
THE AGENTIAL POINT OF VIEW

BY

BEN SORGIOVANNI

Abstract: Agentialist accounts of self-knowledge seek to do justice to the connection between our identities as rational agents and our capacity to know our own minds. There are two strategies that agentialists have employed in developing their position: substantive and non-substantive. My aim is to explicate and defend one particular example of the non-substantive strategy, namely, that proposed by Tyler Burge. In particular, my concern is to defend Burge's claim that critical reasoning requires a relation of normative directness between reviewing and reviewed perspectives. My defence will involve supplementing Burge's view with a substantive agentialist account of self-knowledge.

1. Introduction

Agentialist accounts of self-knowledge are motivated by the intuition that an adequate account of self-knowledge must do justice to the connection between our identities as rational agents – as creatures who generally hold propositional attitudes on the basis of reasons – and our capacity to know our own minds. Typically, the argument is made that *empiricist* views of self-knowledge – views that understand self-knowledge on the model of ordinary empirical knowledge¹ – cannot meet this condition, and so it is concluded that our identities as rational agents vouch safe a form of non-empirical knowledge of our mental states.²

This brief summary of the dialectic between agentialists and empiricists prompts at least the following two questions. (1) What exactly is the relevant connection between our identities as rational agents and our capacity for self-knowledge? (2) Why think that empirical accounts of self-knowledge (hereafter simply *empirical views* or *empiricism*) cannot do justice to this connection?

There are, broadly speaking, two strategies that agentialists have employed in responding to (1) and (2).³ The first strategy involves giving an account, in the course of identifying the relevant connection between self-knowledge and rational agency, of how we can know our mental states non-empirically. Call this way of responding to (1) and (2) *the substantive strategy*, substantive in the sense that it involves identifying a method or route via which we can come to know our mental states. A prominent proponent of this strategy is Richard Moran. According to Moran,

... as I conceive of myself as a rational agent, my awareness of my belief is awareness of my commitment to its truth, a commitment to something that transcends any description of my psychological state. And the expression of this commitment lies in the fact that my reports on my belief are obliged to conform to the condition of transparency: that I can report on my *belief* about X by considering (nothing but) X itself. (Moran, 2001, p. 84)

For Moran, the relevant connection between self-knowledge and rational agency is the fact that for rational agents such as ourselves, to be aware of believing that p is to be aware of being committed to the truth of p. This connection is evidenced by the fact that the question whether I believe that p is transparent to the question whether p, by the fact that I can come to know whether I believe that p by reflecting, not on facts about myself, but on the question whether p.

The second strategy agentialists have employed in responding to (1) and (2) is to argue that the relevant connection between self-knowledge and rational agency can be met only if we can know our attitudes non-empirically, but to argue for this conclusion in a way that leaves it open how exactly such non-empirical self-knowledge might be gleaned. Call this strategy *non-substantive*. It is pursued in, for example, Tyler Burge's 'Our Entitlement to Self-knowledge'. Burge's central argument against empiricist conceptions of self-knowledge is transcendental in form. On Burge's view, the relevant connection between self-knowledge and rational agency goes via our identities as critical reasoners, as creatures that are capable of engaging in 'reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such' (Burge, 1996, p. 98). Critical reasoning is both carried out within, and applicable to, a particular perspective. Let us follow Burge in calling the perspective within which reasoning is carried out the *reviewing perspective* and the perspective to which the reasoning is applicable the *reviewed perspective*. Burge takes it to be a necessary condition for our genuinely reasoning critically that there is a relation of a particular kind between these perspectives. Specifically, in instances of genuine critical reasoning, there is an 'immediate rationally necessary connection between justified rational evaluation within the review, on one hand, and its being *prima facie* reasonable within the reviewed perspective to shape attitudes in accord with that evaluation, on the other' (Burge, 1996, pp. 109–10). He goes on to argue that this condition can't be met unless we can know our attitudes

non-empirically. So, given that we *are* critical reasoners, we *must* be able to know our attitudes non-empirically. Burge's discussion leaves it open, however, how exactly it is that we know our attitudes non-empirically.

These two ways of responding to (1) and (2) – substantive and non-substantive – are typically discussed in isolation from one another. Indeed, the Burgean strategy has received significantly less attention in the literature than has Moran's strategy. I suggest that this is due to unclarity about the view itself. Two questions in particular present themselves: first, why does critical reasoning require a rationally necessary connection of the sort envisaged by Burge; and second, how could there be such a connection? My aim in this paper is to explicate and defend the Burgean view by addressing these two questions. In answer to the first question, I shall argue that a rationally necessary connection of the relevant kind is implied by a norm of critical reasoning. My answer to the second question will involve supplementing Burge's view with an account of self-knowledge that is substantive in the earlier sense – an account that involves identifying a method or route via which we can come to know our mental states non-empirically.

The discussion proceeds as follows. In Section 2, I elucidate the notion of a rationally necessary connection between reviewing and reviewed perspectives and explain why purely observational accounts of self-knowledge seem not to permit such a connection. In Section 3, I elucidate Burge's rationale for the claim that there must be a rationally necessary connection in order for us to be critical reasoners and defend the plausibility of that rationale. In Section 4, I show how supplementing Burge's view with a substantive account of self-knowledge provides an answer to the question of how there could so much as be a rationally necessary connection between reviewing and reviewed points of view.

2. *Normative directness and observational accounts of self-knowledge*

Central to Burge's non-substantive strategy is the claim that human beings are critical reasoners, creatures that are capable of, and who sometimes engage in, critical reasoning,

reasoning that involves an ability to recognise and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. (Burge, 1996, p. 98)

All reasoning requires a capacity to be responsive to reasons. What is distinctive about critical reasoning, Burge thinks, is that it requires a capacity to *reflect* on one's reasons and reasoning. To count as a critical reasoner, a creature must be able to think of reasons *as reasons*, to evaluate reasons

thus conceptualised and to assess one's reasoning in the light of this evaluation.

Critical reasoning requires, Burge maintains, a connection of a special sort between the reviewing and reviewed perspectives. The requisite connection is summarised in the following passage:

If in the course of critical reasoning I reasonably conclude that my belief that a given person is guilty rests entirely on unreasonable premises or bad reasoning, then it normally follows immediately both for the perspective of the review and for the perspective of the reviewed belief that it is reasonable to give up my belief about guilt or look for new grounds for it. (Burge, 1996, p. 110)

Let us say that a reached conclusion is reasonable for a subject if and only if the subject reaches it in a way that is guided by the norms of critical reasoning (Burge, 1996, p. 94. See also Burge, 2013c, p. 197; Burge, 2013a, pp. 493–4). Such norms include, but are not limited to, the rules of deductive and inductive inference. With this clarification in hand, we can understand Burge's point as follows. Suppose I believe that some person, S, has done me wrong. Suppose I subsequently come to believe, first, that q and r and, second, that q and r are strong reasons not to believe that S has done me wrong. Suppose I believe that there are no reasons that weigh decisively against q and r. Suppose that on the basis of these beliefs, I conclude that it is all things considered (hereafter *ATC*) reasonable for me to give up my belief that S has done me wrong. Assuming the steps in my reasoning are themselves reasonable, it is *ATC* reasonable for me to give up my belief that S has done me wrong. What's more, Burge wants to say, insofar as I am reasoning critically, it follows *immediately* and *necessarily* that it is *ATC* reasonable for me to give up that belief. (Exactly why Burge thinks this is a question taken up in Section 3.) According to Burge, it is necessary for critical reasoning that if I reasonably judge that it is (*prima facie*/*ATC*) reasonable for me to believe that p, then it follows immediately and necessarily that it is (*prima facie*/*ATC*) reasonable for me to believe that p.⁴

Let us say that whenever this condition is satisfied, there is a relation of *normative directness* between the reviewing and reviewed perspectives. Let us refer to the requirement that, for critical reasoning to take place, there must be a relation of normative directness between the reviewing and reviewed perspectives as the *normative directness requirement*. And let us refer to the objection that this requirement cannot be met if we can know our mental states only empirically as the *objection from normative directness*. In his 1996, Burge devotes his energies to explaining why the normative directness requirement cannot be met by observational accounts of self-knowledge, accounts according to which all self-knowledge is 'based either on imaging, remembering, or reasoning about *sensed* inner-goings-on, or on observing our own behaviour and hearing about it from others' (Burge, 1996, p. 104). Consider a case in which I am critically reviewing your

attitudes. Suppose I reasonably judge that it is ATC reasonable for you to believe that *p*. Still, it doesn't follow immediately or necessarily that it *is* ATC reasonable for you to believe that *p*. After all, I may be wrong about certain of your beliefs. My beliefs about your beliefs might be reasonable and nevertheless false. Even if I happen to be right about what you believe, the connection is neither immediate nor necessary. It is not immediate for it follows from my reasonably judging that it is ATC reasonable for you to believe that *p* that it is ATC reasonable for you to believe that *p*, only via the following mediating step: my reasonable beliefs about your beliefs, upon which my reasonable judgment that it is ATC reasonable for you to believe that *p* is based, are in fact accurate. It is not necessary, for the connection obtains only if my reasonable beliefs about your beliefs *happen to be* true.

But if the observational account is correct, there is no relevant difference between our knowledge of our own mental states and our knowledge of the mental states of other people.⁵ In both cases, our knowledge is based wholly on the deliverances of our senses. Consequently, if the observational account is correct, in every instance of self-reflection, there is the sort of dissociation between reviewed and reviewing perspectives just described. Even if one were reasonable in judging that it is ATC reasonable for one to believe that *p*, it would not follow immediately or necessarily that it is in fact ATC reasonable for one to believe that *p*. It would not follow immediately or necessarily because one might be in error about one's beliefs. One's second-order beliefs about one's first-order beliefs might be reasonable and yet false. Even in the case in which one happens to be right about what one believes, there is no connection of the requisite kind. In this case, it follows from one's reasonably judging that it is reasonable for one to believe that *p* that it is reasonable for one to believe that *p*, but only because one happens to be correct about what one believes (when one might have been incorrect and yet still reasonable). If the observational account is right, there is never the sort of normative directness between the reviewed and reviewing perspectives that Burge wants to say is required for critical reasoning.

But we *are* creatures that are capable of, and sometimes do engage in, critical reasoning. Consequently, there must be a relation of normative directness between our reviewing and reviewed perspectives. So it must be possible for us to know our mental states non-observationally. This is, in broad outline, Burge's (1996) argument against observational accounts of self-knowledge.

Now, it doesn't follow directly from this argument that it must be possible for us to know our mental states *non-empirically*. After all, being an observationalist is but one way to be an empiricist. One might think that our warrant for self-ascriptions is empirical warrant without thinking that it is perceptual (or quasi-perceptual) warrant. The argument will show that it must be possible for us to know our mental states non-empirically only if it shows that *any* account of empirical self-knowledge is incompatible

with normative directness. But it's not obvious that this is the case. The objection from normative directness applies to any account that in every case, allows that a judgment about one's mental states may be reasonable – guided by the norms of critical reasoning – and yet false. So long as this allowance is made, it will not follow immediately or necessarily from my reasonably judging that it is ATC reasonable for me to believe that *p* that it is ATC reasonable for me to believe that *p*. While any broadly observationalist account seems committed to making this allowance (for the reasons given previously), this is not obviously true of all accounts that might fairly be classified as empirical.⁶ In any case, I will not try to settle this issue here. I will regard the objection from normative directness as an objection to empiricism generally, although there are some forms of empiricism to which it may turn out not to apply.

Brie Gertler (2016) has recently argued that the objection does not apply even to broadly observationalist accounts. Gertler challenges the claim that there cannot be an immediate rationally necessary connection between reviewing and reviewed perspectives in those instances where we know our attitudes on a wholly observational basis. Gertler begins by focusing on another strand of thought in Burge, namely, the thought that in instances of critical reasoning, judgments about what it is reasonable for one to believe are immediately effective within the perspective under review:

Where we know our thoughts or attitudes only by observation, the question of *means* of control – of effective application of reasons to them – arises We must face a question of how, by what means, to make those reasons effective in view of the contingent relation between the point of view of the self-knowledge and rational evaluation, on one hand, and the observationally known attitudes, on the other. Again, this is not critical reasoning. In critical reasoning, such questions of means and control do not arise, since one's relation to the known attitudes is rationally immediate: they are part of the perspective of the review itself. (Burge, 1996, p. 113)

Suppose that after reflection, a juror comes to realise that the evidence on the basis of which he believes the defendant to be guilty is very weak. Suppose, however, that this second-order belief is not sufficient to displace the first-order belief; in the face of his realisation, the juror's belief that the defendant is guilty perseveres. In this case, the juror faces a decision about how to make his second-order belief effective. By engaging in certain sorts of imaginative exercises – for example, by repeatedly imagining the defendant doing whatever it was the defendant says he was doing during the time of the infraction – the juror may eventually extinguish his first-order belief. But, according to Burge, the mere fact that he has to engage in this sort of exercise shows that his reasoning is not critical reasoning properly conceived. Compare Gertler's deliberative juror, for whom the realisation that her belief that the defendant is guilty is based on very weak evidence displaces that belief *directly*, in a way that does not depend on the sorts of imaginative exercises in which the detached juror must engage.

Gertler argues that this sort of direct relation can obtain even if we can know our attitudes only observationally. Observational accounts of self-knowledge, according to Gertler, should be understood as views about the nature of one's warrant for self-ascriptions, namely, views according to which that warrant is perceptual or quasi-perceptual warrant. *Prima facie*, there is no tension between this claim and the claim that an agent can act directly on the basis of her judgments about what it is ATC reasonable for her to believe. Whether or not one can act directly on the basis of such judgments is, after all, not a matter of one's warrant for second-order beliefs, but of the way in which those beliefs affect one's first-order states. What is significant is not 'how the attitude is *known*, but rather how the thinker's reasons *affect* it' (Gertler, 2016, p. 16). As long as the judgment that it is ATC reasonable for one to believe that p (or to cease believing that p) leads directly to one's believing that p (or one's ceasing to believe that p), there is a direct relation between the second-order judgment and the relevant first-order state. And there is no obvious reason why this cannot be the case in instances where one's warrant for the second-order judgment is grounded in evidence gleaned through inner or outer observation.

Gertler goes on to suggest that this direct relation is just the sort of rationally necessary connection that Burge thinks critical reasoning requires:

Is this connection between the reviewing and reviewed perspectives an 'immediate rationally necessary connection', as Burge requires for critical self-reflection? It is 'immediate' in that the realisation that the belief is poorly justified unseats that belief directly And this connection is 'rationally necessary' in the sense that the contents of these perspectives jointly contribute to what is rationally required of the thinker: given that realisation, maintaining the belief would violate rational norms. (Gertler, 2016, p. 13)

In fact, there seems to be an important difference between the sort of directness Gertler discusses in this passage and what I am calling normative directness. Certainly, whenever there is a direct unseating of a first-order belief on the basis of a second-order judgment, there is an immediate relation connecting reviewing and reviewed perspectives. But the relation is not rationally necessary, at least not in the relevant sense. It is certainly true that if one reasonably concludes that it is ATC unreasonable for one to carry on believing that the defendant is guilty, then *from the perspective from which the review is being carried out*, it would be a violation of the norms of critical reasoning to go on believing that the defendant is guilty. From that perspective, it has already been reasonably concluded that it is ATC unreasonable for one to retain the belief. But if we can know our attitudes only observationally, it is not true that it follows necessarily that maintaining the belief would violate rational norms *from the reviewed perspective*. From that perspective, it might be ATC reasonable to retain the belief. This is true for the sorts of reasons we have already considered. If one can know one's

attitudes only observationally, in every case one's second-order judgments about one's first-order beliefs might be reasonable and yet false.

For these reasons, I don't think Gertler has shown that there can be a rationally necessary connection between reviewed and reviewing perspectives of the sort that Burge thinks is required for critical reasoning, if we can know our attitudes only observationally. But we might well ask why critical reasoning requires that there must be a rationally necessary connection. It is to this question that I now turn.

3. *Normative directness and critical reasoning*

In several passages, Burge seems simply to take it for granted that normative directness is necessary for critical reasoning. For example, after noting that normative directness seems impossible if we can know our mental states only observationally, Burge goes on:

But this picture [a picture according to which there is never a relation of normative directness between reviewing and reviewed points of view] is nonsense if it is applied to all judgments about one's own propositional attitudes. For it is constitutive of critical reasoning that if the reasons or assumptions being reviewed are justifiably found wanting by the reviewer, it *rationally follows immediately* that there is a *prima facie* reason for changing or supplementing them, where this reason applies within the point of view of the reviewed material (not just within the reviewing perspective). (Burge, 1996, p. 109)

In passages like this one, Burge appears to *assume* that normative directness is necessary for critical reasoning.⁷ How plausible is this assumption? Recall Burge's characterisation of critical reasoning: 'Critical reasoning is reasoning that involves an ability to recognise and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such.' It's not at all obvious why critical reasoning thus conceived requires normative directness. After all, it's consistent with that conception that I can reason critically about another person's attitudes. I can think critically about someone else's reasoning and my criticism can surely be guided by an appreciation, use and assessment of reasons and reasoning as such. But in instances where I'm reasoning about another person's attitudes, there is *no* rationally necessary connection between reviewed and reviewing perspectives (for the reasons given in the previous section). In such instances, it does *not* follow immediately and necessarily from my reasonably judging that it is ATC reasonable for S to believe that p that it is ATC reasonable for S to believe that p. But then surely there *can* be instances of critical reasoning even where there is no relation of normative directness between reviewing and reviewed perspectives.

Burge might well agree with much of this. However, he argues that reasoning critically about another's attitudes requires that one sometimes reason critically about one's own attitudes:

Critical reasoning must be exercised on itself. Any critical reasoning, even about abstract propositional relations or about the reasoning of others, involves commitments by the reasoner. And genuinely critical reasoning requires an application of rational standards to *those* commitments. A being that assessed good and bad reasoning in others or in the abstract, but had no inclination to apply such standards to the commitments involved in those very assessments, would not be a critical reasoner. (Burge, 1996, p. 100)

Suppose that, in the course of reasoning critically about S's attitudes, I judge that because S believes that p and that if p then q, it is *prima facie* reasonable for S to believe that q. In making this judgment, I rely on certain commitments, for example, the commitment that it is *prima facie* reasonable for one to believe that q, given that one believes that p and that if p then q. Burge's point is that my reasoning only constitutes genuine *critical* reasoning if I am inclined to subject this commitment itself to critical scrutiny.

This seems plausible. But it's not clear that it establishes that normative directness is a necessary condition for critical reasoning. It does so on the condition that I can reason critically about my own attitudes only if there is a relation of normative directness between reviewing and reviewed perspectives. If reasoning critically about another's attitudes requires reasoning critically about my own attitudes, and if reasoning critically about my own attitudes requires that there is the sort of rationally necessary connection between reviewing and reviewed perspectives that constitutes normative directness, then normative directness would appear to be a requirement for critical reasoning generally. But why should we think that one can reason critically about one's own attitudes only if there is a relation of normative directness between reviewed and reviewing perspectives? It seems that I can think critically about my own attitudes in a way that is guided by an appreciation, use and assessment of reasons and reasoning as such, even if my beliefs about my attitudes are based exclusively on observational evidence. Suppose that wholly on the basis of observing my behaviour, I reasonably judge that I believe that S has done me wrong. But after careful reflection, I can think of nothing S has done that would warrant such a belief. So I conclude that it is ATC unreasonable for me to believe that S has done me wrong. Surely, *this* is an instance of critical reasoning. And yet in this instance, there is no relation of normative directness between reviewing and reviewed perspectives, for precisely those reasons discussed in the previous section. Insofar as my judgment that I believe that S has done me wrong is based wholly on observational evidence, that judgment, though reasonable, could be false.

We've been considering the claim that normative directness is a necessary condition for critical reasoning in the context of Burge's explicit

characterisation of critical reasoning. That claim does not seem particularly plausible, given that characterisation. Returning to the passage cited at the beginning of this subsection, however, note that Burge's claim is not that normative directness is *necessary* for critical reasoning. It is in fact stronger: normative directness is partly *constitutive* of critical reasoning. This suggests an alternative interpretation of Burge's view. We've been trying to find grounds for thinking that normative directness is a necessary condition for the very possibility of critical reasoning. But perhaps the claim is, rather, that normative directness is somehow implied by a *norm* of critical reasoning, that there is some norm of critical reasoning in virtue of which necessarily, it is (*prima facie*/ATC) reasonable for one to believe that p, given that one has reasonably judged that it is (*prima facie*/ATC) reasonable for one to believe that p. Perhaps Burge's idea is that it follows immediately and necessarily from my reasonably judging that it's (*prima facie*/ATC) reasonable for me to believe that p that it is (*prima facie*/ATC) reasonable for me to believe that p, precisely because my reasonable second-order judgment necessarily *bears constitutively on it's* being (*prima facie*/ATC) reasonable for me to believe that p.

This interpretation seems preferable considered as an interpretation of Burge's view. Not only does it make sense of the occurrence of 'constitutive' in the passage cited earlier but it also fits with Burge's (1996, p. 101) repeated contention that critical reflection 'add[s] a rational element to the reasonability of reasoning'. (Indeed, this contention is crucial to the broader view which Burge is defending in his 1996.) On the interpretation I'm suggesting, critical reflection adds to the reasonability of reasoning precisely because one's reasonable second-order judgments necessarily bear constitutively on what it is reasonable for one to believe at the first-order level. Indeed, this seems to be precisely Burge's thought in the following passage:

Justifiably finding one's reasons invalid or one's thoughts unjustified, is normally *in itself* a paradigmatic reason, from the point of view of the thoughts being reviewed (as well as from the perspective of the review), to alter them. (Burge, 1996, p. 110)

The occurrence of 'in itself a paradigmatic reason' suggests a constitutive relation between one's reasonable second-order judgments about what it is reasonable for one to believe and what it is in fact reasonable for one to believe.

The interpretation also seems preferable considered as a view about the relation between critical reasoning and normative directness. Plausibly, one's reasonably judging that it is (*prima facie*/ATC) reasonable for one to believe that p *does* make it (*prima facie*/ATC) reasonable for one to believe that p.⁸ Our practice of attributing reasonableness and unreasonableness to others constitutes evidence for this. If someone reasonably concludes that it's *prima facie* unreasonable for them to believe that, say, S has done them wrong, their having reached that conclusion is usually treated as itself *prima facie*

grounds for their jettisoning that belief. If they continue to hold the belief that S has done them wrong, and are unable to offer any rationale for their doing so, it is *on that basis alone* appropriate to criticise them for being unreasonable. The analogous point holds true of instances, first, in which someone reasonably judges that it's *ATC* unreasonable for them to believe that p and, second, in which someone reasonably judges that its *prima facie* or *ATC* considered *reasonable* for them to believe that p. In these cases, too, subject's reasonable conclusions about what it is reasonable for them to believe are treated as bearing constitutively on what it is reasonable for them to believe.

Moreover, the interpretation avoids the sort of problems that arose in connection with instances in which one is critically reasoning about another person's attitudes. In such instances, there is no relation of normative directness between reviewing and reviewed perspectives. Clearly, this counts against the claim that normative directness is a necessary condition for the very possibility of critical reasoning. But it does not count against the claim that normative directness is implied by a norm of critical reasoning. Presumably, the relevant norm concerns self-directed critical reasoning only. Roughly, we can expect it to have the following form: if you reasonably judge that it is (*prima facie*/*ATC*) reasonable for you to believe that p, then you ought (*prima facie*/*ATC*) to believe that p. There is no analogous norm that covers the third-person case; it's not true that if I reasonably judge that it's (*prima facie*/*ATC*) reasonable for Jones to believe that p, then Jones (*prima facie*/*ATC*) ought to believe that p. Consequently, the fact that in instances in which one is critically reasoning about another person's attitudes, there is no relation of normative directness between reviewing and reviewed points of view does not count against the interpretation. It does not count against the interpretation because such instances don't fall under the remit of the relevant norm.

Does the interpretation preserve the force of Burge's objections to the observational account (outlined in Section 2)? One might be concerned that it does not. After all, if my reasonably judging that it's reasonable for me to believe that p really does bear constitutively on it's being reasonable for me to believe that p, then why should it matter whether the second-order beliefs about my first-order states on which the judgment is based are acquired observationally or otherwise? Can't the observationalist about self-knowledge readily agree that reasonable judgments about what it is reasonable for you to believe necessarily bear constitutively on what it *is* reasonable for you to believe?

The answer is 'No'. Burge's thought is that my reasonable judgments about what it is reasonable for me to believe necessarily bear constitutively on what it is reasonable for me to believe only if a certain condition is met – namely, the second-order beliefs about my first-order states on which the judgment is based cannot be both reasonable and false. Only then is there

a necessary constitutive relation between my reasonable judgments about the reasonableness of my believing *p* and the reasonableness of my believing *p*. And this thought seems exactly right. Its denial has some plausibility if we take it for granted that the reviewing and reviewed perspectives are constituents of the self-same point of view. If someone judges that it is (*prima facie*/ATC) reasonable for them to believe that *p*, then, plausibly, it is (*prima facie*/ATC) reasonable for that person to believe that *p*. But I take it that Burge's fundamental point is that, if we allow that the second-order beliefs on which the second-order judgment (that it is reasonable to believe that *p*) is based could be false, then there is no reason to think that reviewing and reviewed perspectives *are* constituents of the self-same point of view. If the second-order beliefs on which the second-order judgment is based could be false (and if the second-order beliefs on which *any* similar second-order judgment from that perspective is based could be false, as they could be if a broadly observational account is correct), then the perspective from which this judgment is enacted does not seem essentially different from the perspective of a separate agent. But another person's reasonable judgments about what it is (*prima facie*/ATC) reasonable for me to believe do not necessarily bear constitutively on what it is reasonable for me to believe.⁹ So why think that, in the case as described, the reasonable second-order judgment that it is reasonable to believe that *p* really does bear constitutively on what it is reasonable to believe, from the perspective under review?

This is, I take it, the fundamental rationale for the thought that my reasonable judgments about what it is reasonable for me to believe necessarily bear constitutively on what it is reasonable for me to believe only if the second-order beliefs about my first-order states on which the judgment is based cannot be both reasonable and false. It's only in virtue of this condition's being met that reviewed and reviewing perspectives are constituents of the self-same point of view. But the reasonable judgments from a token reviewing perspective necessarily bear constitutively on what it is reasonable to believe from a token reviewed perspective only if token reviewing and reviewed perspectives are constituents of the self-same point of view. So the reasonable judgments from a token reviewing perspective necessarily bear constitutively on what it is reasonable to believe from a token reviewed perspective only if the second-order judgments on which they are based could not be false. This rationale applies as forcefully in cases of *prima facie* and ATC reasonableness, respectively.¹⁰

A referee asks whether the relevant norm – the norm according to which if you reasonably judge that it is (*prima facie*/ATC) reasonable for you to believe that *p*, then you ought (*prima facie*/ATC) to believe that *p* – might be in place even in those cases in which the judgment is based on false second-order beliefs, in virtue of its role in establishing reviewing and reviewed perspectives as constituting a single perspective. I don't think such a view is plausible. The suggestion seems to get the order of explanation the wrong

way around. It is not, it seems, in virtue of the norm's applying that reviewing and reviewed perspectives are constituents of a single point of view. Rather, it is because they are constituents of the self-same point of view that the norm applies.¹¹ This order of explanation is to be preferred for at least two reasons. First, if the norm applied whether or not the second-order beliefs on which the relevant second-order judgment is based could be false, and if it was in virtue of the norm's applying that reviewing and reviewed perspectives were constituents of a single point of view, then even in cases where the second-order beliefs could be false – indeed, even in cases in which they were *in fact* false – reviewing and reviewed perspectives would count as constituents of the self-same point of view. But it isn't plausible that reviewing and reviewed perspectives really are constituents of the self-same point of view, in such cases. As noted previously, the relation between the perspectives in these cases does not seem essentially different from the relation between the perspectives of two separate agents.

Second, if it is in virtue of the norm's applying that reviewing and reviewed perspectives are constituents of a single point of view, then we stand in need of some explanation for why the norm applies in some cases and not others. Why must there be a relation of normative directness in instances of self-directed critical reasoning? If we explain the application of the norm in terms of the perspectives being constituents of the self-same point of view, then we can appeal to that explanation in answering this question: there must be a relation of normative directness in these cases because instances of self-directed critical reasoning are instances in which reviewing and reviewed perspectives are constituents of the self-same point of view. But no such answer is forthcoming if we reverse the order of explanation. Why does the norm fail to apply in the case of *this* pair of token reviewing and reviewed perspectives? If it is in virtue of the norm's applying that reviewed and reviewing perspectives are perspectives of the self-same point of view, the answer cannot be because the reviewing and reviewed perspectives belong to separate agents. For that is simply to observe that the norm fails to apply. But if we explain the application of the norm in terms of the unity of perspectives, then this counts as a substantive answer to our question.

For these reasons, we should explain the application of the norm in terms of the unity of reviewing and reviewed perspectives and not the other way around.

4. Normative directness and the constitutive model

4.1. INTRODUCING THE CONSTITUTIVE MODEL

There are at least two questions we might have with respect to the normative directness requirement. First, we might wonder *why* there must be a

rationally necessary connection between reviewed and reviewing perspectives. Second, we might wonder *how* there could so much as be such a relation. My concern in the previous section was to offer an answer to the former question, an answer that was plausible both as an interpretation of Burge and as an interpretation of the connection between critical reasoning and normative directness considered in its own right. In summary, the answer was that there must be a relation of normative directness in those cases where reviewing and reviewed perspectives are perspectives of the self-same agent because one's reasonable judgments about what it is reasonable for one to believe bear constitutively on what it is reasonable for one to believe.

In this section, my focus will be on the second question – how could there so much as be a relation of normative directness between reviewing and reviewed perspectives? My answer will involve supplementing Burge's view with a substantive account of self-knowledge, an account that involves spelling out how it is that we could know our mental states non-empirically. The substantive account on which I'll focus is *reflectivism*. As we shall see, certain ideas that Burge develops in his more recent work on self-knowledge (Burge, 2013c) inform an account of reflection that is both *prima facie* plausible and congenial to the reflectivist view.

First, a note on terminology. Hereafter, I shall often talk of first-order and second-order perspectives in place of reviewed and reviewing perspectives. This is purely to ease discussion. I appreciate that these descriptors are not *strictly* interchangeable. A reviewing perspective may be third-order, for instance, a reviewed perspective second-order. I shall also talk about first-order and second-order *beliefs*, often as a proxy for 'beliefs formed at the level under review' and 'beliefs formed at the reviewing level', respectively. Again, I appreciate that it is perfectly possible for second-order beliefs to themselves be the targets of review.

What made it difficult to see how there could be a relation of normative directness between reviewing and reviewed points of view if the observational account is correct is that the account seemed to allow that in every case, my self-ascriptions might be reasonable and yet false. If we wish to know how a relation of normative directness might be so much as possible, a good place to begin, then, is by asking whether there is a way of thinking about self-knowledge, or about the relation between reviewing and reviewed perspectives more generally, which does not make this allowance – which recognises a class of self-ascriptions that cannot be reasonable and yet false.

There's a certain way of thinking about the relation between first-order and second-order perspectives that can make it appear puzzling how there *could* be a class of self-ascriptions that cannot be reasonable and yet false. According to this way of thinking, first-order and second-order perspectives stand to one another in much the same way as the levels of a building: like the levels of a building, these perspectives are contingently related in a host of interesting ways, but are nevertheless *constitutively distinct*. If we think

about first-order and second-order perspectives in this way, then it is bound to strike us as puzzling how any class of reasonable judgments made at the second-order level could be invulnerable to error. In every instance in which it is reasonably judged at the second-order level that one believes that *p*, it will seem a genuine possibility that one does not in fact believe that *p*. The guarantee that reasonable second-order beliefs will be true will appear no stronger than the guarantee that the levels of a building will align, given that the building is constructed according to the proper procedures.

But we need not think about first-order and second-order perspectives on the ‘building model’. Instead, we might see them as constitutively related. Indeed, the discussion towards the end of the previous section gave us some reason to think that this is precisely how Burge himself is thinking of them. A number of contemporary accounts of self-knowledge develop some version of the ‘constitutive model’, as I’ll call it.¹² An obvious example of such a view is *constitutivism*. Call a first-order belief ‘available’ if the subject is “poised” to assent to its content if the question of whether it is true arises, to use it as a premise in her reasoning, and to be guided by it in her behaviour’ (Shoemaker, 2012, p. 245). According to constitutivism, there is a constitutive relation between available first-order beliefs and those standing second-order beliefs that self-ascribe those first-order beliefs (Shoemaker, 2012, p. 239). Specifically, if one is rational and in possession of the relevant concepts, having an available belief that *p* is constitutive of having a standing second-order belief to the effect that one believes that *p*.¹³ On a view of this sort, it simply isn’t possible for one to have an available first-order belief and yet not believe that one has it, for the former belief is constitutive of the latter.

A second example of a view that develops the constitutive model is *reflectivism*, recently defended by Matthew Boyle. According to Boyle (2011, p. 6), ‘in the normal and basic case, believing *P* and knowing oneself to believe *P* are not two cognitive states; they are two aspects of *one* cognitive state – the state, as we might put it, of knowingly believing *P*’. Crucially, one can knowingly believe that *p* in Boyle’s sense without knowing *that one believes that p*. To knowingly believe that *p* is to be *subjectively aware* of believing that *p*. But to be subjectively aware of believing that *p* does not necessarily involve being in possession of a second-order belief to the effect that one believes that *p*. Subjective awareness is essentially *subjectless*. To be subjectively aware of *X-ing*, for example, is simply to be aware of *X-ing*. Contrast objective awareness, which includes a subject within its remit: to be (self-directedly) objectively aware of *X-ing* is to be aware of *one-self X-ing* (Boyle, 2017a).

Both subjective and objective awareness involve finding credible. But what it is that one finds credible differs, Boyle thinks, in the objective and subjective cases, respectively. To be subjectively aware of believing that *p* involves, simply – *finding it credible that p*. By contrast, to be

(self-directedly) objectively aware of believing that *p* involves *finding it credible that I believe that p*. Crucially, I can find it credible that I believe that *p* without finding it credible that *p*; I can be objectively aware of believing that *p* without being subjectively aware of believing that *p*. Suppose, for example, that I come to know, on the basis of observational evidence, that I believe that *S* is untrustworthy. Further, suppose that on reflection, I consciously decide that there is no good reason to believe that *S* is untrustworthy. In this case, I am objectively aware of believing that *S* is untrustworthy but not subjectively aware of believing it, in Boyle's sense.¹⁴

I said that on the reflectivist's view, one can knowingly believe that *p* without having a second-order belief to the effect that one believes that *p*. Still, there is an important relation between knowingly believing that *p* and knowing that you believe that *p*, according to the reflectivist. A subject who is subjectively aware of believing that *p* and who grasps the concept *belief* can come to form the second-order belief that they believe that *p*, 'not by acquiring some new piece of information, but simply by *reflecting* on what she already knows' (Boyle, 2017b, p. 13). In this case, coming to know that one believes that *p* is simply a matter of making explicit what is implicit in one's subjective awareness. Part of what it is to grasp the concept *belief* is to be able to affect this making explicit – to be able to form the belief that one believes that *p* simply on the basis of reflection on the subjective mode of one's awareness of the belief that *p*.

Could reflection give rise to error? If the nature of reflection really is as Boyle describes – if it really is simply a matter of making explicit what is implicit in one's subjective awareness – then it's plausible that self-ascriptions formed on the basis of reflection on one's subjective awareness of the relevant first-order state are *immune to error through misidentification*. If the self-ascription 'I believe that *p*' is the upshot of reflection on my subjective awareness, I cannot be right that *someone* believes that *p*, but wrong that it is *I* who believes that *p*. After all, the transition from what is implicit in my subjective awareness to an explicit self-ascription does not involve my *identifying* myself – or anyone else – as the person who believes that *p*. (Compare a case in which the self-ascription is the upshot of an inference from the following two premises: That man believes that *p*; and I am that man.) So no *misidentification* is possible. (See Recanati, 2012, on this point.)

Still, might reflection give rise to error about the *content* of one's belief? It may, but given a certain account of reflection – which I sketch below – any error is such as to undermine the reasonability of the resultant self-ascription. If a self-ascription is reasonable, and formed on the basis of reflection on one's subjective awareness, then it cannot involve an error about the content of one's beliefs. Note that it is not my concern to provide a full defence of this account. But I hope it will strike the reader as *prima facie* plausible. The account is informed by two ideas that find expression in Burge's more recent work on the nature of self-knowledge. The ideas

concern what Burge calls *purely preservative memory* and *betokening understanding*, respectively. Specifically, they concern the nature of the warrants that attach to these phenomena.

Reflection on one's beliefs involves holding those beliefs in mind. To this extent, it involves the functioning of memory. The sort of memory involved is what is commonly called *semantic memory* and what Burge calls *purely preservative memory*: roughly, the capacity to recall the mode and content of earlier psychological states (Burge, 2013c, p. 180, n. 63). On Burge's view, one's warrant to believe that *p* on the basis of remembering in the semantic mode that *p* does not depend on one's having reasons for thinking that one's memory is functioning reliably. I do not need evidence that my memory has not malfunctioned in order to be warranted in believing that *p* on the basis of recalling that *p*. My warrant to believe that *p* is, rather, an entitlement – a warrant that derives from the proper functioning of my memory system. Consequently, any failure of purely preservative memory to preserve the content (or mode) of an earlier state is such as to undermine the warrant, and hence the reasonability, of those beliefs formed on the basis of its deliverances (Burge, 2013c, pp. 201, 212).

In the course of reflecting on one's beliefs, one may misremember the content of those beliefs. But, if Burge is right, then to the extent that the error is traceable to an error in one's memory, or to carelessness or a lack of attention on the part of the subject, the error undermines the subject's warrant in – and hence the reasonability of – the resultant belief.

Of course, reflection doesn't just involve the functioning of memory. In those cases where reflection on one's subjective awareness gives rise to the self-ascription of an attitude, it involves, according to Burge, the functioning of *betokening understanding*. Betokening understanding performs a preservative role, but what it preserves is not simply the mode and content of a lower-level state over time, but the representational content of a lower-level state for use within a self-ascription. On Burge's view, self-ascribing a belief (at least in cases where the self-ascription is the upshot of reflection of one's subjective awareness of the belief in question) involves both thinking, and thinking about, the representational content of the lower-level belief:

Self-understanding uses canonical representation of presentational content. Such representation uses the representational content in the very representation of it. When one understands one's belief that Federer will win a match, canonical representation of the belief's representational content (that Federer will win a match) refers to the representational content, and also uses that representational content. (Burge, 2013c, p. 185)

Like semantic memory, betokening understanding fulfils a preservative function: it functions to preserve the content of a lower-level thought for use in self-ascriptions. And like semantic memory, the warrant with which it is associated is an entitlement rather than a justification. In order for the self-ascription 'I believe that *p*' to be warranted, one does not need reasons

or evidence for thinking that one's capacity for betokening understanding is functioning as it should – that it really is preserving the representational content of the lower-level state. Rather, the warrant is an entitlement that derives from the proper functioning of the capacity. But then, as before, any failure of the capacity to preserve the content of the lower-level state undermines one's warrant to the resultant self-ascription, for it is from the proper functioning of that capacity that one's warrant derives.

In the course of reflection, there may be a mismatch between the content of the belief that one self-ascribes and the content of the belief one in fact has. But, insofar as the error is traceable to an error in one's capacity for betokening understanding, or to carelessness or a lack of attention, the error undermines one's warrant in – and hence the reasonability of – the resultant self-ascription. Again, I don't claim to have provided a full defence of Burge's ideas here. But they inform a picture of reflection which is, I think, both *prima facie* plausible and congenial to the reflectivist's project. For our purposes, the crucial point is that the account of reflection that they inform is one according to which a self-ascription formed on the basis of reflection on one's subjective awareness of the relevant first-order state cannot be both reasonable and false.

4.2. NORMATIVE DIRECTNESS AND REFLECTIVISM

Both constitutivism and reflectivism constitute alternatives to the building model. My sympathies, however, lie with reflectivism, but not because it does a better job of making sense of how a relation of normative directness could obtain between reviewing and reviewed perspectives. An essential difference between reflectivism and constitutivism is that the former allows for the possibility of having a first-order belief without having a corresponding second-order belief. And this, I think, is a point of difference in favour of reflectivism. In the discussion that follows, my focus is going to be purely on reflectivism. But what I say about the way in which reflectivism makes plausible the idea of normative directness can, I believe, be extended relatively unproblematically to constitutivism (and indeed to various other models of self-knowledge which in one way or another honour the idea that first-order and second-order points of view are constitutively related). My focus will be on *prima facie* (as opposed to ATC) reasonableness, on demonstrating that reflectivism shows us how it could be the case that it sometimes follows necessarily and immediately from my reasonably judging that it is *prima facie* reasonable for me to believe that *p*, that it *is prima facie* reasonable for me to believe that *p*. I believe that the analogous demonstration concerning ATC reasonableness is also possible, but I shall not try to provide that demonstration here.

We want to know whether there is a class of self-ascriptions that cannot be reasonable and yet false, if reflectivism is true. I said that on the

reflectivist's view, a subject who is subjectively aware of believing that *p*, and who grasps the concept *belief*, can come to form the second-order belief that she believes that *p* simply by reflecting on the content of her subjective awareness. I sketched an account of reflection according to which second-order beliefs thus formed cannot be both reasonable and yet false. But, then, plausibly, for any conceptually sophisticated subject, there is going to be a class of second-order judgments that cannot be reasonable and yet false and that self-ascribe first-order states – namely, those self-ascriptions that are grounded in careful reflection on one's subjective awareness. If I am reflecting carefully, then I will form, on the basis of my subjective awareness, the belief that I believe that *p* only if I do knowingly believe that *p*. So, if my second-order belief is reasonable, my having it guarantees its truth.

Is this sufficient to show that if reflectivism is true, it must sometimes follow necessarily from my reasonably judging that it's *prima facie* reasonable for me to believe that *p*, that it is *prima facie* reasonable for me to believe that *p*? Not quite. Consider that my judgment that I have such-and-such a first-order belief might be true and yet my judgment that this first-order belief is reasonable might be false. I might correctly think that I believe that *q* but incorrectly (yet reasonably) think that I am being reasonable in believing that *q*. Suppose that I *do* incorrectly believe that my belief that *q* is reasonable. And suppose that this second-order belief is a premise on which my judgment that it's *prima facie* reasonable for me to believe that *p* is based. Then it *doesn't* follow immediately or necessarily from that judgment that it is reasonable for me to believe that *p*. It doesn't follow immediately or necessarily because the judgment is based on a false belief about what I reasonably believe.

Suppose it can be shown that reflectivism makes plausible the idea that there is a class of second-order judgments about whether certain of my first-order beliefs are reasonable which cannot be both reasonable and false. Will we then have shown that if reflectivism is true, it must sometimes follow necessarily from my reasonably judging that it's *prima facie* reasonable for me to believe that *p*, that it is *prima facie* reasonable for me to believe that *p*? Yes. If it doesn't follow immediately and necessarily, that can only be because it is an in-principle possibility that either, first, I am wrong about what I believe or, second, I am wrong about whether I am reasonable in believing it. If it is not an in-principle possibility, then necessarily it is *prima facie* reasonable for me to believe that *p*.

So does reflectivism make plausible the idea that there are second-order beliefs about whether certain of my first-order beliefs are reasonable which cannot be both reasonable and false? I want to suggest that it does. Consider the following judgment: I am being reasonable in believing that *p*. There are two ways in which this judgment could itself be reasonable and yet false. First, it could be based on a false, yet reasonable, second-order judgment

that self-ascribes a first-order belief. Second, it could be based on a faulty inference.^{15,16} Let's consider each of these possibilities in turn.

Suppose that my judgment that I am being reasonable in believing that *p* is based on the judgment that I (reasonably) believe that *q* and if *q* then *p*. If I don't in fact believe that *q*, say, but am reasonable in believing that I do, then the judgment that I am being reasonable in believing that *p* may well be false though reasonable. In this case, my judgment that I am being reasonable in believing that *p* would be false in the first way described. But it's not clear that this case is possible, given that reflectivism is true, at least not if I form the belief that I believe that *q* on the basis of careful reflection on the content of my subjective awareness. If I am reflecting carefully on the content of my subjective awareness, I will (for the reasons given previously) form the second-order judgment that I believe that *q* only if I do in fact knowingly believe that *q*. If I don't knowingly believe that *q*, I cannot, on the basis of reflection on the content of my subjective awareness, reasonably form the belief that I do.

We can put the basic point as follows. If reflectivism is true, the judgment that I am being reasonable in believing that *p* cannot be false (yet reasonable) *for the reason* that it is based on a false yet reasonable second-order belief about certain of my first-order beliefs of which I am subjectively aware. For if reflectivism is true, a reasonable belief about certain of my first-order beliefs of which I am subjectively aware cannot be false.

Now, suppose that my judgment that I am being reasonable in believing that *p* is based on a faulty inference. For example, suppose that I have inferred it from the judgment that I believe that *q* and that if I believe that *p* then I ought to believe that *q*. Might my judgment that I am being reasonable in believing that *p*, although false, be itself reasonable? I don't think so and not for reasons that presuppose the truth of reflectivism. Earlier, I said that a reached conclusion is reasonable if and only if it is reached in a way that is guided by the norms of critical reasoning, where these norms include, but are not limited to, the norms of deductive and inductive logic. By definition, a faulty inference violates the norms of reasoning. Consequently, a conclusion reached by way of a faulty inference cannot be reasonable. *A fortiori*, I cannot be reasonable in judging that I am being reasonable in believing that *p* if that judgment is the result of a faulty inference.

So, in summary, it appears that if reflectivism is true, then second-order judgments that, first, self-ascribe, on the basis of careful reflection on one's subjective awareness, particular first-order beliefs and, second, assert, of those first-order beliefs self-ascribed, that they are reasonable, cannot be both reasonable and false. But earlier I suggested that this outcome is sufficient for normative directness. So, if reflectivism is true, there must be a relation of normative directness between first-order and second-order perspectives.

5. Conclusion

The aim of this paper has been to explicate and defend the Burgean argument for agentialism and in particular the claim that critical reasoning requires a relation of normative directness between reviewing and reviewed perspectives. I've argued that, plausibly, normative directness is implied (and that Burge takes it to be implied) by norms governing self-directed critical reasoning. We can come to see how a normative relation between reviewing and reviewed points of view might be so much as possible by supplementing Burge's view with substantive, reflectivist thinking about how it is that we come to know what we believe.

Interestingly, reflectivism also seems well positioned to explain the phenomenon of transparency, which as noted in the introduction to this paper is central to Moran's substantive strategy for defending agentialism. To knowingly believe that *p* is to be subjectively aware of believing that *p*, where part of what it is to be subjectively aware of believing that *p* is to find the proposition that *p* credible. But then it stands to reason that I can find out whether I believe that *p* by asking whether (it is credible that) *p*, for in settling the question whether (it is credible that) *p*, I am settling the question whether I consciously believe that *p*. (See Boyle, 2011, p. 236, for a development of this line of argument.) If reflectivism *does* explain the phenomenon of transparency, then it may prove valuable in thinking about what, if anything, is common to Moran's and Burge's respective arguments for agentialism. But that is a topic for another paper.¹⁷

The Queen's College
University of Oxford

NOTES

¹ By 'empirical knowledge' I mean knowledge our warrant for which is grounded in, for example, evidence or reasons or in a reliable belief-forming process. See Gertler (2016) for a drawing of the distinction between empiricist and agentialist accounts of self-knowledge in terms of empirical and non-empirical warrant.

² Throughout, by 'mental states' I mean the propositional attitudes – beliefs, desires, intentions and so on. The question of how agentialists might account for knowledge of, say, our sensations is an interesting one, but I do not pursue it here.

³ To be clear, I am not suggesting that the agentialist *must* argue for her view in either of these two ways. Perhaps there are ways of arguing for agentialism that don't fall neatly into either the substantive or non-substantive camp, as characterised here. Nor am I suggesting that the only tenable ways of pursuing the substantive and non-substantive strategies are Moran's and Burge's, respectively. I introduce the substantive/non-substantive distinction simply to characterise two common ways in which agentialism has been argued for.

⁴ Granted, this is stronger than the claim that Burge makes in the passage cited directly previously. However, there are passages in which Burge commits himself to this stronger claim. See, for instance, the passage cited in Section 3 of this paper.

⁵ This is not to say that there aren't *any* differences at all. As an anonymous referee points out, even if the observational account is correct, we can expect agents to have better access to the body of background information against which their own beliefs are formed than to the body of information against which someone else's beliefs are formed. Given that what it is reasonable for one to believe is in part a function of one's background information, we can expect closer alignment between an agent's judgments about what it is reasonable for her to believe, and what it is in fact reasonable for her to believe, than between an agent's judgments about what it is reasonable for someone else to believe, and what it is in fact reasonable for that person to believe. I grant this difference, but it doesn't show that a broadly observational account is consistent with normative directness. For it remains the case that if the observational account is correct, my judgments about what it is that I believe – and indeed my judgments about the body of background information against which my beliefs were formed – may always be reasonable and yet false.

⁶ It is not obviously true, for instance, of Alex Byrne's view. According to Byrne, one knows one's beliefs by reasoning in accordance with what he calls the *doxastic schema*: *p*; therefore I believe that *p* (Byrne, 2011). For Byrne, the relevant knowledge-conducting feature of this form of reasoning is its reliability – the second-order beliefs that the inference generates could not easily have been false. In fact, the schema is 'strongly self-verifying' in the following sense: 'If one reasons in accord with the doxastic schema, and infers that one believes that *p* from the premise that *p*, then one's second-order belief is *true*, because inference from a premise entails belief in that premise' (Byrne, 2011, pp. 206–7). Is Byrne's view vulnerable to the objection from normative directness? Second-order beliefs formed in accordance with the doxastic schema cannot be false. So the important question is whether in forming one's belief in such a way one is being *reasonable*. I won't try to settle this question here. But see Boyle (2011) for reasons to think that one is not.

⁷ As the passage makes clear, Burge thinks that something stronger is true – that normative directness is partly *constitutive* of critical reasoning. I address the significance of this way of putting the point presently.

⁸ Note that such judgments are not the *only* things that can make it (*prima facie*/ATC) reasonable to believe that *p*. If, for example, one believes that *q* and that if *q* then *p*, then it is *prima facie* reasonable for one to believe that *p* in virtue of one's extant first-order beliefs. One does not first need to form the relevant second-order beliefs. But if one *does* form the relevant second-order beliefs, then the *prima facie* reasonableness of one's believing that *p* is, as it were, overdetermined.

⁹ Another person's reasonable judgments about what it is reasonable for me to believe *may* bear constitutively on what it is reasonable for me to believe, but only in extenuating circumstances. They do not do so necessarily. They do so only given the satisfaction of certain further conditions, for example, that the person knows me extremely well, that they are well versed with respect to my background information and so on.

¹⁰ In the case of ATC reasonableness, at least, there seems to be a further rationale. Suppose, for instance, that a reasonable judgment from the reviewing perspective that it is ATC reasonable to believe that *p* made it ATC reasonable to believe that *p* from the perspective under review, even though the reasonable second-order beliefs on which that judgment was based could be false. In this case, that judgment would, presumably, make it ATC reasonable to believe that *p* even if some or all of the second-order beliefs on which it was based were *in fact* false. But this is not plausible. If this supposition was true, then what it is ATC reasonable for me to believe would depend *exclusively* on what I reasonably judge it is reasonable for me to believe. Facts about my reasonable first-order beliefs would be irrelevant. But these facts are not irrelevant. Whether it is ATC reasonable for me to believe that *p* depends in part on what my reasonable first-order beliefs in fact are. (See Burge's remarks on the 'buck-stopping status' of first-order states in his Burge, 2013c, pp. 216, 175–8.)

¹¹ This certainly seems to be Burge's own view: '[M]y checking my belief and finding it wanting normally *itself* provides immediate *prima facie* reason to change it from within the

perspective of the review. This is because the first-order and second-order perspectives are the *same* point of view' (Burge, 1996, p. 110, emphasis in original).

¹² See, for example, Shoemaker (1994, 2012), Wright (2001a), Heal (1994) and Boyle (2009, 2011, 2017a, 2017b). I leave to one side expressivist and neo-expressivist views of self-knowledge, although they too constitute an alternative to the 'building model'. Arguably, they do so at the expense of an in-principle distinction between first-order and second order perspectives. For a defence of neo-expressivism, see Finkelstein (2003) and Bar-On (2004). For a discussion of the relation between neo-expressivism and transparency, see Bar-On (2015) and Finkelstein (2012).

¹³ Constitutivism thus understood should be delineated from another view that sometimes gets labelled 'constitutivism' in the literature, the view that there is a constitutive relation between a subject's believing that *p* and their avowing, or being disposed to avow, a belief that *p*. See, for example, Wright (2001a, 2001b). For discussion of the relation between Wright's and Shoemaker's views, see Bar-On (2009).

¹⁴ Note, however, that the mark of objective awareness is not that one *in fact* fails to find it credible that *p*, but that one's awareness is in principle consistent with one's failing to do so.

¹⁵ There is in fact a third possibility: the judgment could be based on the false, yet reasonable, second-order judgment that I am being reasonable in holding some specific first-order belief. Of course, we would then want an explanation of how *that* judgment came to be false yet reasonable. At some point, our explanation must terminate either in a false but reasonable second-order judgment that ascribes a first-order belief or in a faulty inference.

¹⁶ One might wonder about cases of the following kind. Suppose my first-order belief that *p* is in fact inconsistent with some other first-order belief of mine, the belief that *q*, say. But suppose that the inconsistency requires a degree of self-reflection for which I simply don't have the time. In reflection, I reasonably, but falsely, judge that there is no inconsistency and on that basis, that it is reasonable for me to believe both that *p* and that *q*. Is this not a case in which my second-order belief about what it is reasonable for me to believe is both reasonable and yet false, albeit not in either of the two ways just outlined? I don't think so. It is so, it seems to me, only if we permit different standards of reasonability to govern the first-order and second-order perspectives, respectively. The claim, effectively, is that constraints on one's time are sufficient to render reasonable the second-order judgment that it is reasonable for one to believe both that *p* and that *q*, but not sufficient to render reasonable one's believing both that *p* and that *q*. And I don't see why this should be so. If constraints on one's time can affect what it is reasonable for one to believe at the second-order level, I don't see why they can't do the same at the first-order level.

¹⁷ My sincere thanks to Casey Doyle, Lucy Campbell and two anonymous referees at *PPQ* for comments on an earlier draft of this paper.

REFERENCES

- Bar-On, D. (2004). *Speaking My Mind: Expression and Self-knowledge*. Oxford: Clarendon Press.
- Bar-On, D. (2009). 'First-person Authority: Dualism, Constitutivism, and Neo-expressivism,' *Erkenntnis* 71(1, First Person Authority), pp. 53–71.
- Bar-On, D. (2015). 'Transparency, Expression, and Self-knowledge,' *Philosophical Explorations* 18(2), pp. 134–152.
- Boyle, M. (2009). 'Two Kinds of Self-knowledge,' *Philosophy and Phenomenological Research* 78(1), pp. 134–164.
- Boyle, M. (2011). 'Transparent Self-knowledge,' *Proceedings of the Aristotelian Society, Supplementary Volume* 85(1), pp. 223–241.
- Boyle, M. (2017a) 'Self-knowledge, Self-concern and the First-person Perspective' (manuscript)
- Boyle, M. (2017b) 'Transparency and Reflection' (manuscript)

- Burge, T. (1996). 'Our Entitlement to Self-knowledge,' in *Proceedings of the Aristotelian Society*, Volume 96, New Series, pp. 91–116.
- Burge, T. (2013a). 'Epistemic Warrant,' in T. Burge (ed.) *Cognition through Understanding: Philosophical Essays*, Volume 3. Oxford: Oxford University Press, pp. 521–533.
- Burge, T. (2013c). 'Self and Self-understanding: The Dewey Lectures (2007, 2011),' in T. Burge (ed.) *Cognition through Understanding: Philosophical Essays*, Volume 3. Oxford: Oxford University Press, pp. 140–226.
- Byrne, A. (2011). 'Transparency, Belief, Intention,' *Proceedings of the Aristotelian Society* 85(1), pp. 201–221.
- Finkelstein, D. (2003). *Expression and the Inner*. Harvard: Harvard University Press.
- Finkelstein, D. (2012). 'From Transparency to Expressivism,' in G. Abel and J. Conant (eds) *Rethinking Epistemology* (De Gruyter, 2012, pp. 101–118.
- Gertler, B. (2016). 'Self-knowledge and Rational Agency: A Defense of Empiricism,' *Philosophy and Phenomenological Research* 96(1), pp. 91–109.
- Heal, J. (1994). 'Moore's Paradox: A Wittgensteinian Approach,' *Mind*, New Series 103(409), pp. 5–24.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-knowledge*. Princeton: Princeton University Press.
- Recanati, F. (2012). 'Immunity to Error through Misidentification: What It Is and Where It Comes from,' in S. Prosser and F. Recanati (eds) *Immunity to Error through Misidentification: New Essays*. Cambridge: Cambridge University Press, pp. 180–201.
- Shoemaker, S. (2012). 'Self-intimation and Second-order Belief,' in D. Smithies and D. Stoljar (eds) *Introspection and Consciousness*. Oxford: Oxford University Press, pp. 239–258.
- Shoemaker, S. (1994). 'Self-knowledge and "Inner Sense",' in *Philosophy and Phenomenological Research*, Volume 54, pp. 249–314.
- Wright, C. (2001a). 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention,' in C. Wright (ed.) *Rails to Infinity*. Cambridge, MA: Harvard University Press, pp. 291–318.
- Wright, C. (2001b). 'Wittgenstein's Rule – Following Considerations and the Central Project of Theoretical Linguistics,' in C. Wright (ed.) *Rails to Infinity*. Cambridge, MA: Harvard University Press, pp. 170–213.