

Rationality and Acquaintance in Theories of Introspection*

Daniel Stoljar, ANU

Abstract: According to a rationalist theory of introspection, rational agents have a capacity to believe they are in conscious states when they are in them, much as they have the capacity, for example, to avoid obvious contradictions in their beliefs. For the agent to know or believe by introspection, on this view, is for them to exercise that capacity. According to an acquaintance theory of introspection, by contrast, whenever an agent is in a conscious state, the agent is aware of or is acquainted with being in the state, where the background notion of acquaintance is understood to be distinct from sense perception, on the one hand, and belief or knowledge, on the other. For the agent to know or believe by introspection, on this view, is for them to take advantage of the epistemic position they occupy in virtue of being acquainted in this way. These theories are not in conflict; it is possible to hold both an acquaintance theory and a rationalist theory. This paper, however, sets out and recommends a different possibility, namely, that of holding a rationalist theory while rejecting any acquaintance theory.

1. Introduction

In contemporary literature, introspection is often defined minimally as that distinctive way that we come to know or rationally believe that we are in conscious states when we are in them (Dretske 2003). Different theories of introspection may then be understood as different proposals about how to fill out or understand the way in question. This paper contrasts two such theories and suggests that one is more attractive than the other.

The first theory, which I will call the *rationality-based* theory, or the *rationalist* theory for short, focuses initially on rational agents. It says (roughly) that rational agents have a capacity to believe that they are in conscious states when they are in them, much as they have the capacity, for example, to avoid obvious contradictions in their beliefs. For the agent to know or believe by introspection, on this view, is for them to exercise that capacity.¹

The second theory, which I will call the *inner awareness-* or *acquaintance-based* theory, or the *acquaintance* theory for short, focuses initially on conscious states.² It says (roughly) that whenever an agent is in a conscious state, the agent is acquainted with, or is aware of, being in the state, or perhaps with the state itself, where the background notion of

* I presented versions of this material to audiences at the “Locating Subjectivity in the Study of Consciousness at Ashoka University, at IUSS (Italy) Workshop on Analytic Phenomenology, Intentionality and Consciousness, and at the American Association of Philosophy (Eastern Division) in Montreal. I am very much indebted to all who took part on those discussions. I am especially grateful also for written comments from the editors of this volume, as well as Matt Duncan, Andrew Lee, Nicholas Silins, and Adriana Renero (who was the commentator on the paper in Montreal).

¹ I have defended a view like this in several publications; see, e.g., Stoljar 2012, 2018, 2019.

² I will use the phrases ‘acquaintance’ and ‘inner awareness’ interchangeably in this paper; in addition, unless stated otherwise by ‘awareness’ I mean ‘inner awareness’.

acquaintance or inner awareness is understood as distinct from sense perception, on the one hand, and belief or knowledge, on the other. For the agent to know or believe by introspection, on this view, is for them to take advantage of the epistemic position they are in as a result of being acquainted in this way.³

These theories are not in conflict. It is possible to hold both the acquaintance and the rationalist theory. But my focus here will be on a different possibility, namely, holding the rationalist theory while rejecting the acquaintance theory. This conjunction of commitments is attractive in several ways. First, it promises a more general approach to introspection than we otherwise would have. Claims about acquaintance are plausible only in the case of specific sorts of conscious states, namely, phenomenally conscious states (as they are usually called) as opposed to conscious states of other varieties; hence the acquaintance theory has a built-in, and I will argue problematic, restriction that the rationalist theory does not. Second, even in the case of phenomenal consciousness, an assumption of acquaintance has, as we will see, proved controversial. Hence my overall conclusion: the rationality-based theory is preferable to the acquaintance-based theory.

2. The Rationalist theory: Basic Idea

Let us begin by looking more carefully at the two theories under discussion. The rationalist theory starts from a picture of rationality that is familiar, though not entirely uncontroversial (Harman 1986, Lewis 1994, Broome 2013). On this picture, there are certain combinations of mental states that rational agents are in in so far as they are rational. For example, a rational agent will typically not have inconsistent beliefs, at least not obviously so; a rational agent will typically act to satisfy their desires in accordance with their beliefs; and a rational agent will typically form beliefs on the basis of their perceptions: if a rational agent perceives a fox in front of them, for example, they will typically believe that there is a fox in front of them.

A picture of this kind needs to be elaborated and qualified in various ways. I won't do that here; the point rather is to take the picture for granted and then understand the rationalist theory as proposing something similar in the case of conscious states, namely, that a rational agent who is in a conscious state will typically believe that they are in the state. Suppose you

³ A view of this general character is widely represented in the literature; see, e.g., Fumerton 2001, 2019, Kriegel 2009, Gertler 2011, 2012, Nida-Rümelin 2011, 2016, Strawson 2015, Montague 2016, Duncan 2021. But there are important differences among these writers, and I will focus solely here on the specific version of the position set out in section 5 below.

are a rational agent who perceives a fox and suppose moreover that your state of perceiving the fox is conscious—that is, you not only perceive a fox but do so consciously; then, according to the rationalist theory, you will typically believe that you perceive the fox.

It is useful to formulate the view in terms of the notion of a principle of rationality, and the associated idea that a rational agent abides by such principles. A candidate principle of rationality, for example, is a *no-contradiction principle*, according to which (to put it roughly) you should not have inconsistent beliefs, at any rate other things being equal and given background conditions. An agent who abides by this principle will typically avoid contradictions and will have a capacity to do so. From this point of view, the rationalist theory proposes what we might call an *introspection principle*, according to which (to put it roughly) you should believe you are in a conscious state if you are in one, again, other things being equal and given background conditions. An agent who abides by this principle will typically believe they are in psychological states that are conscious and will have the capacity to do so.

Earlier I said that introspection is often defined minimally as that distinctive way in which we come to know or rationally believe that we are in conscious states when we are. The sense in which introspection is distinctive is that it is different both from ordinary sensory perception and from inference. If you know by introspection (and only by introspection) that you perceive a fox, you don't know this by inference from something else that you know or believe. Nor do you know it by perception, which of course is not to deny that you may know many things about what you perceive (i.e., the fox) by perception. The rationalist theory is a theory of introspection because it provides an account of what this distinctive way is: to know or believe something by introspection is to exercise a capacity you have insofar as you are rational agent, a capacity that is at least in principle disjoint from those involved in perception or inference.

3. The Rationalist theory: Matters Arising

There are several issues that arise concerning the rationalist theory as so far introduced.

First, what notion of a conscious state is at issue in the introspection principle? Contemporary literature provides a variety of different possible accounts of a conscious state: phenomenal accounts, higher-order accounts, access accounts etc. An important point

about the rationalist theory is that it is ecumenical as regards these different conceptions.⁴ A common assumption, for example, is that what is uniquely at issue in discussions of introspection is the phenomenal conception of consciousness, on which a psychological state is conscious if and only if there is something it is like to be in it. The introspection principle is consistent with that assumption but does not require it; it permits that consciousness might be understood in addition or instead in other ways. For example, it is consistent with the idea that beliefs are conscious and so introspectable, even if they are not conscious in any phenomenal sense. As we will see, this point is important when we compare the rationalist theory with the acquaintance theory.

Second, what is the content of the belief that you form if you abide by the principle? For example, is it the belief *that you are in a psychological state*, a state which as a matter of fact is conscious; is it the belief *that you are in a conscious psychological state*; or is it the belief that *you are in this state* which as a matter of fact is a conscious psychological state? I will assume in what follows that the first belief is in play, but it is important to note that the others are available as well.

Third, what are the background conditions mentioned in the principle? What I have in mind are roughly the following: (a) the agent has an interest in whether they are in the state in question; that is, it somehow matters to them that they are; (b) the agent is psychologically capable of forming the belief that they are in the state; and (c) the agent has no evidence, or at least no overwhelming evidence, against the hypothesis that they are in the state. In many of the central cases of introspection, background conditions of this kind are in force, so there is no harm in assuming them when formulating the rationalist theory.

Fourth, why formulate the principle in terms of an ‘other things being equal’ clause? This allows for exceptional cases in which you are in a conscious state, the background conditions obtain, and yet it is not required by rationality that you believe you are in the state. If you consciously perceive a triangle in front of you, it may be that you can come to know by introspection that you do so; but if you perceive a chiliagon (Descartes’s famous example of the thousand-sided figure) you may not do so, even if you are rational and the

⁴ There is an exception to this ecumenical attitude, namely, we cannot allow that the notion of consciousness in the introspection principle is understood according to a version of the higher-order thought account that one finds in writers such as Armstrong (see Armstrong and Malcolm 1980) on which (to put it roughly) a psychological state is conscious if and only if the subject of the state knows or believes they are in the state by introspection; that is an exception because it defines the notion of a conscious state in terms of introspection rather than the other way around. I will set aside this account here, but for my own views about it, see Stoljar forthcoming.

background conditions are met—the reason being that you cannot tell in introspection whether you are perceiving a thousand-sided- rather than, say, a thousand-and-one-sided figure.

Fifth, what is it to abide by the introspection principle or indeed any principle of rationality? I assume that abiding by such principles is analogous to abiding by the syntactic principles of the language you learnt in childhood. In both cases, to abide by the principles is not merely to conform to them; that could happen by sheer chance. Nor is it to know the rules explicitly, i.e., in a way that allows you to articulate what they are, to attend to them etc. Rather, it is to know the rules implicitly, in a way that does not by itself allow you to articulate them or attend to them, even in principle. In such cases, you conform to the rules, to the extent that you do, because you know the rules in question in this implicit way.⁵

Sixth, it may seem that to abide by the introspection principle requires you to know that the antecedent of the principle is true, or that the antecedent of some instance of the principle is true; in particular, it might seem that to abide by the principle requires you to know that you perceive a fox. If so, the rationalist theory would be a non-starter, since it would presuppose exactly what it sets out to explain. But on the intended understanding of what it is to abide by the principles of rationality, this is not so.⁶ After all, to abide by the non-contradiction principle does not require you to know that you believe such and such; what it requires is that you know the rule, and that you conform to the rule (to the extent that you do) in virtue of doing so. The same is true in the case of the introspection principle.⁷

Seventh, the rationalist theory tells us in the first instance about rational agents, but what about ordinary agents like you and me? Rational agents are rational by definition, but it is a substantive assumption that we ordinary agents are. Can we make that assumption? My view is that in most cases we can make it—so long as the notion of rationality is not too

⁵ To suppose that rationality is a matter of abiding by certain principles leaves open whether the principles come first in the order of explanation or whether the capacities distinctive of rationality do so. On a principle-first view, a rational agent has the capacities they have because they abide by the principles; on a capacity-first view, a rational agent abides by the principles because they have the capacities—here the principles may be understood as abstractions from the capacities rather than the other way around. I will leave this issue open here.

⁶ It may be preferable to speak of representing or cognizing (in Chomsky's 1986 famous phrase) the rules rather than knowing them; I will set that aside. For a discussion of the analogy between linguistic competence and rationality, see Stein 1997; and for a similar analogy as regards morality, see Mikhail 2011.

⁷ Shoemaker (2009) and Colin McGinn (1982) both argue that introspective knowledge is required to abide by the rules of rationality, but I follow Moran 2001 in finding this argument implausible; see Stoljar 2021a.

idealized. Indeed, it is partly for this reason that the principles I have been discussing are all understood as other things being equal principles and as assuming background conditions.

4. Shoemaker and Armstrong

We may complete this sketch of the rationalist theory by contrasting it with two approaches to which it is at first glance quite similar, namely, Shoemaker's constitutivism and Armstrong's 'broad perceptual' theory (to use Shoemaker's name for it); see, e.g., (Armstrong and Malcolm 1984, Shoemaker 1996, Shoemaker 2009)

Shoemaker's position as I understand it has two main parts. The first part is a thesis Shoemaker calls *the impossibility of self-blindness*. According to this thesis, it is impossible in the widest sense (that is, impossible logically or metaphysically) that you are a rational agent, that you are in some conscious psychological state ψ , that you meet the background conditions mentioned above, and yet you do not form the introspective belief that you are in ψ . The second part is a view about what explains this impossibility. Shoemaker's idea is that what explains it is a view about the nature of the introspective belief that you are in ψ , namely, that to have this belief *just is* (a) to be in ψ ; (b) to be rational; and (c) to satisfy the background conditions. It is for this reason that Shoemaker's position is called 'constitutivism:' the introspective belief is constituted by these other elements, and this fact about it explains the impossibility of self-blindness.

The rationalist view I am recommending is self-consciously inspired by Shoemaker and so is like it in several ways. Like Shoemaker, it emphasizes the connection between introspection and rationality. Like Shoemaker, it accepts the impossibility of self-blindness, the first part of his view.⁸

But the rationalist theory is nevertheless different from Shoemaker's view as regards its second part. Instead of explaining the impossibility of self-blindness by appealing to the nature of introspective belief, it does so by appealing to the nature of rationality. On the rationalist theory, as we have seen, to be rational is among other things to abide by the introspection principle. Hence it is no wonder that if one is rational, and one is in a conscious state, one will believe that one is, at least given background conditions: the principle entails this will happen in the relevant circumstances.

⁸ Since the introspection principle is formulated with an 'other things being equal' clause, the notion of self-blindness that it will rule out as impossible must be formulated in a correlative way. I will ignore this complication.

This difference from Shoemaker means that the rationalist view avoids several problems that his own view confronts. One problem is that what he says about the introspective belief that you are in ψ is implausible. Since you can have a belief with this content without being rational, for example, it is hard to see that being rational could possibly be part of what constitutes that belief.

Another problem is why the introspective belief on Shoemaker's view is justified in the first place (Martin 1998). His view entails that having this belief is necessitated by other elements, including the fact that you are rational, but that by itself does not make the belief justified.

A third problem concerns Shoemaker's own understanding of rationality, on which the introspection principle is *not* one of the principles of rationality; on Shoemaker's view, as I understand it, rationality is limited to such standard principles as the no-contradiction principle, the means-end principle and so on. But this limitation makes it hard to see why self-blindness is impossible in the first place. For suppose you consciously perceive a fox, and are rational in the sense that you lack contradictory beliefs, act to satisfy your desires in accordance with your beliefs, and so on. It doesn't follow from this that you will believe you perceive a fox, even if the relevant background conditions are met.

As regards Armstrong's theory, we may understand this as supposing that there are empirical psychological laws that entail that if you are in a conscious state, you will believe that you are, other things being equal and given certain background conditions. If these laws apply to you, you will typically form the belief that you are in the conscious state if you are, and will have the capacity to do so.

The rationality-based theory I am interested in is logically like Armstrong's. The difference is that for us the 'laws' are principles of rationality whereas for Armstrong they are empirical generalizations.

But this difference makes all the difference, since the rationalist theory is in consequence not subject to the problems facing Armstrong. One problem is that his theory allows the possibility of self-blindness; this is Shoemaker's main argument against the broad perceptual theory. Another problem is that, contrary to Armstrong's position, the mere fact that a belief is lawfully correlated with what makes it true does not make the belief justified (Peacocke 1998, Zimmerman 2006).

We might summarize these remarks by noting that the rationalist theory may be viewed as offering a synthesis of Shoemaker and Armstrong while moving beyond both.

Like Armstrong's theory, it emphasizes the connection between introspection and laws—the difference is that for Armstrong, these are empirical laws. Like Shoemaker's theory, it emphasizes the connection between introspection and rationality—the difference is that it represents introspective belief as an exercise of a rational capacity rather than something that is merely necessitated by among other things the property of being rational.⁹

5. The Acquaintance Theory: Basic Idea

So that in outline is the rationalist theory of introspection; turning now to the acquaintance theory, this starts from a quite different place. The main idea of the theory—at least as I will understand it here—is that you are in a conscious state if and only if you are acquainted with being in the state. Hence if you perceive a fox, and that state is conscious, you are acquainted with your perceiving a fox.¹⁰

One might ask how this theory is a theory of introspection. On the surface, it is a theory about conscious states, and their relation to acquaintance, not about how we come to know or believe that we are in conscious states.

But the acquaintance theory may be developed by adding two subsidiary ideas to the main one just introduced. The first is that acquaintance has the following epistemic property. If you are acquainted with being in some state, you are in an epistemic position to form a justified belief that you are in the state, whether or not you do form that belief. The second idea is that if you are in an epistemic position to form this justified belief, you will form that belief, at least given the background conditions of the sort we discussed earlier: you are interested in whether the belief is true, you are capable of forming the belief, and you have no countervailing evidence.

Putting these ideas together yields a theory of introspection in the intended sense, i.e., a theory of the distinctive way in which you come to know that you are in conscious states when you are in them. On this theory, if you consciously perceive a fox, you will be

⁹ The rationalist theory is not only different from the views of Armstrong and Shoemaker, it is different also from so-called transparency views of the sort defended by (e.g.) Byrne 2018. The main difference with Byrne's view is the introspection principle is not a rule of inference; for discussion, see Stoljar 2019.

¹⁰ As noted in fn. 3, not every proponent of the acquaintance theory will develop their view in the way indicated in the text. One alternative, for example, which I take to be defended in Gertler 2010, 2012, is to say, not that you are in a conscious state if and only if you are acquainted with it, but only that you are in an *introspectable* conscious state if and only if you are acquainted with it. In my view, such a position is best classified as a version of what Shoemaker calls the 'object-perceptual model' and is subject to the criticisms he makes of that model, which I endorse; see Shoemaker 1996. However, I won't attempt to address this version of the view here.

acquainted with your perceiving a fox. If you are acquainted in this way, you are in an epistemic position to form a justified belief that you are perceiving a fox. Then given background assumptions, you will form the belief that you are doing so, and this belief will typically amount to knowledge.

6. The Acquaintance Theory: Matters Arising

Just as several questions arise for the rationalist theory, so too for the acquaintance theory.

First, what is the notion of consciousness at issue here? While the rationalist theory is mostly ecumenical, for the acquaintance theory it is consciousness in the phenomenal sense, rather than any other sense, that is in play. Martina Nida-Rümelin puts this particularly clearly. “To have an experiential property is to be aware of having it”, she writes, and later, “to be aware in the relevant sense of having experiential properties is to be aware of being phenomenally conscious” (2016, p. X); see also (Kriegel 2009, Nida-Rümelin 2011, Strawson 2015, Montague 2016)

Second, while the acquaintance theory concerns phenomenally conscious states, it nevertheless makes a substantive claim about such states rather than a trivial claim. As we have seen, on the phenomenal conception, a psychological state is conscious if and only if there is something it is like to be in it. But the claim that there is something it is like to be in the state does not by itself entail you are acquainted with it. As I have argued elsewhere, to say that there is something it is like to be in a state is (roughly) to say that the subject of the state is affected in a particular way in virtue of being in the state, and in particular feels a certain way in virtue of being in the state (Stoljar 2016). If that is right, the phenomenal conception of consciousness does not say anything about acquaintance. This doesn't mean that the acquaintance theory is false of course, nor that it is wrong to restrict acquaintance to phenomenally conscious states; but it does mean that arguments are required for these claims.

Third, what is the proper object of acquaintance? I said above that on the acquaintance theory you are acquainted with being in the state, but many defenders of the acquaintance theory want to say that you are acquainted with the state itself rather than being in it. There are some complications with this way of speaking. It suggests that one can be acquainted with a state whether or not one is in it, which raises the question of how being acquainted with a state you are in puts you in an epistemic position to believe you are in it. But for present purposes I will set this aside, and allow that we can speak in both ways.

Fourth, several issues arise about how to understand and develop the claim that you are in a conscious state if and only if you are acquainted with being in it. My assumption is

that this claim is necessary if true for the acquaintance theory. But one might still ask what direction of explanation, if any, is at issue here. Are you in a conscious state because you are acquainted with being in it or vice versa? Likewise, one might ask if this claim should be upgraded to an identity claim, as Nida-Rümelin suggests in the passage above, or whether it should remain merely a necessary equivalence. Once again, these issues will not affect the points I want to make.

Finally, how close is the acquaintance theory to a traditional inner sense theory? It is clearly different from Armstrong's view, and so if that is what is meant by an inner sense theory, the acquaintance theory isn't it. But, at least as I am understanding it here, it is also different from what Shoemaker calls the 'object-perceptual' view, a position that in fact deserves the name 'inner sense theory' to a greater extent than Armstrong's.

A key feature of the object-perceptual view is that, unlike Armstrong's view (and indeed unlike both Shoemaker's constitutivism and the rationalist theory), it operates with a three-way conceptual distinction between: (a) the introspective belief to the effect that you are in a particular psychological state; (b) the target psychological state that makes that belief true; and (c) an intermediate perception-like state which allows the agent to form the introspective belief in the first place. This structure is intended to be analogous to ordinary cases of perception in which it is typical to distinguish (a) the perceptual belief to the effect that some physical object in the local environment exists or has some property; (b) the target object that makes this belief true; and (c) an intermediate perceptual state which allows the agent to form the perceptual belief. Since the acquaintance view likewise postulates an intermediate perception-like state (i.e. in the form of acquaintance) it too operates with such a conceptual distinction, and to that extent resembles the object-perceptual model. But the difference is that the object-perceptual model assumes only a contingent connection between the target state and the intermediate perception-like element. By contrast, on the acquaintance theory, the connections here are metaphysically necessary; indeed, the defining feature of the acquaintance theory, as we have seen, is that you are in a conscious state if and only if you are acquainted with being in it.¹¹

7. The Logical Situation

¹¹Again, some versions of the acquaintance theory may assume that the connections described in the text are contingent rather than necessary, and hence would become versions of the object-perceptual model. I am setting aside such versions here.

Now that we have both the theories before us, we may proceed to their comparison. I noted above that the two theories are not in conflict, but in fact it is more correct to say that they are logically independent. You can accept neither or both or merely one or merely the other.

As regards the possibility of accepting neither, this is illustrated by, among other things, Shoemaker and Armstrong. Their positions are committed neither to acquaintance nor the introspection principle.

As regards the possibility of accepting both, a proponent of the acquaintance theory may well accept the rationalist theory but take themselves to be filling it out in a particular way. A good way to illustrate this possibility is to notice an important choice point in thinking about the introspection principle or indeed any principle of rationality: whether to treat it as derived, i.e., a principle that follows from a more basic principles of rationality plus auxiliary assumptions, or whether to treat it as basic, i.e., a principle that is not derived. It is possible for the acquaintance theorist to say that the introspection principle is derived, and moreover that it is derived from a more basic principle that we might call *the acquaintance principle*. On this principle, if you are acquainted with a state you are in, you should believe that you are in that state, other things being equal and given background conditions. An agent who abides by this principle will typically believe that they are in the states they are acquainted with and will have a capacity to do so. If it is a principle of rationality that you will believe that you are in a state if you are acquainted with being in it, and if it is also the case that you are in a conscious state if and only if you are acquainted with being in it, then it is natural to infer that it is a principle of rationality that you will believe you are in a conscious state if you are. In other words, the introspection principle follows from the acquaintance principle together with the acquaintance theory itself.¹²

As regards the possibility of accepting the acquaintance theory alone, while the acquaintance theory may agree with the rationalist theory, there is nothing in the view itself to mandate this.

Finally, as regards the possibility of accepting the rationalist theory alone, this can best be illustrated by looking again at what it means to abide by principles of rationality, such

¹² Even if one adopts the rationalist theory and rejects the acquaintance theory, one still faces the issue of whether to treat the introspection principle as a basic or derived principle. My own preference is to be open minded here. The introspection principle cannot be derived in the way discussed in the text, since that presupposes that the acquaintance theory is true; moreover, it cannot be derived from principles of rationality that concern inference or sense perception, since that would undermine the distinctiveness of introspection. But it doesn't follow that it cannot be derived at all; it remains a possibility that there is some basic principle of rationality that generates all the rest presumably with auxiliary assumptions. However, since this is not a matter that needs to be resolved here, I will set it aside.

as the no contradiction principle. I noted earlier that to abide by this principle it is not required that you know what beliefs you have; the point rather is that you exhibit a particular combination of beliefs as a consequence of knowing or representing the principle. A related point applies to acquaintance: to abide by the no-contradiction principle it is not required that you are acquainted with the belief states you are in. The same thing applies to conscious states if the rationalist theory is true: to abide by the introspection principle, and so form the belief that you are in a particular conscious state, it is required of course that you are in conscious states, but it is not required that you are acquainted with them.

While these possibilities are available in principle, however, in the remainder of the paper I will limit my attention to arguing that the option of holding the rationalist theory alone is preferable to any option that involves holding the acquaintance theory, either alone or in conjunction with the rationalist theory. I will for the most part not distinguish these two acquaintance possibilities, and will continue to speak of the comparison between a rationalist theory (by which I mean that option alone) and the acquaintance theory (by which I mean the acquaintance theory in any form). The justification for this procedure is as follows. First, the option of rejecting both has in effect already been set aside; the problems with Armstrong and Shoemaker illustrate the limits of that possibility. Second, there is no need to deal with the two acquaintance options separately since the arguments I will give against acquaintance concern the theory itself rather than whether it is held on its own or in conjunction with something else.

What then are the arguments to adopt the rationalist theory alone rather than any version of the acquaintance theory?

8. Implausible Constraints

The first argument against the acquaintance theory is that it imposes an implausible constraint on the class of psychological states that are introspectable, i.e., open to introspection. We noted above that the acquaintance theory is limited to phenomenally conscious states, i.e., states such that there is something it is like to be in them. It follows, at least if we take it to be a general theory of introspection, that on the acquaintance theory, a psychological state is introspectable only if it is phenomenal.

But on the face of it this claim is false; there are several examples of states that are introspectable but not phenomenal. Consider belief. You may believe (to borrow an example from Shoemaker) that Sacramento is the capital of California and also know by introspection that you do. It follows that the belief is conscious in some sense, since

introspection tells us about our conscious states rather than about unconscious ones. But it doesn't follow that the belief is phenomenally conscious, i.e., that there is something it is like to be in it. If the belief is conscious but not phenomenally conscious, the acquaintance theory is false.

The rationalist theory imposes no such constraint. As I noted above, the introspection principle that is at the heart of that theory leaves open what notion of a conscious state is in play. Hence there is a lot of scope in that theory for conscious states that are not phenomenally conscious. One possibility here is to draw a distinction between phenomenal consciousness and access consciousness in a variety of senses, and then say that beliefs are access conscious rather than phenomenally conscious; belief will then fall within the scope of the introspection principle so long as it is interpreted in turn as applying to both kinds of conscious states. Alternatively, one might interpret the principle as applying uniformly to access conscious states, and then suggest that phenomenal conscious states are introspectable only if they are also access conscious.

I won't explore these options here; the point rather is that the rationalist theory faces no problem over the possibility of states that are introspectable but not phenomenal. But how might the acquaintance theory respond to that possibility? There are several options here but all have drawbacks.

First, it might be argued that in fact beliefs are either not introspectable or not conscious. But that is to deny obvious facts. It is obvious not only that we believe things but that we know that we do. It is obvious moreover that on many occasions we come to know this in a way that relies neither on outer perception nor inference—and that is the defining mark of introspection.

Second, it might be argued that beliefs are themselves phenomenally conscious or perhaps can be defined in terms of things that are phenomenally conscious, e.g., acts or judgements of various kinds. However, while there are suggestions like this in the literature, I think it is fair to say that they are rather controversial; see, e.g., (Smithies 2019, Pitt 2024). It is preferable, therefore, to have a theory of introspection that is consistent with what Kriegel (2015) calls 'mainstream stinginess' regarding belief, i.e., the view that, while beliefs may be introspectable and so conscious in some sense, they nevertheless lack phenomenal character.

Third, it might be argued that properly understood the acquaintance theory is not a general theory of introspection but is a theory only of the phenomenal case; when we turn to non-phenomenal states, some other theory will have to be worked out. But the problem with this option is what the other theory is going to be. One might of course appeal to the

rationalist theory at this point, but if so, there is no reason why we should not appeal to it uniformly.

Finally, it might be argued that the acquaintance theory might be modified so that it applies to non-phenomenal states as well. Perhaps, for example, phenomenally conscious states bear a necessary connection to acquaintance but access conscious states bear only a contingent connection. The problem with this is that now the theory is indistinguishable from the object-perceptual version of the inner sense view mentioned earlier. As I have noted (fn.10), I follow Shoemaker in thinking that such a view is implausible.

9. Simple Creatures

One problem for the acquaintance theory concerns the possibility of introspectable states that are not phenomenal; a different problem concerns the reverse possibility: phenomenal states that are not introspectable.

Consider essentially simple creatures that, in Bentham's famous phrase, can suffer but can't reason. Such creatures may well be in phenomenal states—for example, they may feel pain. But such creatures will not and cannot believe that they are in such states; by their nature, they have no capacity to do so and no interest in doing so. The acquaintance theory does not entail that such creatures will *form* introspective beliefs. Nevertheless, it seems to be committed to something related, namely, that they *ought* to form introspective beliefs, or at any rate, ought to do so from a purely epistemic or rational point of view. The considerations in favour of this last claim are as follows. First, the acquaintance theory in the form we have been considering it entails immediately that the creatures are in an epistemic position to form introspective beliefs: they are in phenomenal states, after all, and hence by the theory are acquainted with being in such states. Second, to be in an epistemic position to form a belief is to have a good reason to form that belief; hence the creatures have a good reason to form introspective beliefs—indeed they have as good a reason to form such beliefs as we sophisticated creatures often do. Third, the creatures clearly have no epistemic reason against holding such beliefs—what could such reasons be? Finally, as a general matter, if you have a good reason to form a belief, and have no reason for not doing so, then you ought to hold it. But this claim—that essentially simple creatures of the kind we are imagining ought of form introspective beliefs—is, I think, highly objectionable. The reason is that it violates an epistemic “‘ought’ implies ‘can’” principle that plausibly applies in this case, namely, that if you ought epistemically to believe something, there must be at least a sense in which you are able to do or at least able to believe something similar. There is no sense in which essentially

simple creatures are able to form the relevant beliefs—that is in part what it means to be essentially simple. Hence the acquaintance theory goes wrong when it entails that they ought to do so.

Once again, the rationalist theory does not face this problem. The essentially simple creatures we are imagining do not abide by the introspection principle; that too is part of what it means to be essentially simple. But since they do not abide by that principle, they are not in an epistemic position to know or rationally believe they are in pain, even if they feel pain; nor, in consequence, is it the case that they ought to believe this.

How might the proponent of the acquaintance theory respond to the objection that, on that theory, essentially simple creatures ought to form introspective beliefs? One possibility of course is to reject the epistemic “‘ought’ implies ‘can’” principle outright or else to argue that it doesn’t apply in this case. I won’t attempt to assess this option here beyond making the following points. First, the version of the principle invoked here is plausible in part because it is so weak; it says only that if you ought to believe something there must a *sense* in which you are able to do so or at least believe something similar. Second, recent defences of such a principle are in my view compelling; see, e.g., (Wedgwood 2013).¹³ Third, taking matters in this direction is at the very least a cost to the acquaintance theory, a cost that the rationality theory does not incur.

It might alternatively be suggested that the acquaintance theory can be modified so that it does not entail that merely being acquainted with the state puts you in an epistemic position to form the belief. Perhaps, for example, it is being acquainted plus some further enabling condition. However, the problem with this suggestion is what the enabling condition is going to be. Clearly if it is or entails abiding by the introspection principle, we are back where we started.

Finally, the proponent of the acquaintance theory might respond by going on the attack. It may be true that the acquaintance theory has a problem with simplicity, but doesn’t the rationalist theory have a problem of its own with sophistication? In particular, does the rationalist theory not limit the scope of introspection to sophisticated creatures, i.e., those who can abide by the principles of rationality? And doesn’t that represent introspection in general as an overly sophisticated endeavour?

¹³ For further discussion of epistemic ‘ought’ implies ‘can’, see Smithies 2019; for discussion of the principle in the (more usual) moral context, see, e.g. King 2017, 2019 and Southwood 2016.

However, while this is a natural reaction to the rationalist view, there are several ways to blunt it. First, while being in various mental states may happen in simple creatures, forming introspective beliefs requires some cognitive sophistication on anyone's view, since it requires having the concepts needed to form the relevant beliefs; hence it is no objection to the rationalist theory that it attributes some sophistication to agents that can come to know or rationally believe things in this way. Second, the notion of rationality in play here has little to do with conscious reasoning or critical reflection, things that do require some high degree of sophistication. As noted above, being rational is to be understood on the model of understanding a language. But there is no a priori insight into what level of sophistication is appropriate or not in such cases; as I said before to abide by the principles does not mean that you can articulate what the principles are. Finally, as I noted above (fn. 5), the kind of view about rationality that I am adopting can be developed in either a capacity first way or a principles first way; while the principles-first way in my experience provokes a 'too sophisticated' reaction, the capacity first way is less likely to.

10. Memory

I have been concentrating on the idea that the acquaintance theory has several objectionable consequences that the rationalist theory considered alone does not. One is that the theory entails that all introspectable states are phenomenal; another is that it entails that phenomenal states are such that essentially simple subjects ought to believe that they are in them.

But there is a further problem for the acquaintance theory that merits attention. Even if we focus on standard cases of phenomenal states, it is difficult to provide any convincing argument that the subjects of such states are acquainted with them as a matter of necessity as the acquaintance theory suggests. One argument, for example, might be that the acquaintance theory follows directly from the definition of phenomenal consciousness; but as we noted above, this is not so. Another argument is that postulating acquaintance is the only way of understanding introspection. But given the possibility of a rationalist theory, this is not so either.

In fact, the problem of providing a convincing argument for the acquaintance theory is a major issue. Elsewhere I have offered a systematic list of arguments for acquaintance in the literature and argued that none of them is persuasive (Stoljar 2021b). I won't relitigate those points here. Instead I will make a subsidiary point about one of the arguments on this list, recently defended by Uriah Kriegel and Anna Giustina, namely, the memory argument

(Kriegel 2019, Giustina 2022). Kriegel himself goes as far as to say that “no better argument has been produced” for the acquaintance theory.

The memory argument can be understood in various ways; in the paper just mentioned I focused on an abductive version based on propositional memory or what psychologists call ‘semantic memory.’ But both Kriegel and Giustina suggest that the argument is better understood as a deductive and as focusing on what psychologists call ‘episodic memory.’ They offer the following reconstruction:

- P1. A subject can remember an event E only if she was aware of E when E occurred;
- P2. Every conscious state is such that there is some later time at which its subject can remember it;
- C. Every conscious state is such that its subject is aware of it at the time of its occurrence.

Since, as we noted at the outset (see fn. 2), the notion of ‘awareness’ here can be understood as equivalent to acquaintance, the conclusion of this argument is equivalent to the main claim of the acquaintance theory.

How plausible is this version of the memory argument? The problem I will concentrate on is why we should accept P1.¹⁴ According to Kriegel and Giustina, P1 follows from the nature of episodic memory: “The truth of P1”, Giustina writes, “becomes straightforwardly apparent once we distinguish *episodic* from *semantic* memory...The former consists in first-person experiential recollection of events of one’s personal past, while the latter is third-person memory of facts of the world.” (Giustina 2022, p. 5). She goes on to say one cannot recollect events in this way without being aware of them: for example, “the police will rely on my memory to reconstruct the dynamics of a car accident only if I witnessed the accident—i.e., only if I was somehow aware of the accident: they will make no use of my memory if I did not see, hear, or experience the accident in any way” (2022, p.5)

But I think is a mistake to view considerations like this as rendering P1 plausible. The problem is that there are different sorts of events. It may be true that I can only have first-person experiential recollection of *the accident* if I witnessed it and so was aware of it. But it does not follow that I only can have first-person experiential recollection of *witnessing the*

¹⁴ See Silins (forthcoming) for a critical discussion of the argument that focuses on P2.

accident in the same way. How might I have first-person experiential recollection of witnessing the accident without being aware of it? An obvious suggestion is that I may do so if there was something it is like for me to witness the accident. But this by itself does not entail that I was aware of witnessing the accident. As I have mentioned, on the view of ‘what it is like’ that I have defended elsewhere, to say that there was something it is like for the subject to undergo E means only that the subject felt a certain way in virtue of undergoing E. From this it does not follow that the subject was aware of undergoing E.

What this suggests more generally is that the memory argument in the version Kriegel and Giustina advance falters at its first premise. What is plausible is not P1 but something weaker, namely:

P1* A subject can remember an event E only if *either* she was aware of it when E occurred *or* there was something it was like for her to undergo E when E occurred.

And the problem is that P1*, even when combined with P2, does not yield the acquaintance theory; similarly, P1* by itself does not yield P1.

One might reply that this criticism of the memory argument is only as good as the view of ‘what it is like’ phrases on which it depends. Again, I am assuming here that there is something it is like to undergo an event if and only if you are affected in some way by that event. Would it be possible to reject this view in favour of one friendlier to the acquaintance theory? One problem with doing so is that such views are implausible, as I have suggested elsewhere. But another problem, and in the present context a more pressing problem, is that this defence of the memory argument turns now on a theory about ‘what it is like’ phrases, rather than anything specifically to do with memory.

11. Conclusion

In the first part of this paper, I set out the rationalist theory and the acquaintance theory; in the second part, I argued that, while it is possible to hold both theories, the most attractive option is to adopt the rationalist theory and reject the acquaintance theory, since the acquaintance theory faces several major problems that the rationalist theory does not. It is in this sense that the rationalist theory is preferable to the acquaintance theory.

Of course, that the rationalist theory is preferable to one theory does not mean it is preferable overall. There are several major questions for the rationalist theory I will not take

up here. Still, the acquaintance theory is perhaps the dominant one in the literature today; that the rationalist theory is in this way preferable to it is a major point in its favour.

References

- Armstrong, D. M. and N. Malcolm (1984). Consciousness and Causality: A Debate on the Nature of Mind. Oxford, Blackwell.
- Broome, J. (2013). Rationality Through Reasoning. Oxford, Wiley-Blackwell.
- Byrne, A. (2018). Transparency and Self-Knowledge. New York, Oxford University Press.
- Dretske, F. (2003). How Do You Know You are Not a Zombie? Privileged Access: Philosophical Accounts of Self-Knowledge. B. Gertler. Burlington, VT, Ashgate Publishing.
- Duncan, M. (2021). "Acquaintance." Philosophy Compass **16**(3): 1-19.
- Fumerton, R. (2001). "Brewer, Direct Realism, and Acquaintance with Acquaintance." Philosophy and Phenomenological Research **September**: 417-422.
- Fumerton, R. (2019). Acquaintance: The Foundation of Knowledge and Thought. Acquaintance. J. Knowledge and T. Raleigh. Oxford, Oxford University Press: 245-259.
- Gertler, B. (2011). Self-Knowledge. London, Routledge.
- Gertler, B. (2012). Renewed Acquaintance. Introspection and Consciousness. D. Smithies and D. Stoljar. New York, Oxford University Press.
- Giustina, A. (2022). "Nature Does Not Yet Say No to Inner Awareness: Reply to Stoljar." Erkenntnis.
- Harman, G. (1986). Change In View: Principles of Reasoning. Cambridge, MA, MIT Press.
- King, A. (2017). "Ought Implies Can: Not so Pragmatic After All." Philosophy and Phenomenological Research **95**(3): 637-661.
- King, A. (2019). What We Ought and What We Can, Routledge.
- Kriegel, U. (2009). Subjective Consciousness: A Self-Representational Theory. New York, Oxford University Press.
- Kriegel, U. (2015). The Varieties of Consciousness. New York, Oxford University Press.
- Kriegel, U. (2019). "Dignāga's Argument for the Awareness Principle: An Analytic Refinement." Philosophy East and West **69**(1): 144-156.
- Lewis, D. (1994). Reduction of Mind. A Companion to the Philosophy of Mind. S. Guttenplan. Oxford, Blackwell.
- Martin, M. G. F. (1998). An Eye Directed Outward. Knowing Our Own Minds. C. Wright, B. C. Smith and C. Macdonald. New York, Oxford University Press.

McGinn, C. (1982). The Character of Mind: An Introduction to the Philosophy of Mind. New York, Oxford University Press.

Montague, M. (2016). The Given: Experience and Its Content. Oxford, UK, Oxford University Press.

Moran, R. (2001). Authority and Estrangement: An Essay on Self-Knowledge. Princeton, NJ, Princeton University Press.

Nida-Rümelin, M. (2011). "Phenomenal Presence and Perceptual Awareness: A Subjectivist Account of Perceptual Openness to the World." Philosophical Issues **21**(1): 352-383.

Peacocke, C. (1998). Conscious Attitudes, Attention, and Self-Knowledge. Knowing Our Own Minds. C. Wright, B. C. Smith and C. Macdonald. New York, Oxford University Press.

Pitt, D. (2024). The Quality of Thought. New York, Oxford University Press.

Shoemaker, S. (1996). The First-Person Perspective and Other Essays. Cambridge, Cambridge University Press.

Shoemaker, S. (2009). "Self-Intimation and Second Order Belief." Erkenntnis **71**(1): 35-51.

Smithies, D. (2019). The Epistemic Role of Consciousness. New York, Oxford University Press.

Southwood, N. (2016). "Does 'Ought' Imply 'Feasible'?" Philosophy and Public Affairs **44**(1): 7-45?

Stoljar, D. (2012). Knowledge of Perception. Introspection and Consciousness. D. Smithies and D. Stoljar. New York, Oxford University Press.

Stoljar, D. (2016). "The Semantics of 'What it's like' and the Nature of Consciousness." Mind **125**(500): 1161-1198.

Stoljar, D. (2018). "Introspection and Necessity." Noûs **52**(2): 389-410.

Stoljar, D. (2019). "Evans On Transparency: A Rationalist Account." Philosophical Studies **176**(8): 2067-2085.

Stoljar, D. (2021a). Armstrong's Just-so Story about Consciousness. Armstrong's Materialist Theory of Mind. P. Anstey and D. Braddon-Mitchell. Oxford, Oxford University Press: 176-194.

Stoljar, D. (2021b). "Is there a persuasive argument for an inner awareness theory of consciousness?" Erkenntnis: 1-21.

Stoljar, D. (Forthcoming). A Euthypho Dilemma for Higher-order Theories of Consciousness. Grounding and Consciousness. G. Rabin, Oxford University Press.

Strawson, G. (2015). "Self-intimation." Phenomenology and the Cognitive Sciences **14**(1): 1-31.

Wedgwood, R. (2013). "Rational 'Ought' Implies 'Can'." Philosophical Perspectives **23**: 70-92.

Zimmerman, A. Z. (2006). "Basic Self-Knowledge: Answering Peacocke's Criticisms of Constitutivism." Philosophical Studies **128**(2): 337-379.