

REASONING AND SELF-KNOWLEDGE*

MARTIN F. FRICKE

Instituto de Investigaciones Filosóficas - Centro Peninsular en Humanidades
y Ciencias Sociales - Universidad Nacional Autónoma de México
mfcephcis@gmail.com

Abstract

What is the relation between reasoning and self-knowledge? According to Shoemaker (1988), a certain kind of reasoning requires self-knowledge: we cannot rationally revise our beliefs without knowing that we have them, in part because we cannot see that there is a problem with an inconsistent set of propositions unless we are aware of believing them. In this paper, I argue that this view is mistaken. A second account, versions of which can be found in Shoemaker (1988 and 2009) and Byrne (2005), claims that we can reason our way from belief about the world to self-knowledge about such belief. While Shoemaker's "zany argument" fails to show how such reasoning can issue in self-knowledge, Byrne's account, which centres on the epistemic rule "If p , believe that you believe that p ", is more successful. Two interesting objections are that the epistemic rule embodies a mad inference (Boyle 2011) and that it makes us *form* first-order beliefs, rather than *revealing* them (Gertler 2011). I sketch responses to both objections.

KEY WORDS: Reasoning; Self-Knowledge; Transparency; Shoemaker; Byrne.

Resumen

¿Qué relación existe entre el razonamiento y el autoconocimiento? Según Shoemaker (1988), cierto tipo de razonamiento depende de un autoconocimiento: no podemos revisar nuestras creencias racionalmente sin saber que las tenemos, en parte porque no podemos ver que la inconsistencia de un conjunto de proposiciones constituye un problema a menos que estemos conscientes de crearlas. En el presente artículo argumento que esta propuesta es errónea. Una segunda posición, versiones de la cual se pueden encontrar en Shoemaker (1988 y 2009) y Byrne (2005) proclama que es posible razonar desde creencias sobre el mundo hacia un autoconocimiento sobre tales creencias. Mientras el *zany argument* de Shoemaker no logra mostrar cómo tal razonamiento puede resultar en un autoconocimiento, la teoría de Byrne, que se centra en la regla epistémica "Si p , cree que crees que p " tiene más éxito. Dos objeciones interesantes son que la regla epistémica incorpora una inferencia insana (Boyle 2011) y que nos hace *formar* creencias de primer orden, en lugar de *revelarlas* (Gertler 2011). Esbozo respuestas a ambas objeciones.

* An earlier version of this text was presented at the Ninth Annual UT/UNAM Workshop on Reasons in San Antonio in 2014. I would like to especially thank Brian Cutter for acting as a commentator on that occasion. I am also very grateful for the stimulating discussion with the other participants of the workshop. Finally, I am indebted to two referees of *Análisis Filosófico* for some very useful comments.

PALABRAS CLAVE: Razonamiento; Autoconocimiento; Transparencia; Shoemaker; Byrne.

1. The problem of self-knowledge

Knowledge of at least some of one's own mental states has two striking features. First, belief that one is in such a mental state is very likely to constitute knowledge; the possibility of error seems to be greatly reduced compared to belief about other people's mental states or the world in general. Some philosophers have argued that we have infallible knowledge of these states: we cannot go wrong when we believe that we are in them. Others merely attribute an authority to them greater than that which we have when trying to find out what others' mental states are. Here I shall only make the second claim. Although error is possible in beliefs about our own mental states, at least with respect to some of them the probability of such error is much lower than in belief about other people's mental states.

Second, such knowledge is acquired in a special way, which seems to be "more direct" than the one we have for acquiring knowledge about other people's mental states. Of course, more needs to be said about what this special way of acquiring self-knowledge is. Some ideas about this will be discussed in the third and fourth section of this text. To set up the problem of self-knowledge as I understand it, it suffices to agree that such knowledge is not acquired in the same way as we acquire knowledge about other people's mental states. We do not need to observe ourselves or to draw elaborate inferences in order to know our own mental states. Furthermore, the way we have of knowing our mental states is not available for knowing the states of others.

Alex Byrne (2005) summarises these two features of self-knowledge by saying that we have a "privileged" and a "peculiar" access to our own minds. He stresses that privilege and peculiarity can come apart. For example, our proprioceptive access to the position of our own limbs might be thought to be peculiar, because not being available to others, but not privileged, because it can easily be erroneous. On the other hand, we might think of a scientist specialising in some area as having privileged access to facts in this area, without her access also being peculiar, i.e. more direct and not available to others.

The fact that knowledge of some of one's own mental states exhibits both peculiar *and* privileged access is what I would like to call the problem of self-knowledge. We have some idea as to what makes beliefs about the world especially error-proof. If I want to be

sure about whether I still have some milk in the fridge, I examine the fridge. I look closer, from different angles, perhaps use sophisticated instruments such as microscopes, chemical analyses etc. to examine the contents of my fridge, back up my findings with tested theories or consult other experts on the matter. All these are ways of making my beliefs more error-proof. Some of them are also available if I want to have greater certainty about the mental states of other people. Perhaps there are no special instruments for finding out what someone else thinks. However, it is clear that I can improve my accuracy of mental state ascriptions to others by conversing with them, observing them carefully and consulting with others who know them well.

The problem of self-knowledge consists in the fact that our beliefs about some of our own mental states are highly error-proof despite only being based on a peculiar access. They are not based on close observation, backed up by theory, experiment or consultation with other experts; yet they are less prone to error than ordinary beliefs about the world and less prone to error than beliefs about other people's mental states. How can this be, given that none of the elements that ordinarily make beliefs more error-proof seems to be present?

In this paper, I shall examine whether the problem of self-knowledge can be explained by an appeal to features of our powers of reasoning. Sydney Shoemaker claims that self-knowledge is necessary for the kind of reasoning we engage in and understanding this necessity we might hope also to understand how it can combine privilege and peculiarity. Alex Byrne, on the other hand, tries to show that self-knowledge naturally arises out of ordinary powers of reasoning. Both kinds of argument have in common that they do not postulate an inner sense to explain self-knowledge. In what follows I evaluate these arguments and examine whether they help to solve the problem of self-knowledge. I will argue that Shoemaker's arguments are inconclusive and cannot explain the problem. Byrne tackles the problem, but has difficulties explaining knowledge of one's own standing states. Overall, I think that Byrne's account is the more promising approach that should be further investigated. Throughout my discussion, I shall concentrate on the central case of knowledge of one's own present beliefs.¹

¹ For self-knowledge to be privileged and peculiar, the beliefs known can be occurrent or dispositional, but they must be present states.

2. Shoemaker: reasoning requires self-knowledge

One salient feature of our reasoning is that it is rational (in some sense to be defined). According to Sydney Shoemaker, such rationality is only possible if the reasoning subject has self-knowledge. I shall look at four versions of this argument.

The first version says, “[d]eliberation is a self-critical enterprise. One’s beliefs, desires and intentions are up for review, and for this to occur one must not only have them but be aware of having them.” (Shoemaker 1988, p. 187) So here, the rationality of our reasoning is seen as residing in our ability to review our own beliefs, desires and intentions. The argument then seems to rely on the idea that one cannot revise what one does not know one has. I shall not dwell on the point that one might revise things, such as other people’s cupboards, without “having” them in the sense one has beliefs. More importantly, it is not clear that one can only revise knowing that one has the thing revised. It seems that there are many corporeal facts, such as one’s pulse, rate of glucose in the blood, tension of various muscles etc., that are under our “revision”, without us being aware of them. Of course, Shoemaker is talking about the revision of mental states, not the states of such “lower” organs. Does this make a difference? Suppose I notice that it has started to hail and against my earlier intentions I decide not to go for a walk. Nowhere in this simple revision of beliefs and intentions does it seem necessary to self-ascribe these states. It seems that animals such as cats and dogs should be able to perform such forms of reasoning; yet we probably do not consider them capable of self-knowledge. Therefore, it is unclear why we might not revise beliefs, desires and intentions in some kind of automatic way, without being aware of having them.²

² Shoemaker concedes at several points (Shoemaker 1994, pp. 285f.; 2003, p. 399; 2009, p. 39) that some rational revision of beliefs and desires might go on automatically, but insists that there “are cases in which the revision of the belief system requires an investigation on the part of the subject, one that involves conducting experiments, collecting data relevant to certain issues, or initiating reasoning aimed at answering certain questions” (Shoemaker 2009, p. 39) and in these cases he thinks that self-knowledge is necessary for the investigation. I think that Shoemaker is right in that our scientific investigation, for example, explicitly proceeds with the help of second-order knowledge of objectives (intentions), of what is known or surmised (beliefs), of theories under consideration, etc. However, I have two questions: (1) Is it *necessary* for such investigation to rely on second-order knowledge? (2) The second-order knowledge that scientific research uses is fixed in notes or publications. It is clearly not of the peculiar kind and it might be privileged only in a select few specialists. So is it *necessary* for any investigation to rely on the peculiar and privileged self-knowledge

The second version of the argument runs as follows:

Suppose that one's standing beliefs include the belief that P and the belief that if P then Q, and that one now comes up against evidence that Q is false. To see that there is a problem here that calls for resolution, it is not enough to be aware that the propositions "P," "If P, then Q," and "Not-Q" form an inconsistent triad; one must also be aware that these are all propositions one believes or is disposed to believe or has *prima facie* reason to believe. (Shoemaker 1988, p. 187)

Here, the idea seems to be that an inconsistency among a set of propositions is only a problem for me if I know that I believe them (or am disposed to believe them or have *prima facie* reason to believe them). After all, if I do not believe them the coherence of my beliefs is not threatened by this particular inconsistency.

However, this second claim (if I do not believe the set of inconsistent propositions the coherence of my beliefs is not threatened) is no good reason for the first (such inconsistency is only a problem for me if I know that I believe those propositions). In general, the following argument is a *non sequitur*:

- (1) P, only if Q
- (2) Therefore, I can know that P, only if I know that Q.

Consider, for example:

- (3) This figure is a circle, only if all its points are equidistant to a single centre point.
- (4) Therefore, I can only know that this figure is a circle, if I know that all its points are equidistant to a single centre point.

(4) does not follow from (3). I can know that some figure is a circle, for example by comparing it with some sample, without knowing that its points are equidistant to a single centre point. Knowledge is not closed with respect to implications.

However, Shoemaker seems to suggest just such an argument:

- (5) I only have a problem with the inconsistency of a set of

which is our concern here? It is not clear to me that either of these questions has to be answered affirmatively.

propositions, if I believe all the propositions of the set.

- (6) Therefore, I can only know (“see”) that I have a problem with the inconsistency of a set of propositions, if I know that I believe all the propositions of the set.

(6) does not follow from (5).³ I would suggest that we could notice the problematic character (for us) of an inconsistent set of propositions without self-ascribing beliefs about those propositions. For this, it is sufficient to notice that there are good reasons for taking the propositions in question to be true and to notice that they are inconsistent. These facts, if noticed by us, indicate clearly that we have a problem. It is impossible for all the propositions to be true; therefore, our reasons for taking them to be true cannot be the whole story. We now know that we have to dig deeper if we want to know the truth of the matter. However, nowhere in this process do we need self-ascriptions of belief; and in this sense self-knowledge of our own beliefs is not necessary. In fact, even the earlier quote from Shoemaker suggests that it might be sufficient for us to be aware only that the propositions in question are ones that we have “prima facie reason to believe”. However, this is not self-knowledge of belief.⁴

Shoemaker’s third argument for the claim that rational reasoning requires self-knowledge has the form of a *reductio*:

Someone who had no idea what he believed could not entertain the possibility that any specific one of his beliefs was wrong, and could not be led by doing so to initiate activities aimed at testing that possibility. If such a person’s beliefs were inconsistent, and he were aware of the inconsistency between the propositions believed, he would have to think, incoherently, that the facts were inconsistent! (Shoemaker 1988, p. 187)

Let me ignore, for the moment, the idea that self-knowledge enables us “to initiate activities aimed at testing [the] possibility [that some belief we have is wrong]”. Shoemaker claims that if a person has inconsistent beliefs and becomes aware of the inconsistency of the propositions believed, and if, furthermore, she is not aware of

³ The *non sequitur* from (5) to (6) is my suggested reconstruction of the argument in the quoted passage from Shoemaker. Other, more charitable, reconstructions might be possible.

⁴ I am suggesting that knowledge that one has reason to believe that *p* is not, in and of itself, knowledge that one believes that *p*.

having these inconsistent beliefs, then she should think that the facts themselves are inconsistent. Since this conclusion is absurd and since we are able to recognise inconsistency among propositions we believe, we must have self-knowledge of these beliefs.⁵

However, this argument seems to prove too much. Consider a dog expecting food when it hears some specific noise, say the rustling of the food bag. Let us assume that the dog's expectations are then frustrated. Does this mean that the dog takes the facts themselves to be inconsistent? On the other hand, does it mean that it has self-knowledge of its beliefs, thus being able to change them? Shoemaker seems to be saying that a person cannot change her beliefs unless she knows that she has them. But more plausibly, if, say, I perceive that the key is not in the drawer, where I thought it to be, I simply lose my earlier belief and form a new one to the effect that it is not there.⁶ It is not clear why this change of belief should involve the self-ascription of beliefs. In general, it does not seem necessary to self-ascribe beliefs for asking oneself whether things really are as they have presented themselves to be until now. I can ask myself "Is the key really in the drawer?", check the drawer and then form a new belief about the key. If I find that the key is not in the drawer, I can infer that, earlier, I had a false belief. However, this inference seems to be optional, not a necessary precondition of rationally changing one's beliefs.

In his fourth argument, Shoemaker asks himself what the situation would be like if we reasoned without self-knowledge and claims that such a view would fail to account for the idea that we are epistemic *agents*:

[O]ne might try to view deliberation about what to believe as a battle between contending beliefs or inclinations to believe. If this were right, it would seem unnecessary that the deliberator should have knowledge of the contending beliefs and desires; he would merely be the subject of them, and the battleground on which the struggle between them takes place. But this model seems hopelessly unrealistic, in part because it leaves out entirely the role of the person as an agent in deliberation; it represents deliberation as something

⁵ Louis Gallois (1996) employs a similar line of reasoning with the aim of justifying an Evansian procedure for self-ascribing beliefs, where we infer from the fact that *p* that we believe that *p*.

⁶ Why do I lose my earlier belief? I do not know. Thankfully, that is just how our belief-formation usually works. I lose my earlier standing belief that *p*, as soon as I form a new belief that not *p* in response to some new perception or other reasoning.

that happens in a person, rather than as an intentional activity on the part of the person. (Shoemaker 1988, p. 186)

Suppose we accept this argument. As Shoemaker remarks, more argument is needed to show that being “an agent in deliberation [...] essentially involves self-knowledge” (Shoemaker 1988, p. 186).⁷ However, it is also not so clear to me that the conception of the deliberator as the “battleground on which the struggle between [contending beliefs and desires] takes place” is completely misguided. In deliberation, we might intentionally direct our attention towards what we take to be facts and evidential relations. However, the crucial insights that change our beliefs do not seem to be intentional. That is why we have the perceptual metaphor of “insight”. Such “perception” might well be seen as something that happens to us, not something that can be intentionally formed or meaningfully “decided upon”. If this is correct, then it is the idea of agency in deliberation that is problematic, not the idea of the deliberator as a battleground of competing beliefs and desires.⁸

We have seen that Shoemaker proposes a number of different arguments to the effect that rational deliberation about what to believe requires knowledge of one’s own beliefs. All of these arguments can be doubted with good reason. Although we do have self-knowledge of our own beliefs, it is not clear that this is *necessary* for rational deliberation.

However, even if the arguments were successful in their own right, it seems to me that they would still contribute only very little to explaining the problem of self-knowledge as I have characterised it earlier. This is because Shoemaker’s arguments at best only show that we have self-knowledge. Given that we can rationally deliberate about what to believe, Shoemaker tries to show that we must have self-knowledge. However, his arguments do not tell us how such self-knowledge is *possible*.⁹ In fact, it seems to me that they do not even show

⁷ Shoemaker’s former student Richard Moran (2001) might be seen as developing this line of thought.

⁸ A fuller discussion of this point would have to give some account of how to understand our practice of *evaluating* a reasoner’s deliberation, without invoking the notion of agency. If the perceptual analogy is correct, blaming or praising a deliberator might be similar to blaming or praising someone for perceiving or failing to perceive something.

⁹ Was this a foregone conclusion? Why should the fact that self-knowledge is necessary contribute to an explanation of its possibility? Stoneham (1998), for example, gives an argument to the effect that self-ascriptions of belief are necessarily true – roughly, “I believe that p” must have at least all the consequences that “p” has, otherwise it would constitute a reason for changing the first-order belief and thus

that we must have self-knowledge of the kind described earlier. That is, they do not show that we must have a privileged and peculiar access to our own beliefs. Further argument would be needed to demonstrate that the self-knowledge supposedly necessary for deliberation is of the special kind described earlier.

3. Shoemaker's "zany argument": reasoning from "p" to "I believe that p"

The arguments examined so far aim to show that one cannot reason rationally in the way we do without having self-knowledge. However, Shoemaker also proposes a different type of argument not to the effect that we *must* have self-knowledge. Instead, this "zany argument" (Shoemaker 2009, p. 37) aims to show how a subject with normal cognitive capacities and dependent on rational cooperation with other subjects can reason her way from first-order beliefs to true self-ascriptions of these beliefs. Thus, self-knowledge is not shown to be necessary for rational deliberation, but shown to be a possible outcome of it. The idea is that someone with normal cognitive abilities, but – yet – without self-knowledge, could reason as follows:

p. Since *p* is true, it will, *ceteris paribus*, be in the interest of anyone to act on the assumption that *p*, if one is in circumstances (call these relevant circumstances) in which whether one so acts is likely to affect the satisfaction of one's interests. To act on the assumption that *p* is to act as if one believes that *p*. And part of acting as if one believes that *p* is acting in ways that indicate to others that one believes that *p*; for given that *p* is true, it will be in anyone's interest to act this way in relevant circumstances. So acting will help one enlist the aid of others who believe that *p* in the pursuit of one's goals. Others who believe that *p*, and share one's goals, will cooperate with one in ventures undertaken on the assumption that *p*, and since *p* is true such ventures will tend to be successful. Acting in ways that indicate

be incompatible with it, which is absurd. This account, if true, could also contribute to understanding how such a self-ascription can enjoy peculiarity and privilege: it does not matter how the self-ascription comes about (peculiarity); it necessarily makes itself true anyway (privilege). This is not an account to be discussed here; but it serves to show that a theory that claims necessity for self-knowledge might well be thought to contribute to a better comprehension of the problem of self-knowledge as I have defined it at the outset. Alas, this is not the case with the four arguments of Shoemaker's discussed here to the effect that reasoning requires self-knowledge.

to others that one believes that p will include saying, in appropriate circumstances, that one believes that p . Since this applies to everyone, it applies to me. And since I am in appropriate circumstances, I should say that I believe that p . (Shoemaker 2009, p. 37)

As a result of going through this reasoning process, the subject will truthfully self-ascribe the belief that p and generally behave as if she believes that she believes that p . Everything will be as if she has self-knowledge and Shoemaker suggests that we should regard the subject as having self-knowledge.

If we accept this, we will certainly have an explanation of at least part of the peculiar character of self-knowledge. The reasoning process Shoemaker describes is available only for first-person ascriptions of belief. If p and if I am in relevant circumstances, I should say that I believe that p , because this is likely to be helpful in the pursuit of my goals. However, it will not *per se* be helpful also to say that *some other person* believes that p . The other person might not have access to the information that p and if she does not, it will be in the interest of everyone to act on the assumption that she does not. What the other person believes or not will have to be judged on the basis of observing her. By contrast, Shoemaker's reasoning process does not rely on outward observation of oneself. Rather, it goes, only with the help of some general pragmatic considerations, directly from the first-order fact that p to the self-ascription of the belief that p . This explains peculiar access; the possibility of self-ascribing beliefs without recurring to the kind of evidence we would need for ascribing beliefs from a third-person perspective.¹⁰

Next, we can ask whether the "zany argument" also explains privileged access. Is a self-ascription arrived at by the reasoning process Shoemaker describes more likely to constitute self-knowledge than an ascription of belief based on third-person evidence? It seems that more argument is needed to establish such a claim. It could be pointed out, for example, that Shoemaker's reasoning process does not rely on perception. Therefore, it cannot go wrong in the way perceptually based ascriptions of belief can go wrong. Furthermore, it seems that the process does not depend on getting the facts about the world right. It might be

¹⁰ One might hold that Shoemaker only explains *part* of what I have described as peculiar access because one might think that the access he describes is not sufficiently "direct", proceeding, as it does, by way of an inference. However, Shoemaker certainly provides a reason for thinking that it is peculiar in the sense of being different from our access to other people's mental states. This is at least one important aspect of peculiar access.

that the subject is mistaken in thinking that p . Still, her going through Shoemaker's process would produce a true self-ascription of belief. It might also be that the subject judges wrongly that the circumstances are "relevant". They might not be in the sense that it does not matter in the circumstances whether anyone believes that p . None of this affects the truth of the resulting self-ascription of belief. We might say that the truth of such self-ascriptions of belief depends only on the subject's capacity to hold on to her first-order belief content " p " and to redeploy it in the subsequent self-ascription (cf. Fricke 2009).

However, the "zany argument" faces some serious criticisms. To look at two of them, let me distinguish the following steps in Shoemaker's argument:

- (1) p
- (2) In relevant circumstances, (i.e. in circumstances where it matters for the success of one's actions whether or not p is true), it is in the interest of anyone to act on the assumption that p .
- (3) To act on the assumption that p is to act as if one believes that p .
- (4) Part of acting as if one believes that p is acting in ways that indicate to others that one believes that p .
- (5) Acting in ways that indicate to others that one believes that p includes saying, in the appropriate circumstances, that one believes that p .
- (6) Since I am in the appropriate circumstances, I should say that I believe that p .

My first worry is that proposition (3) is doubtful. Acting on the assumption that p is not the same as acting as if one believes that p , especially if this is supposed to include saying that one believes that p . For instance, it might be expedient to act on the assumption that a patient is infected with Ebola, where this means that she is put into the isolation unit of a hospital. Yet, one might not actually *believe* that she has Ebola, perhaps because the relevant tests have not yet been made. Here, it might be important both to act on the assumption that she is infected with Ebola *and* to make clear to others that one believes neither that she is nor that she is not. On the other hand, it might be that it is unlikely that the patient has Ebola, yet it is important, in the circumstances, to act on the assumption that she has it to make sure that there is no risk of further infection in the unlikely event that she does. Again, it might

be important to explain to others that one does not believe her to have Ebola, that one's actions are only a precautionary measure.¹¹

A second objection might be that even though it is a simple chain of reasoning, it is still far too complex to be thought to describe how we actually come to our self-ascriptions of belief. Some might even say that we make self-ascriptions without going through any prior reasoning. In reply to this objection, Shoemaker says that he does not hold "that we come to our self-ascriptions of belief by employing the zany argument. We do not come to them by employing any sort of reasoning." (Shoemaker 2009, p. 38) Shoemaker's idea is that the "zany argument" only serves to "rationalise" our self-ascriptions of belief. This means that the subject does not actually go through the process of reasoning that supposedly leads to the self-ascriptions of belief. However, she is disposed to make such self-ascriptions on the basis of the premises of the "zany argument" and if she did go through the argument, she would arrive at the self-ascriptions.

It seems to me that the idea of merely "rationalising" our dispositions (and capacities) for reasoning is problematic without further explanation. If we do not actually employ the reasoning process Shoemaker describes, how does it explain anything about our self-ascriptions? Is the idea that at some earlier time we did use it and that, with time, it became automatic? On the other hand, is the idea that because it could be used, such dispositions are rational, hence useful and therefore became evolutionarily selected for? Without some such further explanation, the claim that an argument merely rationalises some behaviour, dispositions or capacities, but is not actually employed by the subject seems dubious to me.

Perhaps the most serious objection is that the argument does not seem to give the subject self-*knowledge*. It leads to the subject making a self-ascription only in the sense of *saying* "I believe that *p*", not in the sense of *coming to believe* that she believes that *p*. Thanks to the argument, the subject knows that, given the goals she pursues, it is expedient to affirm that she believes that *p*. However, this does not amount to genuine knowledge about herself. Knowledge about one's beliefs is more than just being able to make a true utterance to the

¹¹ As a referee has pointed out to me, the Ebola case is not analogous with Shoemaker's in that Shoemaker takes it to be a premise that *p*. So unlike in the Ebola case, there is no scenario where it could turn out that not *p*. This does not mean that my criticism of (3) is wrong, but it suggests that Shoemaker could avoid it by formulating his argument more carefully. Perhaps he should replace "act on the assumption that *p*" with "act such as to take into account that *p*".

effect that one has some belief. It is plausible, for example, that to know that one believes that p one must also *believe* that one believes that p . Yet, Shoemaker's reasoning does not seem to lead to such second-order belief. The subject just comes to believe that she should *say* that she believes that p , not that she actually believes it. Perhaps Shoemaker is right in affirming that the subject's behaviour, if guided by the "zany argument" is indistinguishable from ours with regard to expressions of self-knowledge. However, we might ask, as Byrne does in discussing a different version of the argument, "Why hasn't Shoemaker just outlined a strategy for *faking* self-knowledge?" (Byrne 2005, p. 92)

Let me briefly look at an earlier version of Shoemaker's argument. In "On Knowing One's Own Mind" (Shoemaker 1988), he points out that a subject with normal cognitive and conceptual abilities should be able to recognise the awkwardness of uttering Moore-paradoxical sentences, i.e. sentences of the form " p , but I don't believe that p ". Shoemaker characterises this awkwardness in pragmatic terms. Assertions, if sincere, express beliefs. Therefore, the first conjunct of the Moorean sentence, if asserted sincerely, expresses the subject's belief that p . However, the second conjunct says that the subject does not have this belief. Therefore, the first conjunct expresses a belief whose existence refutes what the second conjunct says. This means that one could not hope to get one's audience to accept the first conjunct as an expression of one's belief and the second conjunct as true. However, it is the pragmatic purpose of assertions to be taken both as true and as expressions of the utterer's beliefs. Since Moore-paradoxical sentences cannot be taken both as true and as expressions of the subject's belief, asserting them is self-defeating. All this should be evident to anyone with normal cognitive and conceptual capacities. Therefore, anyone with such capacities should both be puzzled when hearing others utter Moore-paradoxical sentences and avoid uttering such sentences herself. However, avoiding Moore-paradoxical sentences, while still making assertions about the world, can lead directly to self-ascriptions of belief. The reasoning could go like this:

- (1) p
- (2) In relevant circumstances (where it matters for the success of my actions whether or not p is true), I should say/assent to " p ".
- (3) One should never assert a sentence of the form " x , but I don't believe that x ".
- (4) In relevant circumstances, I must not assert " p , but I don't believe that p ".

- (5) In relevant circumstances, I must not assert “I don’t believe that p ”.
- (6) In relevant circumstances, I must not deny “I believe that p ”.
- (7) In relevant circumstances, I should say/assent to “I believe that p ”.

The first premise and conclusion of this argument are the same as in the earlier one. However, where the earlier argument said, in essence, “If p and if this helps to advance the realisation of your goals, say ‘I believe that p ’”, the argument from Moore’s Paradox gives the subject a more specific reason for assenting to “I believe that p ”. The reason is that if you are disposed to assert that p , then it is pragmatically self-defeating not to be prepared to assert, “I believe that p ” as well. Unlike in the first argument, here the assertion “I believe that p ” is not just a trick to get others to believe something that helps me along with my goals. Rather, it is an utterance that I have to be prepared to make in order to be coherent in what I say. Therefore, I think the argument from Moore’s Paradox is somewhat more convincing. However, it is zany all the same in that it does not lead to genuine self-*knowledge*, but only to a belief about what I should be prepared to *say*.¹²

We have seen that Shoemaker’s “zany argument” would explain, if successful, how peculiar and privileged access could go together in self-knowledge. The argument connects the premise “ p ” and the conclusion “I believe that p ” with the help of pragmatic considerations about how to further one’s goals or avoid incoherence in one’s utterances. As the argument is unsuccessful, it might be useful to examine so-called transparency theories of mind, which attempt to connect the two propositions in a more direct way.

4. Byrne’s transparency theory: inferring “I believe that p ” from “ p ”

Transparency theories of self-knowledge (especially those that focus on knowledge of one’s own beliefs) take their inspiration from a remark of Gareth Evans’s:

[I]n making a self-ascription of belief, one’s eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone

¹² I have discussed the argument from Moore’s Paradox in somewhat greater detail in Fricke (2012).

asks me ‘Do you think there is going to be a third world war?’, I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question ‘Will there be a third world war?’ I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure I have for answering the question whether p . (Evans 1982, p. 225)

Evans observes here that we do not inspect inner items and examine what contents they have when we try to answer questions about our beliefs. Rather, we think about the world and our thinking about the world reveals to us what it is that we believe. Therefore, we might say that self-knowledge is transparent to a consideration of the world or, as Richard Moran puts it:

[A] first-person present-tense question about one’s belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world. (Moran 2001, p. 62)

So how could the idea of transparency help us to explain the problem of self-knowledge? In what follows I shall briefly look at the theory proposed by Alex Byrne.

Byrne interprets Evans as saying that we come to our self-ascriptions of belief through a process of reasoning. We infer from the fact that p that we believe that p . Such reasoning can be interpreted as following an epistemic rule that has this form:

(BEL) If p , believe that you believe that p (Byrne 2005, p. 94)

Byrne stipulates that we follow an epistemic rule of the form ‘If conditions C obtain, believe that p ’ if and only if we believe that p because we recognise that conditions C obtain. Therefore, following (BEL) requires recognising that p and inferring from this fact that one believes that p . Byrne also uses Gallois’s “doxastic schema” to describe the rule for this inference:

$$\frac{p}{\text{So, I believe } p}$$

(Gallois 1996, p. 46; cf. also Byrne 2011b, p. 204f.)

It is immediately obvious that such an inference is not valid. The premise does not even make the conclusion particularly probable. So why should it be thought that this epistemic rule/inference schema can help us explain self-knowledge?

Byrne's general attitude to this question seems to be that the success of the rule in bringing about true self-ascriptions of mental states is sufficient to justify its peculiarities. Using the rule requires recognising that p . However, recognising that p is, or involves, forming or retrieving the belief that p . So if someone recognises that the antecedent conditions of the rule obtain, i.e. if she recognises that p , then she believes that p . It follows that the self-ascription which results from an application of the rule is true.

Does the rule explain the peculiarity of our access to our own beliefs? It does, because applying the rule does not require observing oneself or otherwise gathering evidence about oneself from which one could conclude what one's beliefs are. It can do without most of the elements that would make our ascriptions of beliefs to other people reliable. It just involves a direct inference from a fact that does not even need to be about oneself to a self-ascription of belief. Moreover, it is a rule only for oneself and one's own beliefs. It does not yield knowledge of other people's beliefs.¹³

Does the rule also explain why we are privileged in our access to our own beliefs? We have already seen that correctly applying the rule requires recognising that p and that this entails that the resulting self-ascription is true. There is no equivalent rule for ascribing beliefs to other people. Consider (BEL-3):

(BEL-3) If p , believe that Fred believes that p (Byrne 2005, p. 96)

Even if I correctly apply this rule, i.e. I recognise that p , it is far from guaranteed that the resulting belief ascription turns out correct. I would not dismiss (BEL-3) completely (classify it as a bad rule, as Byrne seems inclined to do), because it expresses a kind of charitable default assumption that other people, largely, have the same beliefs as I do. However, it is clear that (BEL-3), even if applied correctly, will often result in false ascriptions of belief to Fred. Of course, in general, ascriptions of belief to other people often rely on observation, theories

¹³ Perhaps using (BEL) is not a direct or immediate way of acquiring knowledge of one's own beliefs. However, not relying on perception, it does seem to be more direct than the acquisition of knowledge of other people's beliefs.

and, possibly, more complex inferences. By contrast, (BEL) does not require observation, theoretical knowledge or substantive inferences. As far as it does not, there are fewer possibilities for self-ascriptions of belief based on (BEL) to go wrong.

Byrne points out, furthermore, that I do not even have to apply (BEL) correctly in order to arrive at a true belief-ascription. I might be wrong in thinking that p . Therefore, I do not actually satisfy the antecedent condition of the epistemic rule. If I still use the rule to form the belief that I believe that p , Byrne says that I have merely *tried* to follow (BEL), but failed to recognise an antecedent condition. However, even merely trying to follow (BEL) in this sense still produces true self-ascriptions of belief. Even if I am mistaken in thinking that p , if on the basis of this supposed fact I conclude that I believe that p I will be making a true self-ascription of a false belief. Byrne describes epistemic rules where this is the case as “strongly self-verifying”.¹⁴

Finally, we might also point out that the only capacity necessary in order to arrive at a true self-ascription of belief using (BEL) is to be able to hold onto and redeploy a thought content. We must not let our thought content “ p ” transform into some other content when we prefix it with “I believe that”. Other than this capacity, we do not need any further perceptual or inferential skills (cf. Fricke 2009).

We should note that if (BEL) correctly describes our method for self-ascribing beliefs, then it is possible to have beliefs without knowing that one has them. (BEL) might simply not have been applied to them yet. It should also be possible for self-ascriptions to fail. (BEL) is simple and not subject to errors that can occur in other-ascriptions of belief. However, this does not mean that there is no room for error at all. For whatever reason, the thought content of the first-order belief “ p ” might be corrupted on its way to being redeployed in a self-ascription of belief, which would then turn out to be false.¹⁵ (That such false self-ascriptions might in turn have a habit of bringing about the first-order belief that

¹⁴ The following is an epistemic rule that is only self-verifying, but not strongly so: “(KNOW) If p , believe that you know that p ”. Following (KNOW) and correctly recognising that p makes it true that one knows that p . However, merely trying to follow (KNOW), where one fails to recognise a true fact that p , will not lead to true self-ascriptions of knowledge. (Cf. e.g. Byrne 2011b, p. 206)

¹⁵ In the doxastic schema, the subject would start with the premise p , but then draw the false conclusion “I believe that q ”. She fails to hold on to the content of her premise. This might seem contrived, especially for simple propositions. However, it is hard to see why it should be impossible. It would represent a failure of a capacity basic for drawing inferences and, in this sense, a failure of rationality. (Cf. Fricke 2009, p. 8)

they falsely ascribe and thereby a habit of becoming true after all, is another story.¹⁶)

A further remark is that an account of self-knowledge based on (BEL) does not stipulate special faculties, such as an inner sense. Ordinary reasoning capacities in combination with (BEL) should be enough to produce self-knowledge. In this respect, the account avoids making claims that Byrne labels as “extravagant”; instead, it is “economical”. If correct, certain partial cognitive failures, such as a failing uniquely of self-knowledge, because some inner sense ceases working, while all other cognitive capacities remain functioning, would seem to be unlikely.¹⁷

I should like to mention two interesting criticisms of Byrne’s theory. First Matthey Boyle has characterised the inference from “*p*” to “I believe that *p*”, which is implicit in Byrne’s epistemic rule (BEL), as “mad” (Boyle 2011, p. 230):

To believe that I believe *P* is to hold it *true* that I believe *P*. Being a reflective person, I can ask myself what grounds I have for holding this true. The answer ‘*P*’ is obviously irrelevant. I am asking what shows that the proposition *I believe P* is true, and a modicum of rational insight will inform me that, even if it is true that *P*, this by itself has no tendency to show that I believe it. (Boyle 2011, p. 230)

Boyle’s view is that an inference is not just “a reliable tendency to pass from one believed content to another” (Boyle 2011, p. 231). Rather, the premises of the inference must in some sense *show* the truth of the conclusion. Clearly, this is not the case when we pass from “*p*” to “I

¹⁶ Cf. Stoneham (1998) for a theory according to which sincere self-ascriptions of belief necessarily make themselves true. His theory lacks an account of how these self-ascriptions come about in such a way as to cohere with our pre-existing beliefs. In this, it might be complemented by a transparency theory of self-knowledge.

¹⁷ As I said in the introduction, throughout this text I concentrate on the central case of knowledge of one’s own beliefs. I should mention, however, that Byrne makes a considerable effort to generalise his account to cover knowledge of one’s own intentions (2011b), visual perceptions (2012a), desires (2012b) and current thoughts (2011a). In all these cases, Byrne suggests that there are inferential schemas (or epistemic rules) similar to (BEL) that enable us to acquire self-knowledge in a peculiar and privileged way. However, this is not to say that in each case the privilege (the comparative error-proofness) exists to the same degree as in the case of knowledge of belief. Byrne acknowledges that the inferential schemas for acquiring knowledge of one’s own intentions and desires, for example, are defeasible by additional knowledge in ways that (BEL) is not.

believe that p ". One consequence of this view is that even if I followed the epistemic rule (BEL) and thus formed a self-ascription of belief, I should – if I subsequently reflect on the self-ascription – abandon the self-ascription for not being well supported:

[A] belief, once formed, doesn't just sit there like a stone. What I believe is what I hold true, and to hold something true is to be in a sustained condition of finding persuasive a certain view about what is the case. Even if we grant that a disposition to pass from one content to another could deposit various arbitrary beliefs in my mind, those beliefs would be unsustainable if I, understanding their contents, could see no reasonable basis for holding them true. (Boyle 2011, p. 231)

This criticism puts a strong internalist constraint on the formation of second-order beliefs. If there is a basis on which a rational person forms a belief about her beliefs, then it must be a basis that shows *to the subject* that the second-order belief is true.

Boyle's own account shows that it is difficult to find such a basis. His solution is the claim that first-order beliefs from the very beginning go together with second-order knowledge that we have them. Such second-order knowledge is normally tacit. When we reflect on our first-order states, it can become explicit in a self-ascription of belief. It is not clear whether the account proposed can avoid an infinite regress of ever higher-order tacit knowledge.

What are we to make of Boyle's criticism? On the one hand, it might be responded that Boyle's internalist constraint seems too strong, that we do not necessarily discard a belief that we find ourselves with just because we have nothing that *shows* its truth. On the other hand, it might also be responded that we do have an understanding of why (BEL) produces true self-ascriptions of belief: when something turns out to be true (in my view) then I form the belief that it is so. As far as I am rational, my finding something true is what makes me believe it. Therefore, the inference from what is true (when I answer this question) is bound to reveal what I believe.

The second objection to Byrne's account is that his method for producing self-ascriptions of belief does not *reveal* first-order beliefs, but rather *creates* new first-order judgments. So even if the resulting self-ascriptions of belief turn out to be true, they are so because we have made them so following (BEL), not because we have had the beliefs now self-ascribed all along.

Brie Gertler (2011) has developed this objection in some detail. She points out, for example, that I may not have a belief as to whether p . If I now employ (BEL) to determine whether I believe that p or not, I will ask myself whether p is true. Here I might well come to form a belief as to whether p . Suppose I form the belief that p . Following (BEL) I will then proceed to the self-ascription “I believe that p ”. This self-ascription of belief will be true. However, it is true only because the first-order belief was formed in response to the application of (BEL). (BEL) cannot reveal the absence of belief.

There might also be a problem with current beliefs. Suppose I try to use (BEL) to find out whether I currently believe that p . I proceed by asking myself whether p , find that it is indeed the case that p and then correctly self-ascribe this belief. The problem is that my asking myself whether p might be what prompts me, perhaps after a new examination of my available evidence, to form the belief that p . This might be so even though before applying (BEL) I did not believe that p or I believed that not p . Therefore, although the resulting self-ascription of belief is true, it does not reveal or detect a pre-existing current belief. Rather, it changes it, creates a new belief and self-ascribes it.

Suppose we wish to know whether we have a dispositional belief that p . Gertler points out that if we want to find out by using (BEL), we must take care not to change the dispositional belief by bringing in new evidence. Applying (BEL), we ask ourselves whether p . If we produce an answer by looking outward at the evidence we have *now*, we might come to a different answer than what we already believe dispositionally. Therefore, the application of (BEL) does not reveal the dispositional belief unless we restrain ourselves from taking into account new evidence. If we thus only rely on what we already have in the mind, the “internal”, our answer to the question of whether p might reflect our dispositional beliefs. However, it is only by concentrating on the internal and blocking external influences out that (BEL) can be made to reveal dispositional beliefs. This means that the application of (BEL) presupposes some sort of self-knowledge, namely regarding the distinction between what is “internally believed” and what is true in light of “external” influences, rather than just bringing self-knowledge about.

How serious is Gertler’s objection? It is clearly true that asking oneself whether some proposition is true can stimulate one to form a belief that one did not have before or to examine one’s evidence anew and change beliefs that one had until then. Therefore, if the application of (BEL) always requires asking oneself whether p is true, then it clearly is not a good method for revealing antecedent beliefs.

However, it is not so clear to me that we necessarily have to conceive of the application of (BEL) in this way. If we have (BEL) at our disposal, this might simply mean that we have a tendency to infer from facts that we come across that we believe them. In other words, for (BEL) to be operative it is not necessary that one goes through a process of *asking* oneself some first-order question. It is sufficient for one simply to *have* a first-order belief. Having the first-order belief and having (BEL) as an epistemic rule at one's disposal, one might then be disposed to self-ascribe the first-order belief by applying (BEL). Here, there is no prompting of an answer to a first-order question. It is simply that an inferential routine is available for execution. If we have self-knowledge that has not been acquired in response to some questioning (be it of first-order or second-order matters), then that might be because we went through the inferential routine described by (BEL) when a first-order belief was available for it. It seems that there is no reason to think that such self-knowledge could not be of antecedently existing current or dispositional beliefs. – However, even such unprompted and unprompting use of (BEL) does not provide one with self-knowledge regarding the absence of beliefs. If we know that we do *not* have a belief on some subject, this knowledge must have come about through some method other than (BEL).

5. Conclusion

I have looked at three different conceptions of the relation between reasoning and self-knowledge. On the first conception, self-knowledge is a precondition for a specific kind of reasoning, namely for rational adjustment of one's beliefs in light of new evidence or newly discovered relations of coherence. I have tried to show that the arguments for this kind of claim are inconclusive. It might well be possible rationally to adjust one's beliefs in the way we do without ever self-ascribing them.

According to the second and third conception, we can reason from first-order beliefs to second-order beliefs. Shoemaker's "zany argument" says that a subject who finds that p will, in relevant circumstances, also find it useful, for the realisation of her goals, to self-ascribe the belief that p . In another, more convincing, version, the argument says that a subject who finds that p and is disposed to assent to " p ", will realise that her own assertion will only be coherent if she is also disposed to assent to "I believe that p ". Both versions of the argument fail to bring about self-*knowledge* in the subject.

The third conception of the relation between reasoning and self-knowledge claims that we might assume that we have an epistemic rule

at our disposal that permits a direct inference from first-order beliefs to self-ascriptions of these beliefs. Self-knowledge is then the product of reasoning with the help of this epistemic rule.

The first conception, even if it were true, would not contribute significantly to an explanation of the two characteristic and problematic features of self-knowledge: privileged and peculiar access. At best, it only establishes that we have self-knowledge; not how it is possible. The second and third conception do allow for an explanation of the two features of self-knowledge. On balance, and despite two interesting objections, the third conception seems to offer a more successful account of the problem of self-knowledge.

References

- Boyle, M. (2011), "Transparent Self-Knowledge", *Proceedings of the Aristotelian Society Supplementary*, 85 (1), pp. 223-241.
- Byrne, A. (2005), "Introspection", *Philosophical Topics*, 33 (1), pp. 79-104.
- Byrne, A. (2011a), "Knowing that I am Thinking", in Hatzimoysis, A. (ed.), *Self-Knowledge*, Oxford, Oxford University Press, pp. 105-123.
- Byrne, A. (2011b), "Transparency, Belief, Intention", *Proceedings of the Aristotelian Society Supplementary*, 85 (1), pp. 201-221.
- Byrne, A. (2012a), "Knowing What I See", in Smithies, D. and Stoljar, D. (eds.), *Introspection and Consciousness*, Oxford, Oxford University Press, pp. 183-209.
- Byrne, A. (2012b), "Knowing What I Want", in Liu, J. and Perry, J. (eds.), *Consciousness and the Self: New Essays*, Cambridge, Cambridge University Press, pp. 165-183.
- Evans, G. (1982), *The Varieties of Reference*, Oxford, Clarendon.
- Fricke, M. F. (2009), "Evans and First Person Authority", *Abstracta*, 5 (1), pp. 3-15.
- Fricke, M. F. (2012), "Rules of Language and First Person Authority", *Polish Journal of Philosophy*, 6 (2), pp. 15-32.
- Gallois, A. (1996), *The World Without, the Mind Within: An Essay on First-Person Authority*, Cambridge, Cambridge University Press.
- Gertler, B. (2011), "Self-Knowledge and the Transparency of Belief", in Hatzimoysis, A. (ed.), *Self-Knowledge*, Oxford, Oxford University Press, pp. 125-145.
- Moran, R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, Princeton University Press.
- Shoemaker, S. (1988), "On Knowing One's Own Mind", *Philosophical Perspectives*, 2 (Epistemology), pp. 183-209.

- Shoemaker, S. (1994), "Self-Knowledge and 'Inner Sense'. Lecture II: The Broad Perceptual Model", *Philosophy and Phenomenological Research*, 54, pp. 271-290.
- Shoemaker, S. (2003), "Moran on Self-Knowledge", *European Journal of Philosophy*, 11, pp. 391-401.
- Shoemaker, S. (2009), "Self-Intimation and Second Order Belief", *Erkenntnis*, 71, pp. 35-51.
- Stoneham, T. (1998), "On Believing that I am Thinking", *Proceedings of the Aristotelian Society*, 98, pp. 125-144.

Received: March 2, 2016; revised: July 1, 2017; accepted: September 11, 2017.