behavior in contextually salient occasions. Moreover, we usually infer from these data to their likely causes, such as the love for a given person or the envy for her success, etc. Indeed, our application of this little theory may often take place in rapid and almost unnoticeable ways, but only because we are already proficient in applying it. Indeed, genealogically or in new, unexpected cases it will require time and effort and possibly help from a third party. For we may well be at a loss about how to interpret the pool of data about ourselves we may have collected. That is to say, we may need the intervention of another person to be in a position to infer that our characteristic feelings and behavior are signs of love or envy. Moreover, a lot of our third-personal self-knowledge, such as affective forecasting or knowledge of one's deep-seated dispositions will depend on simulating relevant aspects of a given situation to see how we would react to it, thereby acquiring some insight into our own nature and character. Reading novels and watching movies can achieve similar results insofar as we may identify with the protagonists or be prompted to simulate salient aspects of the plot to see how we would react if we found ourselves in those situations.

Finally, it should be stressed that, contrary to the kind of self-verifying self-ascriptions that have commitments as contents, in all cases in which psychological self-ascriptions substitute more instinctive forms of behavior, there is however limited room for error. Due to slips of the tongue, or to somewhat impaired cognitive conditions, a

subject could actually voice sensations, basic emotions or perceptions she is not actually enjoying. Yet, constitutivism can take care of these possibilities by appropriately specifying the relevant C-conditions. By contrast, when the self-ascription of dispositions or of non-basic emotions is at stake, there is no default presumption that a subject should be authoritative with respect to them. For she will be as exposed to error as she would be if she were applying her psychological theory in order to get knowledge of another person's mental states.

At least since Shoemaker's work, an account of self-knowledge has been taken to have a bearing on the perplexing yet fundamental phenomenon of Moore's paradox—the paradox, that is, consisting in judging "P, but I do not believe it" or "I believe that P, but it is not the case that P". Accordingly, in the Appendix, the proposed account of commitments and their distinctively first-personal self-knowledge is brought to bear on it. In particular, it is claimed that only by countenancing propositional attitudes as commitments can Moore's paradox so much as exist. By contrast, if one took its doxastic conjuncts to express (the lack of) beliefs as dispositions, the paradox would, surprisingly, disappear. Indeed, the case of a self-deceived subject who discovers her self-deception can perfectly well illustrate the point. For one may find oneself in a position in which one would coherently assert "I believe that my husband is unfaithful to me, but he is not"; where the first conjunct expresses a disposition one has found out by observing one's own behavior and by inferring to its likely cause, and the second conjunct expresses one's belief as a commitment, given one's knowledge of one's spouse's loyal behavior. By contrast, it would seem that if, by uttering (or judging) that very sentence, one where trying, through its first conjunct, to express a commitment, its second conjunct would actually undo it. This, in fact, would generate a Moorean paradox. Thus, the interesting an novel result is that the existence of Moore's paradox can be secured only by countenancing essentially normative mental states such as commitments.

Hence, to conclude: what goes by the name of "self-knowledge" is a blend of disparate factors. Sometimes psychological self-ascriptions actually constitute the corresponding first-order mental states and while one cannot fail to "know" them, it is not because one entertains a particular epistemic relation to one's first-order mental states. Rather, it is because the self-ascription brings them about and is therefore necessarily authoritative. Some other time, they are alternative ways of giving expression to mental states, which can exist independently of them, resulting from being drilled to substitute their immediate expression with the relevant linguistic behavior. Still, under appropriately specified C-conditions, being in a position immediately to self-ascribe them and being correct in one's self-ascription are guaranteed to hold a priori and as a matter of conceptual necessity. Finally, in many cases, self-knowledge is actually the result of the application to one's own case of a little psychological theory or of simulative strategies, or indeed of an inferential deployment of highly dispositional psychological concepts, or, lastly, it is obtained through testimony. Only in these latter cases would self-knowledge be the result of some kind of cognitive achievement and the term "knowledge" would, accordingly, express an epistemic relation between a subject and her own mental states. In all other cases, by contrast, the term "knowledge" would rather signal the fact that there is no room for error, when self-verifying self-ascriptions are at stake, or at least not in the normal run of cases, when we are dealing with self-ascriptions of sensations, basic emotions, perceptions and perceptual

experiences. Either way, self-knowledge is valuable either because of its constitutive links with ("thick") rationality, concepts' possession, and, at least in some cases, responsible agency; or else, because it can help us have a better, more integrated and unitary life. Small surprise, then, that Western philosophy since its inception appropriated the *dictum* of the oracle of Delphi "Know thyself".

References

Armstrong, D. (1968). *A materialist theory of the mind*. New York: Routledge.

Bar-On, D. (2004). *Speaking my mind. Expression and self-knowledge*. Oxford: Oxford University Press.

Bilgrami, A. (2006). *Self-knowledge and resentment*. Cambridge: Harvard University Press.

Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research, 77*(1), 133–164.

Boyle, M. (2011). Transparent self-knowledge. *Aristotelian Society Supplementary, 85*(1), 223–241.

Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society, 96*, 1–26.

Burge, T. (2011). Self and self-understanding. The Dewey lectures (2007-2011). *The Journal of Philosophy, 108*, 6–7.

Cassam, Q. (2014). *Self-knowledge for humans*. Oxford: Oxford University Press.

Coliva, A. (2016). Review of Quassim Cassam *Self-Knowledge for Humans*. *Analysis, 76*(2), 246–252.

Coliva, A. (2018). Self-knowing interpreters. In P. Pedrini & J. Kirsch (Eds.), *Third-person self-knowledge, self-interpretation and narrative*. Dordrecht: Springer, forthcoming.

Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.

Fernández, J. (2013). *Transparent minds*. Oxford: Oxford University Press.

Goldman, A. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences, 16*, 15–28.

Gopnik, A. (1983). How we know our minds: the illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–15.

Gordon, R. (1995). Simulation without introspection or inference from me to you. In M. Davies & T. Stone (Eds.), *Mental simulation: Evaluations and applications* (pp. 53–67). Oxford: Blackwell.

Gordon, R. (2007). Ascent routines for propositional attitudes. *Synthese, 159*, 151–165.

Heal, J. (2002). First person authority. *Proceedings of the Aristotelian Society, 102*, 1–19.

Lycan, W. (1996). *Consciousness and experience*. Cambridge: MIT Press.

Moran, R. (2001). *Authority and estrangement*. Princeton: Princeton University Press.

Peacocke, C. (1999). *Being known*. Oxford: Clarendon Press.

Peacocke, C. (2003). *The realm of reason*. Oxford: Clarendon Press.

Shoemaker, S. (1996). *The first person perspective and other essays*. Cambridge: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

Wright, C. (2001). *Rails to infinity*. Cambridge: Harvard University Press.