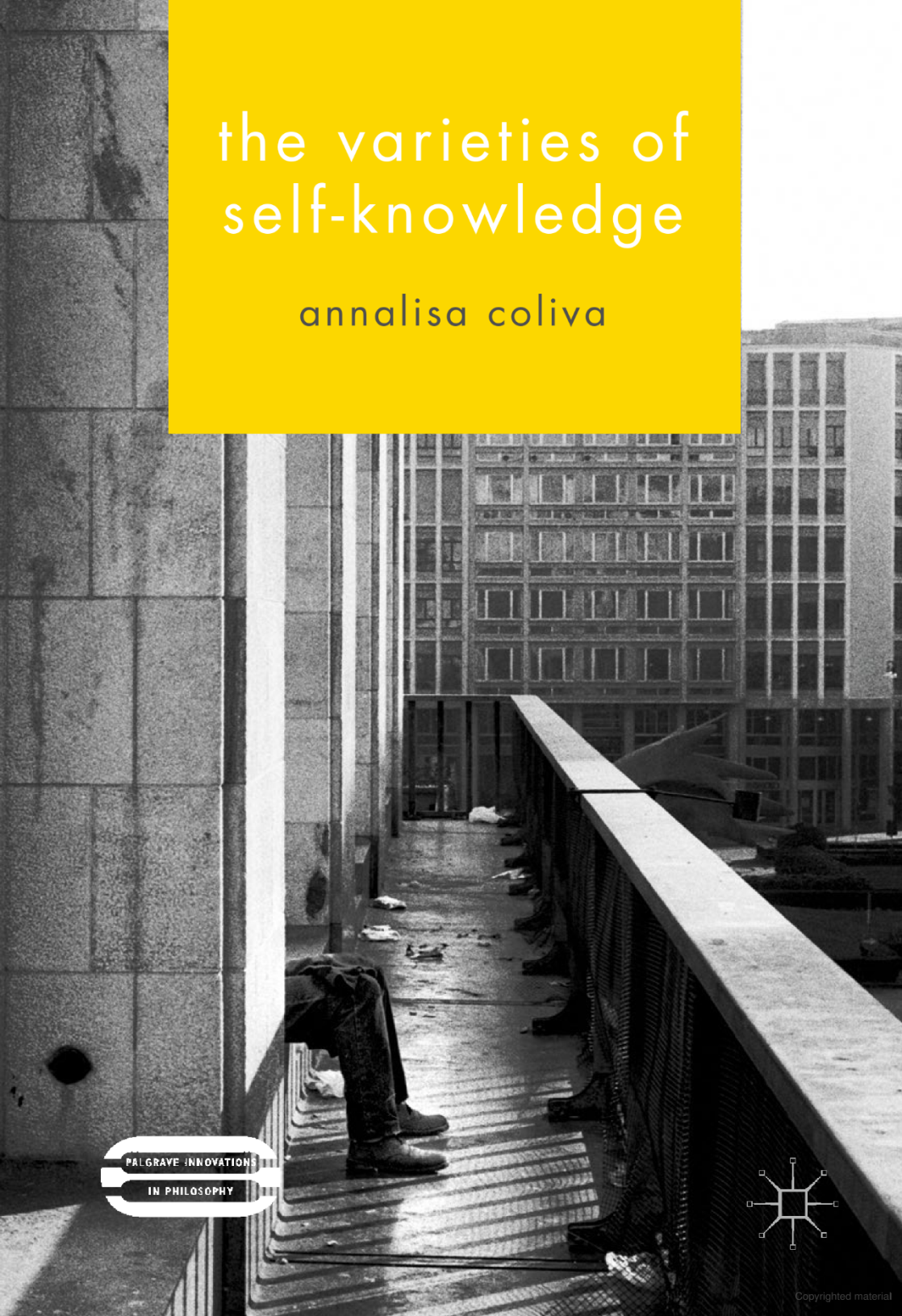
# the varieties of self-knowledge

annalisa coliva



PALGRAVE INNOVATIONS IN PHILOSOPHY

Annalisa Coliva

# The Varieties of Self-Knowledge

palgrave
macmillan

Annalisa Coliva
Department of Philosophy
Irvine, California, USA

Palgrave Innovations in Philosophy
ISBN 978-1-137-32612-6 ISBN 978-1-137-32613-3 (eBook)
DOI 10.1057/978-1-137-32613-3

Library of Congress Control Number: 2016936088



Cover illustration: © Vincenzo Dragani / Alamy Stock Photo

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Macmillan Publishers Ltd. London

# Contents

# 1

# Introduction

The main and novel idea presented in this book 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 realise that we enjoy a given mental state by reflecting on our behaviour 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—that is, as dispositions and as commitments—hence, 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, although it is clear that sensations and at least some emotions have



A. Coliva, *The Varieties of Self-Knowledge*,
DOI 10.1057/978-1-137-32613-3_1

a distinctive phenomenology and can also be had 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, and so on. 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 be constituted at least in part 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.

Even if one were convinced that first-personal self-knowledge is ultimately an epistemic relation between a subject and a proposition or a state of affairs, one should be open to the possibility that the methods whereby one gets to know in a first-personal way one's own mind can vary depending on the kind of mental state at issue. Hence, one should be open to pluralism about the methods whereby we gain self-knowledge, which go from being inferential, or even based on simulative routines, when third-personal self-knowledge is at stake, to possibly a variety of means when first-personal self-knowledge is at stake, depending on the kind of mental state one would know in such a way.

However, as will become apparent, in the case of so-called first-personal self-knowledge, there is no real epistemic relation between a subject and a proposition or state of affairs. So, to talk of "knowledge" in this connection is somewhat a misnomer, brought about by habit—in particular, by the philosophical habit of using that term and of mistaking certain conceptual truths, which entrain, in appropriate conditions, the indubitability of one's own psychological self-ascriptions, as in fact due to a peculiar epistemic relation holding between a subject and her own mental states.

Thus, the kind of pluralism propounded in this book is both a pluralism of methods when genuine knowledge of one's mind is at stake—that is, when third-personal self-knowledge is at stake—and of "states", as we may put it, for lack of a better term, when we are dealing with first-personal self-knowledge. Thus, in that connection, the term "self-knowledge" is in fact used as a shorthand for a set of conceptual truths

which can be variously redeemed. Hence, in some cases, through a variety of methods, a state of knowledge of one's own mind is obtained and expressed through the relevant psychological self-ascriptions. In some other cases, the relevant self-ascriptions (which in some cases may even be superficially identical to the ones which express third-personal self-knowledge) express a different and non-epistemic kind of state, which varies from merely showing or exhibiting one's mind to bringing about the relevant first-order mental states.

It is fair to say, however, that although 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[1] and therefore are 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.[2] 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.[3] Sometimes, in this vein, they realise that the attempt to generalise their preferred accounts stumbles, in particular, against the asymmetry between first- and third-personal self-knowledge; as a consequence, they are led to denying it or to making it a difference in degree rather than in kind.[4] 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 inter-

[1] A lot of them start out by presenting themselves as offering accounts of our knowledge of just this or that kind of mental state.

[2] A notable exception is Boyle 2009, 2011a. His kind of pluralism, however, is more limited than the one defended in this book, 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. No such priority claim is defended in this book.

[3] Several authors whose views will be considered in the following chapters are subject to this criticism, as we shall see.

[4] Gopnik 1983 and Cassam 2014 are a case in point.

ested 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.[5] Conversely, those who offer an account that works mostly for third-personal self-knowledge, and who realise 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.[6] Hence, the implicit 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.

Therefore, the present book 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 those kernels of truth they have, when they have them. In the last chapter, they are put at the service of a pluralistic account of self-knowledge, both of methods, in particular when third-personal self-knowledge is at stake, and of states, since, as anticipated, in many cases of so-called first-personal self-knowledge, the relevant psychological self-ascriptions do not depend on, and do not express, the obtaining of a genuine epistemic relationship between a subject and her own first-order mental states, as it were.

In the second chapter, titled "Varieties of Mental States", we introduce the variety of mental states we enjoy. We explore and propose a systematisation of the complex geography of the mental. We first distinguish 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

[5] Most theorists whose views will be discussed in the following chapters do that.

[6] Cassam 2014, who defends inferentialism, is a case in point. For a critical assessment, see Coliva 2015a.

specification. We then move on to propositional attitudes and distinguish between beliefs, desires and intentions as dispositions and as commitments. Whereas the former may be independent of judgement 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 "judgement-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, we consider the complex case of emotions. As is well known, there are, nowadays, a number of different and competing accounts, which range from identifying emotions with sensations to equating them with evaluative judgements or with perceptions of values or finally with felt bodily attitudes. It will be argued that none of these theories seems entirely satisfactory, although a detailed treatment of each falls beyond the scope of this book. It will be claimed that if this is the case, it is really tempting to consider emotions as *sui generis* mental states, sharing some features of other mental states while not reducing to any of them. On this view, common sense would, after all, be right in considering emotions to be different from all other mental states and in grouping them under one special category.

In chapter 3, titled "Varieties of Self-Knowledge", we turn to 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 being 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 it 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 to the best explanation starting from their own observed behaviour and possibly further aspects of their own psychology. In fact, each element in this triad admits of different readings and in the chapter we go to some length in discussing them. Furthermore, their domain of application has to be properly limited and we spend time showing how that should be done. 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 display either 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 (Christopher Peacocke, Dorit Bar-On and Quassim Cassam, just to mention a few) who, *contra* Sydney Shoemaker and Tyler Burge, 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 will also be defended against possible objections stemming from recent findings in cognitive sciences and from scepticism about knowledge of the content of our own propositional attitudes deriving from the endorsement of semantic externalism. 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 behaviour. 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, if and when they are, 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.

Furthermore, some theorists take the rise of content externalism to be incompatible with at least authoritative self-knowledge regarding one's

current propositional attitudes, for, if externalism is correct, a subject may think of being thinking a water thought, say, when she is in fact entertaining a thought about twater, due to her actual causal connections with an environment in which lakes, rivers and seas are in fact filled in with XYZ, rather than $H_2O$. Let us grant, for the sake of argument, that externalism is correct. Let us further suppose that our subject is actually thinking a twater thought, unbeknownst to her. Still, she would seem to have essentially first-personal access to it, even if she may be wrong about its actual content. Hence, she would still have transparent access to the fact that she is entertaining a thought, rather than a hope or a wish; and her access would still be groundless—that is, it would be based neither on observation nor on inference. Finally, she would still be authoritative with respect to its seeming or apparent content. (Some theorists would call it "narrow content" and would happily acknowledge its existence alongside with "wide content", but we need not take a position about it here.)

In keeping with the characteristic aspects of the *Palgrave Innovations in Philosophy* series, the volume then presents and critically discusses various accounts of such privileged self-knowledge that have been proposed, with special emphasis on contemporary versions of each of these theories. Hence, in the fourth chapter, titled "Epistemically Robust Accounts", we start by considering the inner-sense account of self-knowledge. This model tends to equate self-knowledge with 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 and William Lycan, 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 will be that the model presupposes a crude form of reliabilism that severs the constitutive connection between self-knowledge, rationality and concepts' possession.

We then turn to inferentialist accounts of self-knowledge. The inferentialist model 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, who has developed a "theory-theory" account. Within the first 4 years of life, children acquire and develop a little theory of the mind, which they apply to both themselves and others, in order to (self-)ascribe mental states starting from the observation

of overt behaviour (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 and Robert Gordon. 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 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 fifth chapter, called "Epistemically Weak Accounts", we introduce and assess various models, which are united in claiming that self-knowledge is indeed a kind of modest, yet genuinely cognitive achievement, while trying to avoid construing self-knowledge either as due to the operations of an inner scanning mechanism or as a form of inferential knowledge. For this reason, they are called "epistemic" models. They come in various fashions, however. Some of them can be traced back to some remarks by Gareth Evans in *The Varieties of Reference*. According to Evans, in order to know our own beliefs, we need only

to look outward, see whether we can answer "yes" to the question as to whether P is the case, and then preface P with "I believe". Recently, Evans's insights have been developed especially by Richard Moran but also by André Gallois, Jordi Fernández and Alex Byrne. We pay special attention to Fernández's and Moran's more thorough accounts. Both are found wanting, even though for different reasons. The former is criticised 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 latter, 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.[7]

Significantly different, yet still epistemic, accounts have been proposed by Peacocke and Burge. 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 judgement that P, *qua* such a *judgement* 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 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

[7] For a criticism of Byrne's development of Evans's ideas, see Boyle 2011a.

fine-grained to license a specific psychological attribution, it cannot be appealed to in order to explain self-knowledge along the lines proposed by Peacocke. Furthermore, it is claimed, against Peacocke's account, 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 judgements vis-à-vis beliefs), it would actually presuppose self-knowledge rather than explain it.

In addition, Burge's account is criticised mainly either for implausibly claiming that "thin" rationality requires knowledge of the kind of attitude one is enjoying or 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 criticised in the previous chapter.

In Chap. 6—"Expressivism About Self-Knowledge"—we move on to 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 by means of cries and laughter or other behavioral manifestations. After presenting and critically examining Wittgenstein's approach, which is at the origins of expressivist positions, as well as of some aspects of constitutive ones (that are reviewed in Chap. 7), we dwell on Dorit Bar-On's recent and powerful defence of that model. Though generally sympathetic to that approach, we highlight the fact that, after all, it seems much better suited to account for our knowledge of sensations rather than of propositional attitudes, and certainly it cannot be generalised 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 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 a lot of dispositional mental states we enjoy. Furthermore, difficulties emerge as soon as one tries to combine expressivism with the view that first-personal 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 and basic emotions; moreover, it can be extended to our "knowledge" of our own perceptions and can offer interesting insights about the nature and the acquisition of several psychological concepts. These insights will be built upon in the final chapter of the volume.

In the seventh 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 (see Chap. 3), 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 characterised 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 pioneering work, through Crispin Wright's and Jane Heal's

linguistic version of constitutivism, up to Akeel Bilgrami's 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 characterisations 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 characterised by reference to subjects who possess normal intelligence and rationality and are endowed with the relevant psychological concepts. According to Wright,[8] 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 either for 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.

We then introduce a metaphysically robust brand of constitutivism, which is claimed to hold for only 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, and so on, on the basis of evaluating (or at least of being able to evaluate) evidence in favour 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",[9] 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

[8] Heal 2002 too defends this position, but our exposition will focus on Wright's more thorough account.

[9] 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".

two ways of forming commitments, either by judging their contents or 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 categorise. This account, in 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 ongoing first-order mental states, which are integral to those very first-order mental states, rather than as different mental states entered by making 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 behaviour. Yet one's self-ascription of the commitment is actually correct, even if one happens to behave in ways which run contrary to it because of one's counter dispositions.

In the eighth and final chapter, called "Pluralism About Self-Knowledge", a pluralist account of self-knowledge is put forward. As the discussion in Chap. 7 makes apparent, constitutive accounts can hold in their full-blooded version for only 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 achieved through a variety of methods. Hence, we sometimes know them 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 behaviour 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. Sometimes, we gain knowledge of our minds by relying on other people's judgements about us. Our self-knowledge is therefore achieved though testimony. Finally, knowledge of our dispositional mental states 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 behaviour and mental states under a concept by self-consciously exploiting its characteristic notes.

Strong constitutive accounts have limited purchase also because, contrary to what some of their supporters hold, they do not extend to past self-ascriptions of propositional attitudes as commitments, which are known, when they are, on the basis of 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 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,[10] 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, with respect to their own mental states and are immediately in a position to self-ascribe them without observing either their own mental states or their overt behaviour. These first-order mental states, however, can exist independently

[10] *Pace* McDowell 1994.

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 is once again 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 behaviour. 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 the 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, we 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", and so on. 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. While basic emotions like fear can be conceptualized and expressed just like expressivism recommends, with more complex emotions, such as jealousy or envy, we usually know them by attending to a complexity of factors, such as their characteristic phenomenological aspects (if and when they have them) as well as our own behaviour 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, and so on. 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 behaviour are signs of love or envy. Moreover, a lot of our third-personal self-knowledge, such as affective forecasting or knowledge of our 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 behaviour, there is, however, limited room for error. Owing 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 behaviour 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 behaviour. By contrast, it would seem that if, by uttering (or judging) that very sentence, one were trying, through its first conjunct, to express a commitment, its second conjunct would actually undo it. This, in fact, would generate a Moorean paradox. The interesting and 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 although 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 therefore is necessarily authoritative. Some other time, our psychological self-ascriptions 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 behaviour. 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 may depend on testimonial evidence. 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 signal rather 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 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".[11]

## Bibliography

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

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

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

Coliva, A. (2015a). Review of Quassim Cassam *Self-knowledge for humans*, *Analysis, doi: 10.1093/analys/anv078*.

Gopnik, A. (1983). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1–15. Reprinted in Goldman, A. (ed) (1993). *Readings in philosophy and cognitive science*. Cambridge MA: MIT Press.

Heal, J. (2002). First person authority, *Proceedings of the Aristotelian Society*, *102(1)*, 1–19. Reprinted in (2003). *Mind, Reason and Imagination* (pp. 273–288). Cambridge: Cambridge University Press.Please provide page range for Heal (2002). 1–19

McDowell, J. (1994). *Mind and world*. Cambridge MA: Harvard University Press.

[11] According to the style of *Palgrave Innovations in Philosophy* series, while retaining the ambition of presenting a novel account, the volume also contains a detailed discussion of some prominent contemporary theories of self-knowledge. However, it does not address the issue of the first person, even though having a concept of oneself is necessary in order to make psychological self-ascriptions. The topic is extremely complex and will deserve a volume in its own right. I hope to be able to write it before too long.

# 2

# Varieties of Mental States

In this chapter, we explore and propose a systematisation of the complex geography of the mental. We first distinguish between sensations and perceptions (§1). We then move on to propositional attitudes and distinguish between beliefs, desires and intentions as "dispositions" and as "commitments" (§2). Finally (§3), we consider the complex case of emotions, whose nature still escapes philosophical consensus. After presenting and criticising several contemporary accounts, we put forward a borderline view of emotions.

This overview shows that we enjoy a variety of mental states, whose intrinsic features are extremely different. This paves the way to the claim, at the heart of this book, that single, all-encompassing accounts of how we can know our minds are unlikely to be successful, for what that knowledge is about is a heterogeneous mix. Just as not many will expect a uniform account of how we can gain knowledge of such diverse objects as truths about physical objects, moral truths and mathematical ones, say, so it is unrealistic to think that mental states, whose intrinsic features are very dissimilar, should be known in the same way, simply because they are all mental states. It would be like thinking that a uniform account is owed of how we know that there is a hand where we seem to see it or that there



A. Coliva, *The Varieties of Self-Knowledge*,
DOI 10.1057/978-1-137-32613-3_2

about perception empirically specific" (2010, p. 87). For perceptual psychology contributes law-like generalisations that explain "the processes by which perceptual states with specific veridicality conditions are formed from specific types of proximal stimulations" (2010, p. 88) as well as cases of perceptual illusion. The difference between veridical and illusory perception generally depends on differences in the actual, occurrent distal antecedents of a given type of proximal stimulation. Hence, causal relations between perceptual states and their *representata* are presupposed by scientific explanation. Moreover, the central problem of perceptual psychology—the so-called "underdetermination problem", according to which the same proximal stimulations are compatible with several different physical causes—is solved when the principles that govern the formation by perceptual systems of *veridical* perceptual states are discovered. Veridical perceptual states, in turn, are individuated by their relations to environmental entities. Hence, the solution of the central problem of perceptual psychology presupposes anti-individualism. Furthermore, according to Burge, the laws that govern perceptual systems are never attributable as acts to the perceiver, not even implicitly. They are computational formation principles, "inaccessible to consciousness and not under the perceiver's control" (2010, p. 94). They operate at the subpersonal level, although their results are constitutively attributable to the whole perceiver, despite not being necessarily conscious. Hence, once more, perceptual psychology is anti-individualist, insofar as it does not require a subject to be able to represent the conditions which make perception possible. Finally, perceptual systems are domain-specific, (partially) encapsulated from other cognitive systems, although they can interact with other systems, and are shared across a wide number of species. All these aspects further support the view that perceptual psychology is deeply committed to anti-individualism.

## 1.1 The Objectivity of Perceptual Representation

The crucial issue addressed by Burge is what it means to say that perception affords an objective representation. "Objective" as used here connotes being a product of objectification, which "is formation of a state with a representational content that is *as of* a subject matter beyond

idiosyncratic, proximal, or subjective features of the individual" (2010, p. 397), comprising entities in one's physical environment and also one's own body. According to Burge, in order to perform objectification, the system must discriminate one shape from the other, but also shapes from other relevant elements, which are environmentally salient and could have an impact on the needs and activities of the perceiver—similarly, for the perception of bodies, which must be discriminated from events, properties, and so on. Perception is still objective even if the perceptual system is incapable of discriminating these elements from illusions, proximal stimulations, abstract kinds, undetached entity parts, and so on. For the latter do not figure as relevant alternatives in a causal account of the formation of the perceptual states or figure in natural biological explanations of functional individual needs and activities. Hence, "the perceiver's objectifying discriminatory abilities determine the nature and content of his perceptual abilities only within this larger environmental and ethological framework" (2010, pp. 407, 466).

Another fundamental facet of objectification, according to Burge, consists in the exercise of perceptual constancies, which allow us, for instance, to perceive a colour as the same even if it is presented in different ways, like when a white wall is perceived as having the same colour although it is unevenly illuminated. Again, perceptual constancies are at work when we perceive a given object as the same while we move further away (or nearer) to it, thus undergoing different proximal stimuli. According to Burge, perceptual constancies are necessary and sufficient for the system's being a perceptual system (2010, p. 413).

## 1.2 Perceptual Contents

Perceptions have representational contents, according to Burge. The latter are abstract kinds that fix conditions under which a psychological state is veridical. All perceptual representational contents are structured—that is, they have singular and general elements. The latter perceptually indicate certain types or attributes—roundness, to the right of, and so on—and attribute them to particulars. Burge calls them "perceptual attributives" (2010, p. 380). Perception, however, singles out also particulars: not only

bodies or events but also specific, contextually determined instances of properties and relations. These singular elements are labelled "singular perceptual applications". Both perceptual attributives and singular perceptual applications are semantically relevant: the former can rightly or wrongly indicate types or attributes or rightly or wrongly attribute them to particulars; the latter, in contrast, could fail to refer.

A close examination of perceptual psychology supports the view that the elements of perceptual contents are not objects and properties but perceptual modes of presentation of them. Specific objects and properties are relevant only in order to determine whether a given perceptual representational state is veridical. Moreover, according to Burge, there is a structural difference between perceptual and propositional content. The former necessarily involves singular, context-determined elements, which are categorised or grouped from a contextually bound perspective. What is not yet present, however, is the separation of attributions from singular reference, to arrive at propositional predication. "A capacity for such a separation is a central aspect of achieving the specific context independence and generality that are embodied in pure attribution, propositional thought and rational inference" (2010, p. 541). Moreover, the content of perception is similar to a map or a sketch from an egocentric perspective.[2] This is not the form of a proposition. In addition, while the transformations of perceptual states do not depend on the individual, the transformations among propositions—for instance, in inference—are normally acts by the individual. Furthermore, there is no perception of logical constants, while real propositional contents involve logical operations.

Finally, according to Burge, perceptual attributives are limited; they concern shape, spatial relations, colour, motion, texture, possibly danger, food, conspecifics, and so on. Burge calls these attributives "perceptually basic" (546). Perceptual beliefs containing only conceptualisations of perceptually basic attributives are called "basic perceptual beliefs" (*ibid.*). Many of our perceptual beliefs employ concepts, which go beyond the range of perceptually basic ones, like the concepts of baseball bat, CD player, and so on. However, according to Burge, "in any particular appli-

[2] In this respect, Burge's account of the content of perception is similar to Christopher Peacocke's scenario content. Cf. Peacocke 1992, Chap. 3.

cation (…) the broader type of perceptual beliefs ultimately relies on conceptualisations of basic perceptual attributives" (*ibid.*). In propositional thought about perceived particulars, the singular elements are inherited from perception and embedded in an inferential structure, which may involve also quantificational elements, although it need not do so.[3]

## 1.3 Sensory States and Sensations

Sensory states are different from perceptual ones in that they are not objective. That is to say, they do not represent external elements of reality as such. Take smell, for instance. In many cases, it seems to afford a manifold of stimuli without attributing them to some relevantly stable distal cause. For such a reason, sensory states do not have veridicality conditions. They are not correct or incorrect, true or false representations of something out there. They are merely subjective variations in proximal stimuli.

Sensory states, however, are not necessarily sensations. That is to say, there need not be a subjectively conscious phenomenal aspect to them. A creature may be hard-wired such that her sensory systems may register variations in temperature or pressure on her skin, so as to give rise to certain bodily movements, say, without her being conscious or experiencing a sensation of cold or heat or of increased or decreased bodily pressure. Similarly, perceptions, according to Burge, need not involve a conscious, phenomenal element, even if they are objective representations with veridicality conditions. Therefore, blind sight, for instance, would be a case of perception. Of course, this is not to say that this conscious, phenomenal aspect is not present in many, or even most, human perceptions.

[3] Burge compellingly criticises Elisabeth Spelke's claim that bodies are not represented as such in perception. Moreover, he convincingly argues that cohesion, solidity, boundedness and spatio-temporal continuity are properties, which can be represented as such in perception. According to Burge, the ability to discriminate three-dimensional figures from a background and to represent them as cohesive and bounded, together with the ability to track objects perceptually over time (although not necessarily in motion or behind occlusions), is "constitutively necessary to visually representing bodies as such" (2010, pp. 456, 458–459). By contrast, he thinks that perceptual attributions of solidity are not necessary to that end, even if they are sufficient for it. Notice, moreover, that, according to Burge, the ability to perceive bodies as such is not necessary for objective perceptual representation, although it is central to the development of our conceptual system.

It means merely that it need not be present for a creature to be able to perceive aspects of her environment.

Furthermore, it should be noticed that even if bodily sensations are—by definition—felt in one's body, they are not representations of it and its physical properties. I can feel pain *in* my knee, but this does not provide me with a representation *of* the knee—that is, of *its* location, extension and shape. Nor do they have veridicality conditions. Hence, I may be mistaken about where my pain is located and even feel it in a limb I no longer have. The sensation, however, remains, notwithstanding its erroneous localisation. Of course, there may be (quite unusual) cases in which it seems to one as if one is in pain when one is not or in which one takes a tickle for a pain. But that does not hinder the fact that if there is a given sensation, then one would be enjoying it, even if one were mistaken regarding its localisation or its conceptualisation.

One aspect Burge does not touch upon, but which is worth mentioning in this connection, is the presence of qualia—that is to say, the specific phenomenal aspect of a given sensation. It seems quite intuitive that a painful sensation feels different from a sensation of cold or heat or that the smell of coffee feels different from the one of chocolate or vanilla. It seems compatible with everything we have been saying so far that, for creatures who can enjoy sensations, they may also feel different to them, without requiring any conceptualisation on their part, while their identification as sensations of pain (as opposed to sensations of cold or heat) or as smells of coffee (as opposed to smells of chocolate or vanilla) would depend on the exercise of concepts.

## 2 Two Kinds of Propositional Attitudes: Dispositions and Commitments

Human beings not only enjoy sensations and perceptions but are capable of thoughts. They believe that the sun will rise tomorrow, they desire to live a pleasant life, they hope their children will flourish, and so on. They are therefore capable of having various attitudes with respect to different propositional contents they can entertain, thanks to the possession of the relevant concepts. That is to say, one cannot believe, desire or fear that

on its appraisal. For, *ex hypothesi*, these creatures do not have the concepts necessary to grasp those contents, let alone the ones needed to assess the evidence in their favour.[6]

(ii) Mental states that are attributed to subjects to make sense of their behaviour, of which they themselves may be entirely *ignorant*. The latter class of mental states may comprise, but is not exhausted by, *unconscious* mental states of a Freudian kind, which, however, can be operative in shaping a subject's behaviour.[7] The idea here is that a subject may form these beliefs and desires as dispositions, as reactions to experiences undergone in early infancy and be entirely oblivious to them, while they do shape much of her overt behaviour. This is not to say that she may not acquire knowledge of them through therapy, say. Clearly, however, that would be a case of third-personal self-knowledge. Deep-seated biases would be another case in point, like gender preferences in offering certain kinds of job preferably to male (or female) candidates. Again, one may act on the basis of such a bias and could eventually recognise it, but if one did recognise it, the first-order mental state, if still in place, would remain dispositional, at least in the normal run of cases. Indeed, one may continue to act on its basis while sincerely judging the opposite.[8]

(iii) Also several propositional attitudes that are neither biases nor Freudian mental states but are self-attributed on the basis of an act of *self-interpretation*, by finding them out through the observation of one's own behaviour and other immediately self-known mental states, will fall into this category. For self-interpretation, when successful, makes one aware of a mental state that is already there yet is not "one's making" but rather something one finds oneself saddled with. A nice example, though not a case of propositional attitude, is provided by Jane Austen in her novel *Emma*, when the protagonist finds out about her love for Mr. Knightley,

[6] This makes it disputable that they could have the relevant beliefs as well, if those depended on having the concepts necessary to grasp the propositions which constitute their contents. If one were in the grip of such preoccupations, then a-conceptual creatures could at least be granted with proto-beliefs, desires and intentions. See Dummett (1996, Chap. 12).

[7] There may also be mental states which are attributed from a third party to make sense of a subject's behaviour, which are unconscious yet are not of a Freudian nature. The example discussed in (iii) would be a case in point if, instead of being self-ascribed, the mental state were ascribed by another person.

[8] We will discuss this possibility in the context of our treatment of self-deception.

her long-lasting friend, by reflection and inference on her own immediately available feelings of jealousy at the prospect that Mr. Knightley could return another woman's feelings.[9] Since this example will be discussed again in the following, it is worth quoting it in full:

> Emma's eyes were instantly withdrawn; and she sat silently meditating in a fixed attitude, for a few minutes. A few minutes were sufficient for making her acquainted with her own heart. A mind like hers, once opening to suspicion, made rapid progress. She touched—she admitted—she acknowledged the whole truth. Why was it so much the worse that Harriet should be in love with Mr. Knightley than with Mr. Churchill? Why was the evil so dreadfully increased by Harriet's having some hope of return? It darted through her, with the speed of an arrow, that Mr. Knightley must marry no-one but herself.

As the converse of self-interpretation, (iv) there may be mental states, which one can *predict*—either through inference or simulation—will assail one, in given circumstances, which, however, *will not be within one's direct control*. Perhaps, owing to one's long-lasting self-observations or through an act of simulation, one will know that if one were to work in an unsupportive environment for a while, one would start losing one's self-confidence and believing that one's work is meaningless or of poor quality. The characteristic feature of these mental states—in this case, the belief that unless one's work gains some kind of external recognition it is not worthy—is that one would seem to find oneself *saddled with them*, even if one were rationally able to find reasons that should make one think differently. This does not contrast with (c), that is the fact that one will not be held rationally responsible for these mental states. For indeed one would not be held rationally responsible for having them but only for not trying to get rid of them, when rationally unmotivated, once becoming aware of them through self-interpretation.

[9] The example is presented and discussed in Wright 1998, pp. 15–16, borrowed from Tanney 1996. Analogous examples could easily be construed for the case of propositional attitudes. Giorgio Volpe has kindly pointed out to me that also Schopenhauer, in *On Freedom of the Human Will*, holds the view that a person's character traits are known to her through reflection and inference on her past behaviour.

Another example could be the one of (v) propositional attitudes formed on the basis of habit. My actions show that I confidently believe that there is a floor behind the door of my bedroom. Or one's surprised reaction at the sight of a black sheep in a field, after seeing only white sheep throughout one's life up to that point, can show that one has so far believed as a disposition, and as a result of habit, that all sheep are white. Indeed, one might even know otherwise, but one's surprise would betray one's dispositional belief.

There may be more instances of propositional attitudes as dispositions, but the important point is that these would all be propositional attitudes manifested in the relevant first-order dispositions to act, while meeting conditions (a–c). Let us now turn to a different kind of propositional attitudes.

## 2.2 Propositional Attitudes as Commitments

Manifestly, adult human beings also have propositional attitudes that depend on a judgement based on the *assessment* of the evidence at subjects' disposal and that, for this reason, are within their control and for which they are held rationally responsible. Call them "intentional mental states as *commitments*" or "*judgement-sensitive* mental states".[10] Although the word "commitment" may have become common currency in philosophical literature nowadays,[11] there is still little agreement among its users about its meaning. For our purposes here, what is essential to commitments and makes them, in effect, very close to "judgement-sensitive" beliefs, desires, intentions, wishes, hopes and so on is the following:

(a') that they are the result of an action—the mental action of *judging* that P is the case (or worth pursuing/having)—on the subject's part,

[10] Cf. Bilgrami 2006, p. 213; Scanlon 1998, Chap. 1; Moran 2001, p. 116.

[11] Bilgrami makes extensive use of the term; Robert Brandom too, although he is more interested in stressing the social dimension of commitments than the former (or indeed myself). Furthermore, it is not my contention, somehow built in to the very notion of a commitment, that one should have knowledge of all the logical consequences of one's own beliefs and further propositional attitudes. As Bilgrami points out (2006, pp. 371–372, fn. 7, but see also pp. 376–377, fn. 20), the origin of the use of this term to refer to intentional states (or at least to a class of them) goes back to Isaac Levi.

on the basis of considering and hence of *assessing* evidence for P (is worth pursuing/having)[12];

(b') that these mental states are (at least[13]) *normatively constrained*—that is, they must respond to the principles governing theoretical and practical reasoning;

(c') and, in particular, they are so constrained (also) *from the subject's own point of view*;

(d') that they are mental states for which the subject is held *rationally responsible*.[14]

Hence, propositional attitudes as commitments depend on a subject's *deliberation* with respect to P, in the case of belief, of "P ought to be pursued" and of "It would be good (for me) if P were the case" in the case of desires and intentions, hopes and wishes, and so on, based on considering and evaluating evidence for P or for "P ought to be pursued", and so on. Judging that P is the case is constitutive of believing that P as a commitment. One cannot have or initiate the latter without the former.

[12] This is the main difference between the present account of commitments and Bilgrami's. For, in his view, commitments are not dependent on a subject's judgement.

[13] One may even hold that they are intrinsically normative and not merely—as it were, externally—constrained by normative principles. This is indeed the view that I favour and that will be put to use in the diagnosis of Moore's paradox (see Appendix). There is no need to take a stance on it at this stage, though, for the less committal view would still do, in order to mark out propositional attitudes as commitments from propositional attitudes as dispositions.

[14] This is the constraint Bilgrami identifies as essential to commitments, from which, on his view, (b') and (c') follow. However, he gives a moral or evaluative twist that it is best to resist. For, in his view, would one be held not only *rationally* responsible for one's commitments but also *accountable* at large. For instance, one might be *reproached* or *resented* for having certain commitments (Cf. Bilgrami 2006, p. 226). However, specified in the way Bilgrami characterises it, (d') is not sufficient to mark out the contrast between commitments and dispositions, because one can criticise or be criticised, and accept to be criticised, also (for) one's own dispositions, such as the disposition to smoke, or, to take a more loaded example, for wanting to get rid of other male opponents as a result of an unresolved Oedipus complex. But, surely, neither mental state is the result of a subject's action, for which one can be held rationally responsible, although one may be considered "badly"—in Bilgrami's extended sense of the term—for having it. It is then not by chance that, as a matter of fact, Bilgrami ends up endorsing the view that "we do have transparent self-knowledge of mental dispositions" (Bilgrami 2006, p. 287). I find this conclusion unpalatable, for, surely, when we do get knowledge of our unconscious mental states we obtain it through a process of self-interpretation or of analysis (that may or may not be guided by a therapist) relevantly similar to the ways in which we may come to attribute mental states to others. So, it seems to me that whatever knowledge we may eventually gain of our unconscious mental states, it is not "transparent" and is actually grounded in observation and inference.

Yet beliefs as commitments involve also dispositional elements, which are not involved in the mere act of judging that P is the case. For example, one ought to be disposed to use P as a premise in practical and theoretical reasoning and accept criticism or be self-critical if it were shown that P is not supported by sufficient evidence and, similarly, in the other cases of propositional attitudes as commitments and their relations to judgement.

Moreover, the evaluation of evidence may not always be explicit. For we sometimes form beliefs (and other mental states) as commitments immediately on the basis of this or that available evidence. However, one ought to be disposed to produce such evidence, were one requested to, and ought to withdraw from one's belief as a commitment (or other propositional attitudes of that sort), were it shown that the original evidence is either bad or insufficient evidence in its favour. Furthermore, to say that propositional attitudes as commitments are the result of a subject's deliberation based on considering evidence should not involve us in any form of doxastic voluntarism, as we will soon see. Moreover, it should readily be acknowledged that there will be forms of local holism between mental states as commitments. Indeed, viewing also desires, intentions and other propositional attitudes, beside beliefs, as commitments, tightly connects them with *believing* that their contents are worth pursuing or would be good for one if actualised.

In addition, these mental states are (at least) normatively constrained in the sense that they are subject to norms governing practical and theoretical reasoning. In the case of belief, as already remarked upon in passing, should countervailing evidence come in, a subject *ought* to withdraw from holding P, "P ought to be pursued" and "It would be good (for me) if P were the case" and so on. Thus, one ought to withdraw from one's belief that P is the case or from one's desire/intention/hope/wish that P should obtain. In the practical case, it is clear that practical syllogisms are sensitive not only to the propositional contents entertained but also to the way in which these contents are entertained. Hence, for instance, if a subject believes that it is raining, desires not to get wet and happens to carry an umbrella, she ought, *ceteris paribus*, to open it, whereas she ought not to have that behaviour did she merely wish it was raining.

Finally, since such "oughts" would have to be appreciated by the subject herself, were she not to withdraw from her beliefs and further propositional