

Final Version: O'Brien, L. (forthcoming), *Philosophy and Phenomenological Research*.

***Self-knowledge, Agency and Force*¹**

1. Introduction

My aim in this paper is to articulate further what may be called an agency theory of self-knowledge. Many theorists have stressed how important agency is to self-knowledge, and much work has been done drawing connections between the two notions.² However, it has not always been clear what *epistemic* advantage agency gives us in this area and why it does so. I take it as a constraint on an adequate account of how a subject knows her own mental states and acts, that it construe the known mental states and acts realistically and as independent of their self-ascription, and that it deliver genuine epistemic standing to the knower. The main task of the paper will, then, be to explore how our having rational agency with respect to our mental states may be able to secure genuine epistemic warrant for our self-ascriptions of states or acts independent of the ascriptions. This task will be carried out by focussing on the question of what account we should give of our knowledge of what I call our acts of judging. In the remainder of this section, I will do a little to clarify what is meant by that question. Section 2 will attempt to introduce us to elements of the best way to approach the question by considering some alternative strategies. Section 3 is devoted to forming some idea of what *kind* of warrant we are looking for when considering how agency might give us self-knowledge. Section 4 aims to present a suggestion as to how agency gives us the kind of warrant identified over our acts of judging. Section 5 deals with some objections.

When we ask about our knowledge of psychological self-ascriptions of the form: I φ that P , there are three aspects of the question that our answers can focus on, corresponding to the three components of the ascription. I can ask: (i) how do I know that I φ that P ? (ii) how do I know I φ that P ?, and (iii) how do I know I φ that P ?³ The last question, which asks how I know that P is the content of my attitude, has been discussed extensively in recent years. It has, for example, been asked how, given non-individualism about content, do I know that I have an attitude with respect to water rather than twater?⁴ I will for the purposes of this paper simply assume that we can explain the authority a subject has over the contents of her attitudes by using the fact that, in knowing what attitude she has, the subject is re-deploying an assumed ability to think P .⁵ It is the first two questions, particularly the second, that will concern me: How do I know that it is me, rather than someone else, who has the attitude towards P , and more particularly, how do I know that I φ , rather than ϕ , that P ? It is in attempting to answer these questions that I think the agency theory of knowledge proves its worth.

I will call the relation in which a subject stands to a content, 'the force' of the subject's attitude. There are of course very many relations in which a subject can stand to a content P . A subject can, for example, believe, think, entertain, judge, question, doubt, fear, desire, want, wish, fancy, that P . To consider all these different forces would involve us in more extensive a task than could reasonably be attempted here.⁶ My concern will be with only those forces we can call purely cognitive forces. I will further narrow the focus by considering those forces that characterize mental acts or events rather than states, and which we can call assertoric.

We can divide the purely cognitive acts of mind into three main groups as follows:

(i) Non-committal (acts of mere thinking):

The thought that P can be entertained, apprehended, understood.

(ii) Assertoric (acts of judging):

The thought that P can be judged, denied, questioned, doubted.⁷

(iii) Supposition (acts of supposition):

The thought that P can be supposed, assumed, fancied, imagined.⁸

Acts of merely thinking involve no stance by the subject with respect to the truth of P . They are, as Frege put it early in his early work, 'mere complex[es] of ideas'.⁹ Acts of judging imply a stance on the part of the subject as to the truth of P . In particular, they imply a commitment on the part of the subject in relation to the actual truth of P : a commitment to P being true, or not being true, or being unsettled or being unlikely true. Acts of supposition also imply a stance on the part of the subject to the truth of P . For, in assuming P the subject treats P as if true. However, in contrast to acts of judging, they do not imply a commitment on the part of the subject in relation to the actual truth of P . They are cases in which the subject acts as if they had commitment with respect to the actual truth of P when they do not.¹⁰ The task adopted in this piece will be to consider the question of how we know the *assertoric* force of our attitudes.

2. How do we know the assertoric force of our attitudes?

So how is it that we know that we are judging, denying, questioning and doubting?

Do we know by knowing the marks that distinguish them? What marks are those?

It is clear that attitudes with different force have different causal roles with respect to behaviour. I act very differently if I judge *P* to be true than if I deny it. It is also pretty clear that the basis of my knowledge that I am judging *P*, rather than denying *P*, is not based on my observing that I am prepared to act as if *P*, rather than not. While the immediate authority we have over our mental states may at times have been exaggerated, it is clear that my knowledge that I am judging that *P*, rather than denying that *P*, does not typically have to await the observation of any action on my part.

Attitudes with different force also have different rational roles. My judgement that *P* will give rise to different future judgements and denials than my denial that *P*. It would clearly be circular to suggest that I know that I judged *P* at time *t* because of knowing that I judged *Q* at time *t'*. But perhaps my knowledge that I judge, rather than deny, *P* is based on my tracking the content transitions: could I know that I judged *Q*, by tracking the transition from the contents '*P*' and '*P* implies *Q*' to '*Q*'? Again, sensitivity to content alone will not ground such knowledge. I need to know first that I hold those contents to be true.

The suggestion that we know that we believe *P* by being sensitive to its phenomenological marks has never had much appeal. The suggestion with respect to our mental acts and activities has, I think, more to be said in its favour. It does seem to be a feature of those mental phenomena we describe using the vocabulary of action and activity that they are phenomenologically accessible to us, that they figure in the stream of consciousness. For William James most thought is accompanied by varying degrees of striving and relief. The attempt on the part of a subject to reach an opinion

is accompanied by a feeling of effort, of striving and the resulting judgement by a sense of triumph, calm or easy relief. Perhaps this way of seeing things provides us with the marks by which we are able us to know that we are questioning rather than judging?

James distinguishes three kinds of mental activity:

- (a) Bare activity
- (b) Directed Immanent Activity
- (c) Directed activity with resistance.

Bare activity is described as having 'no definite movement, no actor and no aim. Mere restless zigzag movement'.¹¹ It is hard to know what in our mental lives is characterised as 'barely active': perhaps dreaming or random thought impressions. What is clear is that it is less than idly thinking things through, for of the other two kinds of activity James says:

activity...is either aimless or directed. If directed it shows tendency. The tendency may or may not be resisted. If not we call the activity immanent, as when the body moves in empty space by its momentum or our thoughts wander at their own sweet will. If resistance is met its agent complicates the situation. If now, in spite of resistance, the original tendency continues, effort makes its appearance, and along with effort, strain or squeeze.¹²

So in thinking, for James, 'the experiencer...feels the tendency, the obstacle, the will, the strain, the triumph, or the passive giving up'¹³. In unimpeded rational thought one

can assume that the subject moves effortlessly through a series of judgements. In questioning, and enquiring the subject experiences a tension, and striving that is only relieved by the forming of a judgement or denial. So, the move from questioning to judging is a move from restless straining to relief. It might be thought that the experience of tendency with absence of 'strain or squeeze' is our experience of judgement, and that it is our experience of this straining and relief that enables us to know that we move from questioning to judging.

I trust that I will not meet much more resistance, in finding this account of our knowledge of the force of our attitudes wanting, than I would if rejecting the Humean suggestion that we could know the force of our attitudes by discriminating the degree and nature of their forcefulness or vivacity.¹⁴ The problems are extensive: what account am I to give of my knowledge of these marks of my mental activities? Are they sufficiently tightly correlated with the attitudes themselves for knowledge of them to secure the authority we have over our judgements? Judgements do not always come with relief, but can occasion high emotional strain □ how am I able to distinguish one kind of relief from another? But it is not my aim to settle against such an account here. Rather, for all its prima facie problems, I want to suggest that there is something right about the Jamesian story and that is that the answer to how we know the force of our attitudes flows from a consideration of them as active.

What of the suggestion that we know the force of our attitudes in virtue of a causal mechanism that reliably converts the force of a first order state or act to the content of a second order state or act? So that we know that we are φ -ing that P , because generally, φ -ing that P causes the belief that I am φ -ing that P . I will not argue against

such an account here. I take it that enough critical work has been done elsewhere to make unnecessary an extensive discussion of this suggestion. It has been a well made charge that causal models of this kind do not secure the rational authority and control we have over our mental states.¹⁵ These accounts precisely leave out the rational agency that is critical to explaining self-knowledge. It is however worth pointing out that once we make clear that the same content can occur with different forces, it becomes obvious that the causal theorist will have to postulate a causal mechanism sensitive, not just to the content of the attitude, but also to its nature. The causal theorist must explain how we get from questioning that *P*, rather than, say, denying that *P*, to the judgement 'I question that *P*'. We cannot make do with a simple re-deployment of content that was operative at the first level and a single self-ascription operator.

If the causal theorist has in general been found to tell too mechanistic a story about self-knowledge, what we can call, the conceptualist about self-knowledge tells one that is too intellectualised.

The conceptualist, as I envisage him, argues that our concepts of our mental acts are governed by certain rules, the application of which by the subject can secure for them knowledge of those mental acts. In knowing the concepts 'judge', 'question' and 'deny' the subject grasps rules that govern the concepts. As such they know that it is acceptable for someone to judge 'I judge that *P*' on finding *P* to be true, to judge 'I deny that *P*' on finding *P* not to be true, and to judge 'I doubt *P*' on finding the truth of *P* to be doubtful. And the subject knows that it is not acceptable for someone to judge 'not-*P* and I judge that *P*' or '*P* and I deny that *P*', or '*P*, but I doubt *P*'. So, on this

story, one knows, for example, what one judges in virtue of using something like the following rule: whenever one finds P to be true, one is ipso facto in a position to judge 'I judge that P '. Let us call this simple rule for the self-ascription of judgement RJ.

On the face of it applying RJ will do little to help us explain our knowledge of the assertoric force of our acts of judgement but rather seems to rely on it. The successful application of the rule *by the subject* seems to rely on their ability to know that they have satisfied the antecedent clause. So, it will require the subject to know that they find P to be true which does not seem very different from what we are trying to explain. It is also difficult to see how our conceptualist will be able to frame rules that are supposed to guide our attempts to come to know what we are denying or questioning, that are not similarly non-explanatory. Suppose the conceptualist were to offer rules similar to RJ for denying and questioning or doubting, let us say:

(RD) Whenever you find P not to be true, you are ipso facto in a position to judge 'I deny that P '.

(RDt) Whenever you find the truth of P unsettled or unlikely, you are ipso facto in a position to judge 'I question or doubt that P '.

However, again, we would then owe an explanation of the subject's knowledge that she finds P not to be true, or to be unsettled or unlikely. It seems that if an attitude concept figures in the antecedent clause of the rule, the subject is going to need to know they have the attitude if they are to apply the rule. To get reference to the attitude out of the formulation of the rules one is tempted to offer something like:

Whenever P is true, you are ipso facto in a position to judge 'I judge that P '. But that is simply false.

It appears that abstract mastery of the concepts and the rules that govern them, without an assumed capacity to know our judgements immediately and spontaneously is not sufficient to explain our normal self-knowledge. This can be further brought out by considering George. George is a conceptual sophisticate, who has full mastery of the rules that govern his psychological concepts but knows about himself only via observation of his behaviour. George observes himself asserting P and on that basis holds that he finds P to be true. Given an application of RJ, George can come to know 'I judge that P '. George does not, it seems safe to claim, know his judgements as we know ours. Suppose he realises that RJ only gives self-knowledge when applied on the basis of sincere assertions, and starts to worry about whether his assertion was sincere. The application of RJ it seems, can only model the way we know our judgements if it is assumed that we know we are judging in the way that we are seeking to explain.¹⁶

Of course there is a real reason to doubt that George is a genuine possibility, to doubt that he can properly be said to be judging at all if he needs to appeal to such sources of self-knowledge. But imagined George does suggest that the source of ordinary self-knowledge does not lie in the rules that a subject takes to govern her concepts. If George *were* possible we might think he would have that source still available.

There is a further worry with the conceptualist as thus construed. It is far from clear that the abilities to formulate or apply such rules as the conceptualist takes to be the

means of self-knowledge, are even necessary for self-knowledge. As long as we are considering, and for the most part we have been, what allows a subject knowledgeably to judge 'I ϕ that P ' we can assume that the subject has the concept ' ϕ ' and so has some tacit grasp of the rules that govern it. However, such a subject need not have these rules ready to hand. To think of the application of those rules as the means by which a subject gains effortless self-knowledge is likely to be a mistake. Further, we might think that a highly deliberative subject should be granted knowledge or awareness of her mental states even prior to full-blown mastery of the psychological concepts of judgement etc. Consider Peacocke's example of a child who believes that his toy is in the cupboard, on the basis that this is where it is normally kept. The child can come to find his reason insufficient and revise his belief by having the purely world directed realization that Auntie is staying and puts toys in the wrong place. He may not have the conceptual sophistication to be granted full blown mastery of the psychological concepts of judgement etc. but he clearly exhibits sensitivity to his reasons and to the nature of his attitudes and may be thought of as having self-knowledge of some kind.

However, there is clearly something right about appealing to the rules and conceptual connections that the conceptualist brings in when attempting to give an account of self-knowledge. The mistake was to think of self-knowledge as the result of the subject themselves grasping and applying the rules. The important thing is to see the rules as rules that the subject acts in conformity with. In particular, it is true that when a subject judges, or finds, P to be true they have, standardly, ipso facto what is needed to know that they judge P , which they can then express using a self-ascription if they have the requisite concepts. This is different from when a subject forms an ulcer, or a

depression on her chair cushion. In the former case the subject doing something, judging that P , seems to secure, or be able to secure, knowledge that she did that thing. This connection is the heart of what I will call 'agency theories of self-knowledge'. But while this may be in some way right it is not clear how and why *doing ϕ* gives us *knowledge* that we are doing ϕ ?

We might, at this point, think of a pared down conceptualist agreeing that, for instance, while the concept of judgement is governed by rules such as RJ, possession of the concept of judgement does not require the subject to know or be able to frame such rules. Rather, they may say, conceptual mastery partly consists in the subject's being, in general, prepared immediately to self-ascribe her judgements when judging. Further, the suggestion is, this is all that is needed in order for the subject to be granted the ability, in general, to know her own judgements.¹⁷

I think it is probably right to say that mastery of this kind, with these application conditions, is sufficient for knowledge of ones judgements. However, unlike in the case of our earlier conceptualist, it is hard to see that we have here an *explanation* of such knowledge or that we have reason to suppose that the subject's conceptual mastery is the significant source of self-knowledge. We know that having the concept is sufficient to secure knowledge of her judgements, but it is most natural to assume that the source of such knowledge lies in facts about the judging itself which are reflected in the concept. Let us go back to Peacocke's example of the child. It seems plausible that the phenomenon of reason led judgement puts in place essential epistemic elements prior to mastery of the concept of judgement, which would function as part of any subsequent explanation of the child's capacity to know his own

states of mind. Whether there is available an account of self-knowledge that is rooted in the nature of the attitudes known and their relation to the knowing subject, rather primarily in features of the subject's self-ascriptions will concern as for much of what follows.

3. Epistemic Framework: Rational Entitlement

Before attempting to explicate what I take to be the core element of the agency theory of self-knowledge, I want to back up a little to get clear about what would count as success in this area.

The suggestions raised so far have, for the most part, attempted either to provide us with an account of what justification the knowing subject might have to think that they are ϕ -ing that P , or have contended that the reliable correlation of the subject's belief that they are ϕ -ing with their ϕ -ing will suffice for self-knowledge. However, in recent years perhaps the most significant move to be made in epistemology in general, and in discussion of self-knowledge in particular, has been the development of the notion of epistemic entitlement. Theorists have moved towards an epistemology that, on the one hand, denies that subjects have knowledge only if their beliefs or judgements are justified by inference or evidence, and on the other hand eschews the idea that mere reliability, with no rational connection between what is known and the knower, can secure warrant or knowledge. The key move made is to allow that there are transitions that are *rationally* warranted but that are not justifiable with reference to reasons statable by, or comprehensible to, the knowing subject.¹⁸ I will, following Burge, take warrant to be the genus under which entitlement and justification fall as sub-species, and take warrant to be necessary for knowledge. Assuming that warrant

will not accrue from justification, the question we can now set us ourselves is: what is it about the transition from ϕ -ing that P to judging 'I ϕ that P ' that makes it one to which the subject is rationally entitled?¹⁹

Let us take any transition from B (Base) to C (Conclusion). B and C are mental states or acts. Even within entitlement theories there are critical differences in how warrant is supposed to accrue to C. Some accounts of warrant by entitlement are relatively conservative in looking for some aspect of B that will play substitute in the role played by content in warrant by justification. Others allow that warrant by entitlement does not have to be explicated with reference to B in particular, but can come from the role played by C. We can call those theories that explain warrant primarily with reference to B, 'bottom-up' theories, and those that explain warrant with reference to the rational, epistemic or other role of C, 'top-down'.

1. Bottom-up Theories.

A bottom-up theorist holds that C is warranted in virtue of the content and/or nature of B. Such an account allows that while the subject might not take B as the basis of an inference or as evidence for the concluding state, nevertheless the transition from B to C is explained as warranted due to features of B and its content.²⁰

2. Top-down Theories

The top-down theorist holds that C is warranted either in virtue of in some way making itself true or in virtue of its conceptual, rational, epistemic or other role.

Accounts such as Wright's (Wright (1998)) that claim that making the judgement that I ϕ that P is part of what constitutes my judging that P is offering a top-down warrant.

What of the pared down conceptualist discussed earlier? He holds that the source of self-knowledge lies in the fact that mastery of the concept of judgement means being reliable in ones self-ascriptions. To hold that this fact constitutes the warrant needed to make her self-ascriptions knowledgeable can be seen as a top-down view. Burge and Shoemaker hold that the role that self-knowledge plays as a pre-condition of critical reasoning is itself sufficient to explain what makes our self-ascriptions of belief states ones to which we are entitled. (Burge (1996), Shoemaker (1996)). A theorist who held that such entitlement is the only kind of entitlement for our self-ascriptions would be a top-down theorist. Bilgrami (Bilgrami (1998)) who argues that we must accord a subject self-knowledge given the role it plays in our being seen as responsible agents can be seen as offering a top-down warrant for it. So called 'framework judgements' might also be thought to be judgements that accrue top-down warrant: we are entitled to them in virtue of the role they play in our being critical epistemic beings at all.

3. Mixed Theories

(a) One kind of mixed theorist holds that the concluding state secures independent and distinct kinds of warrant: both from its rational, epistemic, or other role in the system and from the content and nature of the base state. I take it that anyone who recognises that there are forms both of bottom-up and top-down warrant may think that both kinds of warrant will play a role in most knowledge states.

(b) A more complex kind of mixed theorist will hold that the concluding state secures warrant only from the content and nature of the base state, *but* will hold that the nature of the base state, and thus its ability to function as giving warrant to the

concluding state, depends upon its rational, epistemic or other role in the system. This is a theorist who does not recognise sources of top-down warrant that are independent of the warrant conferring role played by the base state. The warrant of the concluding state accrues only indirectly from its broader role.

In aiming to give an account of rational entitlement in a given case we want to explain why the transition from B to C is one the thinker is rational in making. We can take it that a transition is one the thinker is rational in making if it tends (in normal circumstances) to lead to true beliefs and if its tendency to do so is explained by constitutive or relatively a priori features of the *nature and content* of the states or acts of the subject involved. In aiming to give a bottom-up, or mixed, explanation of the subject's entitlement we will want an explanation of the entitlement to hold C to draw on essential features of the nature or content of B.²¹

Given the options on offer it is clearly sensible to aim to explore whether there is available an account of our entitlement to judgements about the force of our attitudes that is bottom-up. It does not close off the possibility of also developing an account of top-down warrant, either as additional warrant, or if we are unable to give any bottom-up warrant. Further, a significant advantage to giving a bottom-up or mixed, rather than merely a top-down, explanation, of why a conclusion is one to which the thinker is rationally entitled is that it meets the traditional demand of theories of warrant by justification, that warranted transitions need to be both truth conducive and explicable as rational *from the subjects point of view*. Whilst denying that we need construe 'being rational from the subject's point of view' as the demand that the subject have beliefs or evidence that justify her conclusion, the bottom-up entitlement

theorist does aim to supply an account of the subject's entitlement to their conclusion that is rooted in essential features of the state or act that is the grounds of the transition.

If this is right we need to explain why the subject is entitled to the belief 'I am judging that *P*' given that she is judging that *P*, in terms of features of the nature or content of the act of judging. As we stressed at the beginning, there are, three components of the judgement 'I am ϕ -ing that *P*'. Once one has accepted the demand for some bottom-up account of warrant by entitlement one is pressed by three questions correlative to those three components:

(a) What in the content and/or nature of B explains why the transition from it to *the content* of C is one to which the thinker is entitled?

(b) What in the content and/or nature of B explains why *the first person element* of C is one to which the thinker is entitled?

(c) What in the content and/or nature of B explains why *the attitude element* of C is one to which the thinker is entitled?

With a clearer idea of what might count as success when giving an account of our knowledge of our attitudes I want to turn to what I have called agency theories of self-knowledge. According to the agency theory of self-knowledge it is the exercise of agency that secures, or is able to secure for a subject, knowledge that she did that thing. I want to argue that we can explain how agency gives us knowledge by

showing that we can explain how attention to features of agency enable us to provide answers to these three questions.²²

4. Agency and Authority

With the task set of aiming to answer these questions we can now return to ask how agency is supposed to secure knowledge. What *epistemic* advantage does agency give us that automatic self-intimation, or conceptual mastery, does not? The essence of the agency theory is that the rational connection between the pre-suppositions of rational agency, and the self-ascriptions we are concerned with, is of a kind that automatically makes those self-ascriptions ones to which the subject is entitled in the absence of evidence or reasons.

The puzzle we face in giving an account of why the transition from the act of judging '*P*' to the act of judging 'I judge that *P*' is one to which the subject is rationally entitled, is not that it is hard to see how the transition is reliable. Given that, on such a transition, the means by which the self-ascription is reached is that which is self-ascribed, self-ascriptions made on that basis will track the truth. It is rather that it is hard to see why it is rational: from the subject's perspective the transition seems to cross a gap between radically different and unrelated contents. The subject judges, say, that the sky is blue. How is it in any way rational for her to judge 'I judge the sky is blue' on that basis. How is it that the transition, from a content about the sky, to a content which involves the first person and her attitudes, is one that is rational for the subject? The agency theory aims to step in at this point to show that the supposed gap between the content '*P*' and the content 'I judge that *P*' can be closed for the subject. The suggestion is that the gap is closed for the subject when we think of the

occurrence of the judgement with the content '*P*' as a product of rational agency, that is, in the context of a process of the subject determining what attitude she will adopt by a consideration of what is true. The idea is that such a process already pre-supposes first person and attitude elements in way that makes the subject entitled to self-ascribe any attitude she forms as the result of such a process. What are the relevant pre-suppositions? First, it is a pre-supposition of rational agency that the subject has possibilities available to her such that *she* is able *immediately* to exercise them. Second, it is a pre-supposition of rational agency that, in this case, the subject is led by a consideration of what is true in exercising one of these possibilities rather than another. Third, we can assume that a subject has a practical knowledge of these possibilities in order to be able to realise one, rather than another, on the basis of a consideration of what is true.

Elements of these three points can be seen as expressed in the following from Moran:

One must see one's deliberation as the expression and development of one's belief and will, not an activity one pursues in the hope that it will have some influence on one's eventual belief and will. ...Were it generally the case that the conclusion of his deliberation about what to think about something left it still open for him what he does in fact think about it, it would be quite unclear what he takes himself to be doing in deliberating. It would be unclear what reason was left to call it deliberating if its conclusion did not count as making up his mind.²³

Moran talks about deliberation, which may call for more than I have in mind in talking about being guided by a consideration of what is true. However, setting that aside, we can see the quotation as suggesting the view that a subject who self-ascribes an attitude that she formed guided by her consideration of what is true is entitled to take the attitude to be *hers*, because hers are the only attitudes that can be taken immediately on the basis of such a consideration. Further, a subject who self-ascribes an attitude, guided by her consideration of what is true, is entitled to take the attitude as being an attitude of ϕ -ing, because ϕ -ing is the attitude, which she has practical knowledge of as a possibility and which her consideration of what is true immediately led her to adopt.

Applying these thoughts to the problem with which we are concerned we can say that one knows that one is judging, rather than denying, questioning or doubting *P*, because the following are true:

1. Being a rational agent means determining ones attitudes on the basis of reason.
2. Determining one's assertoric attitudes on the basis of reason means determining ones attitudes by a consideration of what is true.
3. Determining one's attitudes by a consideration of what is true pre-supposes that judging, denying, questioning or doubting are options one can *immediately* implement in a given instance on the basis of such consideration.
4. It is only one's *own* attitudes that one can immediately form or change on the basis of such consideration.
5. The *force* of the attitude, with respect to *P*, that one immediately forms or changes on the basis on a consideration of what is true is determined by one's conclusion with

respect to the truth of *P* i.e. by whether one's conclusion is *P*, not-*P*, *P* is unsettled or unlikely.

6. A subject exercising reason over her thought must have a practical knowledge of her options to judge, deny or doubt as (a) things can be done and (b) as things that can be done by her.²⁴

4.1 Forming a judgement

Let us try to get a clearer view of the position by considering what is going on in a case where a subject forms a judgement. What is it about the subject's judging that makes her not only reliable, if she also judges that she is judging, but also entitled to the second order judgement? What makes the transition from judging that *P* to judging 'I judge that *P*' a rational one for the subject? We can suppose that forming a judgement subject to reason, that is subject to a consideration of the truth of *P*, is a task that makes sense relative to the capacity of a subject to ask herself the question whether *P*?

Rational agents can make judgements simply on the basis of perceptions and memories, and can make snap judgements without any explicit prior interrogative phase. However, it seems clear that in order for their attitudes to count as subject to reason, it is permissible to think of them as attitudes formed as an answer to the question 'whether *P*?'. A subject who was not able to ask or enquire, but only to judge or believe could not be thought to be a subject who was guided by her capacity to reason.²⁵ And, note, that a subject counts as judging guided by reason, in the sense being invoked here, just in virtue of forming her attitude on the basis of a consideration of whether *P* is true. It does not mean that her reasons for finding *P* true

are good ones or are sufficient to establish the truth of *P*, nor does it mean that the reasons she brings to bear are all the reasons she could bring to bear, nor need her reasons be very sophisticated: they may amount to no more than that she accepts what she perceives to be true. All it means is that judging that *P*, rather than denying *P*, doubting *P*, or continuing to question *P*, is what her considering whether *P* actually lead her to do in answering the question whether *P*.

So, how does this relate to the subject's entitlement to self-ascribe her attitudes? Suppose *X* asks herself whether *P*. *X* considers whether *P* is true, and judges *P* as a direct result. Her judgement that *P* is the result of her considering whether *P* is true. The exercise of considering whether *P* is true only makes sense against the pre-supposition that one can take a view with respect to *P*, that is, it pre-supposes that one is engaged in the activity of forming ones own attitudes. The activity of attitude formation pre-supposes a practical knowledge of the attitudes one can adopt: judging, denying, doubting and questioning that *P*. Concluding that 'P is true', on considering whether *P* is true, is therefore equivalent to the subject realising the practically known possibility of judging that *P*. In judging 'I judge that *P*', on the basis of her conclusion that 'P is true', the subject obviously judges something true. However, she also judges something she is entitled to take as true because her self-ascription is based on her conclusion that 'P is true' and her conclusion that 'P is true' is the result of a consideration of whether *P*, which *is* a consideration of what *attitude* to take with respect to *P*. If we can assume, as we have, that the subject has knowledge of whether the content of her attitude is *P*, or not-*P*, or *P* is unsettled or unlikely, and we can assume that the subject's judgement is an immediate response to her consideration as to the truth of *P*, with the practical knowledge that that is a consideration as to what

attitude to hold with respect to *P*, then in knowing that the content of her attitude is *P*, the subject is immediately entitled to hold that she is judging that *P*.

So, to go back to our earlier three questions, the suggestion is that we can answer the question as to 'what makes *X* warranted in judging 'I judge that *P*?' as follows:

(a) *X* is warranted in taking *P* to be the content of her attitude because she can re-deploy an assumed ability to think *P*. (Re-deployment)

(b) *X* is warranted in taking *herself* to be judging that *P* because the possibilities of judging *P*, denying *P*, questioning *P* or doubting *P* against which her asking herself whether *P* makes sense, are possibilities for *her* to judge, deny, question and doubt *P*. It is only *her* attitudes that are the attitudes that stand to be altered by a consideration of the truth of *P*.

A subject with the requisite concepts might be able to articulate an a priori connection between the project of aiming to the answer the question whether *P* by the evaluation of the truth of *P*, and the power to bring about an attitude change in *herself* consonant with the deliverance of such an evaluation. However, the entitlement does not rest on the subject having the requisite concepts or articulating the connection.

(c) *X* is warranted in taking *herself* to be *judging* that *P* because *X* has concluded that *P* is true which is, in the context, equivalent to her realising the practically known possibility of judging that *P* on the basis of a consideration of whether *P*. I have suggested that, for a rational agent, determining whether *P* on the basis of an

assessment of *P*, is equivalent to determining whether to exercise the practically known possibilities of judging, denying, questioning or doubting *P*. In finding *P*, to be true the subject realises the possibility of judging *P*, in finding *P* not to be true the subject realises the possibility of denying *P*, in finding *P* to be unsettled the subject realises the possibility of continuing to question whether *P*, and in concluding that *P* is possible, but not likely, the subject is realising the possibility of doubting that *P* etc. This means that the following are transitions to which a subject judging, denying, questioning or doubting as an answer to 'whether *P*?' is entitled.²⁶

(a) X judging that *P* (finding *P* to be true) as an answer to 'whether *P*?' is entitled to the judgement 'I judge that *P*'.

(b) X denying that *P* (finding *P* not to be true) as an answer to 'whether *P*?' is entitled to the judgement 'I deny that *P*'.

(c) X questioning or doubting whether *P* (finding *P* to be unsettled or unlikely) as an answer to 'whether *P*?' is entitled to the judgement 'I question or doubt that *P*'.

A subject with the requisite concepts would be able to articulate an a priori connection between the project of aiming to answer the question 'whether *P*', and the formation of an attitude change consonant with the result of a consideration of that question.

However, the entitlement does not rest on the subject having the requisite concepts, articulating the connection or applying (a), (b) and (c) as rules to her thought. A subject who lacks the requisite concepts can be taken to be knowledgeable with respect to the content and nature of their attitudes in virtue of the fact that the subject

must be assumed to have a practical knowledge of *what* attitude her reasons can lead her to adopt and the equivalence between answering whether *P* and the formation of one of those attitudes.

4.2 Revision.

A rational subject who reviews her cognitive commitments, say in the light of a newly made judgement, is led by reason immediately to revise only those attitudes that are hers, and only those that are in conflict with the new judgement. This means revising those attitudes consonant with both their content and their force. Different revisions are required depending on the force and contents of existing attitudes. The identification of those attitudes that stand to be revised, and the nature of the revision will be knowable to the subject in a way that correlates with the guidance of the subject's reason. If, on finding that *P* conflicts with *Q*, the subject judges *P* and denies *Q*, she is entitled to judge that she judges that *P* and now denies that *Q* because her judging that *P* was the possibility she was led by her reasons to realise.

A subject with the requisite concepts would be able to articulate an a priori connection between the project of aiming to revise her attitudes in the light of her reason and her prior knowledge of the content and the force of her attitudes. However, a subject who lacks the requisite concepts can be taken to be knowledgeable with respect to the content and nature of their attitudes in virtue of the fact that the subject must be assumed to have a practical knowledge of *what* attitudes she can form and revise given that she does form and revise them guided by her reasons.

Although the account offered here aims to account for our knowledge of our assertoric acts, rather than the more commonly dealt with belief, it clearly utilises elements of extant accounts. Like Evans' the account holds that one is in a position to know what one judges with respect to *P* simply by answering the question whether *P* but tries to offer more in the way of an explanation of why that is.²⁷ It is also much influenced by Moran's account of self-knowledge. In particular, it shares the key assumption that self-knowledge can be seen in central cases as product of rational agency. However, the present account has the particular aim of showing how such agency can provide a warrant by entitlement for self-ascriptions, and insists that a subject can have such a warrant without any conceptual mastery of the pre-suppositions of agency on which their entitlement rests. The account also has elements in common with constitutive accounts such as Heal's (Heal (2002)). We have such self-knowledge, according to Heal, in virtue of the fact that answering the question 'what do I believe?' *constitutes* the fact of which we were enquiring. But the account given here is non-constitutive. It is no part of the story that subjects are authoritative over their attitudes in virtue of the self-ascriptions of their attitudes constituting the attitudes they ascribe. It is not our decisions under the title of 'what do I believe' that lay down our judgements. Rather, it is as common sense tells us, forming a view with respect to *P* that determines our judgements. However, forming a view with respect to *P* presupposes conditions of agency that are such as to entitle the subject to the judgement that they are judging, denying, questioning or doubting *P* when they are. If they ask the question what they have judged, having judged that *P*, the conditions of such judgement mean that they are able to gain a knowledgeable answer. The account of self-knowledge, and the picture of entitlement adopted, most obviously reflected here are those offered by Peacocke. However, while Peacocke

explains how knowledgeable self-ascriptions are possible, in the absence of evidence or inference, by appealing to the idea that a conscious state can stand as the basis for its own ascription, the account offered here tries to explain such a possibility in terms of the conditions supposed by rational agency.²⁸ While Peacocke's account might seem to have the advantage of applying to the self-ascription of all of our mental states (or rather of all those which can be conscious) it is never very clear what account is being given of consciousness and so not clear how the normative connections required for warrant are explained. To the extent that the account offered here has succeeded in showing how the distinct elements of the self-ascription are warranted by distinct elements pre-supposed by rational agency the account offered here takes a step forward.

5. Objections

5.1 Voluntarism

It might be objected that in appealing to agency the account adopts an overly voluntarist conception of judgement. But it is important to stress that the account does not need to see judgements as something that the agent can merely will to do. The suggestion is not that the agent knows that they are judging because the agent can make true, by an effort of will, that fact that they are accurately self-ascribing their judgement. One way in which one's being an agent *can* play a role in explaining one's knowledge of something is that as an agent one can, in some cases, choose to make true the thing that one claims knowledge of. I can claim to know that my daughter will be picked up from school at 3.30pm because it is within my voluntary control to make it happen. I can know that she will be picked up at 3.30pm by going to pick her up at 3.30pm. I think, however, that it is entirely implausible to suppose that I know

that I am judging that *P*, because it is within my power to choose to make it true that I am judging that *P* (although it may be in my power to make it true that I come to view with respect to *P*). My judgements are under my rational control, but not under the kind of voluntary control the model supposes. And the suggestion as to why we are knowledgeable with respect to our judgements does not rest on this idea. Agency plays a role in explaining our knowledge of our judgements not because I can decide to judge that *P*, and so make true my belief about what I judge. Rather, agency is important because in order for it to be the case that I can be guided by a consideration of the truth of *P* to judge that *P*, it has to be that I judge that *P* against some capacity to know the attitudes I might be led immediately to form.²⁹

5.2 Judgement without evaluation

It might be objected that the account is able to explain our knowledgeable self-ascription of assertoric attitudes only in cases where they are made on the basis of a consideration of whether *P*. What, it will be said, are we to make of those cases where we judge, doubt etc. immediately on what seems to be no basis at all?³⁰ How do such cases affect the story about why we are entitled to claim knowledge of such attitudes? If the complaint is that there may be creatures who believe or judge in some attenuated sense in which their judgements are not guided by, or subject to, reasons, but who simply form representations capable of being veridical and which guide behaviour, I have made it clear that I am happy to acknowledge that the account may not deliver self-knowledge in those cases. However, I think, we would not want to deliver self-knowledge in those cases. So, let us assume that the complaint is that we, and perhaps less sophisticated reasoners, often form judgements without reasons seeming to come in to play.

It might be supposed that the suggestion offered, in terms of a bottom-up warrant by entitlement, has the advantage over a top-down story in making the subject's view of things play a role in explaining the warrant for her self-ascription, but only at the cost of limiting such cases to those in which the subject is actively reasoning. But I think that this cannot be quite right. It is a commonplace that even when a subject's knowledge rests on justification, say when a conclusion rests on her believing the premises and principles required for its demonstration, she does not actually have to run through the demonstration, or the premises and principles in order to be credited with knowledge of the conclusion. Indeed, she does not even have to be conscious of having any reasons for her conclusion. A subject knows just if her conclusion is one she could (holding her abilities fixed) have reached had she explicitly brought to mind what were in fact her reasons. It seems to me that we can say that the case where a subject asks whether *P*, and judges accordingly can provide us with a paradigm case in which the normative connections are clearly manifest. However, we can add that the normative connections and with it the entitlement, carry over to those cases in which a subject seemingly forms a judgement on the basis of nothing. If it is genuinely a *judgement* that the subject forms then it is a commitment to the truth of *P* that is subject to a consideration of whether *P*. Therefore, it must be such that had the subject asked herself whether *P*, while being in the same rational state but having access to the reasons she in fact has for truth of *P*, she would have answered as she did. For this reason we can take the entitlement to hold in cases in which the subject judges seemingly on the basis of nothing.³¹

5.3 *Belief*

It would be more than unwise to attempt to produce an account of how we know such dispositional attitude states as belief, as well as attempting to shed light on our knowledge of such acts of mind as our judgements. A proper account would need to be given of the relation between the two. For example, there seem to be conditions on genuine belief that do not hold in the same way for judgement. We doubt that someone has a belief, even if they declare that they have it, if it is not something that they go on to act on the basis of. While we might think it essential that judgements normally function to lay down beliefs that will direct future action, we are perhaps happier to allow cases of fly by night judgement that bring in no long term dispositional commitments.³² However, that said, an adequate account of our knowledge of our beliefs may well rest on adequate account of our knowledge of our judgements conjoined with an account of the causal, rational and conceptual connections between belief and judgement. The immediacy, authority and spontaneity known to characterize our knowledge of our beliefs will very likely come from the immediacy, authority and spontaneity that characterises our knowledge of our judgements and which seems to follow from them being products of rational agency.

References

- Bell, D. (1987), 'The Art of Judgement' *Mind* 96:221-224.
- Bilgrami, A. (1988) 'Self-Knowledge and Resentment' in MacDonald *et al.* (1998).
- Burge, T. (2003), 'Perceptual Entitlement', in *Philosophy and Phenomenological Research*.
- (1998), 'Reason and the First Person' in MacDonald *et al.* (1998).
- (1996), 'Our Entitlement to Self-knowledge', *Proceedings of the Aristotelian Society* 96, 91-116.
- (1993) 'Content Preservation', *Philosophical Review* 103: 457-488.
- Dretske, F. (2000), 'Entitlement: Epistemic Rights Without Epistemic Duties?' *Philosophy and Phenomenological Research* 60: 591-606.
- Frege, G. (1980), *Translations from the Philosophical Writings of Gottlob Frege* (ed) Geach and Black 3rd edition.
- Heal, J. (2002) 'On First Person Authority' *Proceedings of the Aristotelian Society* 102, 1-19.
- Hume, D. (1978), *A Treatise of Human Nature*, Edited by L.A Selby-Bigge and P.H.Nidditch, Oxford: Clarendon Press.
- James, W. (1976) 'The Experience of Activity' in his *Essays in Radical Empiricism* Cambridge, Massachusetts: Harvard University Press.
- Gallois, A. (2000), 'First Person Accessibility and Consciousness', *Philosophical Topics* 28.2: 101-146.
- Lawlor, K. (2001), *New Thoughts about Old Things: Cognitive Policies and the Ground of Singular Concepts*. New York: Garland.
- MacDonald, C., Smith B. & Wright C. (eds.), (1998), *Knowing Our Own Minds: Essays of Self-Knowledge*. Oxford: OUP.
- Moran, R. (2001), *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Pryor, J. (2000), 'The Sceptic and the Dogmatist', *Noûs* 34:4, 517-549
- O'Brien, L. (2003), 'Moran on Agency and Self-Knowledge', *European Journal of Philosophy* 11.3: 375-390.
- Peacocke, C. (1999), *Being Known*, Oxford: OUP.

(1998), 'Conscious Attitudes, Attention and Self-knowledge' in MacDonald *et al.* (1998).

(2002), 'Three Principles of Rationalism, *European Journal of Philosophy*, 10: 375-397.

Roessler, J. (1999), 'Perception, Introspection and Attention', *European Journal of Philosophy* 7.1:47-64.

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

Sosa, E. (1991), 'Intellectual Virtue in Perspective' in *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.

Velleman, J. D. (2000). 'On the aim of belief' in *The Possibility of Practical Reason*, 244–281. Oxford: Oxford University Press.

Williams, M. (2000), 'Entitlement: Epistemic Rights Without Epistemic Duties?' *Philosophy and Phenomenological Research* 60:607-612.

Wright, C. (1998), 'Self-Knowledge: The Wittgensteinian Legacy', in MacDonald *et al.* (1998).

¹ This paper has been presented at research seminars in Bristol, Nottingham, Stirling, St Andrews and London. My thanks to audiences for very helpful comments. I have also benefited from written comments from José Luis Bermúdez, and David Bain and discussion of the paper with Jen Hornsby, Sarah Patterson and Sarah Richmond. I also owe much to Tyler Burge, Richard Moran and Mark Sacks for related discussion.

² Moran (2001) is probably the most sustained treatment of the connections and it is clearly the greatest influence on the account offered here. But see also Bilgrami (1998), Gallois (2000) and Roessler (1999).

⁴ But see Lawlor (2001) for a somewhat different problem about knowing the contents of our attitudes.

⁵ The suggestion is primarily Burge's. It is developed by others, most notably in Peacocke (1999, Chapter 5)

⁶ I will also not consider seeing, hearing, smelling, and so on, that *P*, nor cases of visually, aurally or tactilely imagining *P*. Such cases are likely to be subject to an importantly different account of self-knowledge to that appropriate to purely cognitive states or acts.

⁷ Although these four basic kinds of assertoric act will be the focus in what follows it is important to acknowledge that there is significant oversimplification in their identification. There may be categories that are not represented – degrees of acceptance perhaps – and the categories that are acknowledged may not pick out only one kind of mental act.

⁸ I take much of this classification from Bell (1987). I will treat judgement and denial as distinct attitudes: if there are contents with indeterminate truth-values then denying that *P* may not be equivalent to judging that Not-*P*.

⁹ See Frege (1980, p.2).

¹⁰ On this point see Velleman (2000).

¹¹ James (1976, p.82)

¹² James(1976), p.84

¹³ James (1976), p.84.

¹⁴ For Hume "An opinion or belief is nothing but an idea, that is different from a fiction, not in the nature, or the order of its parts, but in the manner of its being conceiv'd. But when I would explain this manner, I scarce find any word that fully answers the case, but am obliged to have recourse to everyone's feeling, in order to give him a perfect notion of this operation of the mind. An idea assented to feels different from a fictitious idea, that the fancy alone present to us: and this different feeling I endeavour to explain by calling it superior force, or vivacity, or solidity or firmness or steadiness" Hume (1978, p. 629)

¹⁵ See Shoemaker (1996), Burge (1996), Moran (2001), Heal (2002).

¹⁶ George is, of course, named after Shoemaker's George (1996, p. 35). However, I am not sure that they are the same George. Shoemaker's George lacks any capacity for internal self-observation, it is less clear whether he is also stipulated to gain knowledge of himself only through external observation and the application of such rules to his behaviour.

¹⁷ George, for example, may not be credited with the concept as he is not in general prepared to self-ascribe his judgements, he is only prepared to when he expresses them.

¹⁸ This position has been carved out by Burge in number of papers: Burge (1993) with respect to memory and testimony, Burge (1996) with respect to self-knowledge, and 'Perceptual Entitlement', forthcoming, with respect to perception. However, it was already in play in Sosa (1991). It is also developed in Dretske (2000) and Peacocke (1999) and (2002).

¹⁹ Note that using 'warrant' and 'justification' in this way is merely a terminological decision. There would be no substantial disagreement with someone who wanted to insist that one could have justification for a belief that *P* that was not justification by reason or evidence. See Williams (2000).

²⁰ Bottom-up entitlement is most familiar from discussions of perceptual entitlement: where it is claimed that a transition from a perception with non-conceptual content to a perceptual belief with related conceptual content is warranted in virtue of the nature of perception and perceptual content. See Burge (2003), Peacocke (2002) and Pryor (2000)

²¹ I take the essentials of this gloss on rational entitlement from Peacocke (2002). It is also in play in Peacocke (1999, pp.235-245.)

²² I am assuming that we can explain knowledge in the conclusion if we can explain our entitlement to such transitions.

²³ Moran (2001, p. 94)

²⁴ Note that I do take such practical knowledge to require the subject to have the concepts and beliefs that would be required to articulate it. Rather, it is plausible to think that the content of our practical knowledge will be non-conceptual in much the way that our perceptual knowledge can be.

²⁵ Note that we may want to allow that simpler animals can be ascribed beliefs in the form of non-analogue representational states which control and regulate behaviour without any capacity to wonder whether *P*.

²⁶ I have already mentioned the extent to which matters have been oversimplified by talking about only the four central cases of judging, denying, questioning and doubting. The effects of such oversimplification become clearer here. First, because it is not clear that finding *p* unsettled is always to continue to question *P*. It might rather be a decision to take up a neutral attitude to *P* or perhaps to prescind from taking any attitude to *P*. Second, we might want to acknowledge acts of mind corresponding to different conclusions with respect to the likelihood of *P*. I am supposing that the basic story will be able to accommodate such possibilities with the appropriate fine-graining.

²⁷ However, it is unclear exactly what aspect of agency Moran takes self-knowledge to depend on. See O'Brien (2003).

²⁸ On a view where a cognitive attitude's being conscious is a matter of its being subject to rational control, this account comes out as a version of Peacocke's.

²⁹ It is not my aim to offer an account of what I called non-committal and suppositional acts of thought in this piece. However, it is worth noting that we are not similarly inclined to resist the idea that acts of supposition are voluntary to a significant degree. Unlike assertoric acts of thought that involve an attitude to the actual truth of P , acts of supposition involve a subject treating P as if true for some purpose. This suggests that in such a case it may be that I know that I assume that P because it is up to me to assume P and there is no question of my not being able to do what I will. In contrast, acts of merely thinking that P seem to invoke agency to a lesser degree than assertoric thought. To the extent that merely thinking involves mere contents passing through one's mind, with no attitude to their truth or falsity, one might wonder whether it is the job of a theory of the knowledge of the *force* of our attitudes to explain such knowledge. It is tempting to say that whatever explanation we give of how we answer the question of how we know the content component of our self-ascriptions can do duty here. Suppose the right account of how we know the content of our attitudes will establish that if I φ that P I can know that I am thinking that P , whatever φ . Then knowledge that I am merely thinking that P will be something I can have by thinking that P and being entitled to the conclusion that I am not taking an attitude to the truth of P .

³⁰ There is an issue here of what to think about judgement formed on the basis of perceptions or memories. Are we to think that perceptions and memories, themselves, can be reasons? I will not take a view on this, but do hold that in so far as subject a subject forms her judgements with the option of taking, or not taking, her perceptions or memories, at face value, we can count them as knowable by the subject in the way that judgements guided by reasons are knowable.

³¹ I am assuming that we can bracket any effects that actually questioning whether P or that explicit rationalizing may produce.

³² Note that it might also be argued that whether or not a mental action counts as a genuine judgement that P , rather than a whim, say, will also rest on whether the subject actually forms a belief that P . In which case if a subject takes herself to have judged P , but has not – in virtue of failing to form any belief P – then she obviously cannot know that she judged P . It is false. It is worth noting that this is compatible with denying that she needs to know that she formed a belief in order to know that she judges when she does and compatible with holding that she knows her belief in virtue of knowing her judgement. If belief is indeed an essential consequence of genuine judgement this is just what we would expect.
