

Knowing Beliefs, Seeking Causes
Krista Lawlor
(American Imago vol65 no.3 2008)

Knowing what one believes sometimes takes effort—it sometimes involves seeking to know one’s beliefs *as causes*. And when one gains self-knowledge of one’s belief this way—that is, through *causal self-interpretation*—one engages in a characteristically human kind of psychological liberation. By investigating the nature of causal self-interpretation, I discover some surprising features of this liberty. And in doing so, I counter a trend in recent philosophical theories, of discounting the value of self-knowledge in projects of human liberation.

Wo Es war, soll Ich werden
(Where id was, there ego shall be)
New Introductory Lectures

I would rather say, "Where superego was there ego shall be."
Hanna Segal

1. Know thyself

Famous commandments notwithstanding, one can wonder about the value of self-knowledge. Spinoza, for example, holds that self-knowledge is the key to a moral life and personal liberation. Certainly, something seems right about the idea; but Spinoza’s doctrine rests on the claim that knowing one’s passions is itself sufficient to free one from them: “An affect which is a passion ceases to be a passion as soon as we form a clear and distinct idea of it.” And this claim is certainly dubious.¹ As modern psychoanalysts know, we should be less than sanguine about the power of self-knowledge all by itself to liberate one from unwanted desires. One might come to know about one’s infantile desire but that no means suffices to make the desire vanish.

Recent trends in philosophy share in pessimism about the value of self-knowledge for human liberation. Some philosophers cast doubt on the power of self-knowledge to effect important personal ends, and focus instead

on the power of *practical* elements of the self in creating valued conditions like freedom of the will and autonomy. On Harry Frankfurt's view, for instance, freedom and autonomy arise with structural alignments of desires and volitions—a picture that seems to leave epistemic reflection without much work to do.² This voluntarization of various projects of self-realization comes full circle in Richard Moran's work, where *self-knowledge* is itself the object of voluntaristic re-construal.³ Self-knowledge, on Moran's view, is itself best understood as a *practical*, and not an epistemic achievement, made when one takes a stand on what one desires or believes. On such a view, purely epistemic—in contrast to purely practical—self-knowledge is not of primary interest.⁴

My aim here is to provide some counterbalance to these recent trends. We do have an interest in self-knowledge—first and simply because our beliefs, desires, impulses, fears and wishes are of fundamental concern to us, and we have an interest in knowing about what is of most fundamental concern to us; second, because distinctive kinds of human suffering result from finding oneself with prejudices one would disown, or acting upon beliefs or desires one has no knowledge of;⁵ and finally because, as I will argue here, self-knowledge can produce psychological change of a liberating kind. Just how, and just what kind of liberation are my concern in this paper.

Briefly, to foreshadow: I will be concerned to identify a kind of psychological liberation effected by *knowing one's beliefs*. In some cases, I will argue, knowing what one believes involves a distinctive self-critical posture, a posture one adopts when one seeks to know one's beliefs *as causes*. And adopting and sustaining this self-critical posture produces important kinds of cognitive psychological freedom. Knowing what one believes, when one's knowledge is gained through self-criticism, has special value.

I should also note that in what follows, I will be restricting attention to what we might call *quotidian self-knowledge* of our beliefs. The contrast I intend is with, on the one hand, what we might call *deep self-knowledge*, of the kind one seeks in therapy, the kind investigated and produced in psychoanalysis, concerning unconscious mental states; and on the other hand what we might call *basic self-knowledge*, of the kind philosophers lately have been concerned to give account, whose chief feature is its effortless nature—for example, knowing that one now believes that one is awake.⁶ Quotidian self-knowledge lies on a spectrum between deep and basic self-knowledge.⁷ On the one hand, quotidian self-knowledge is generally more

easily gained than deep self-knowledge, but not so easily gained as basic self-knowledge—that is, it is not a routine by-product of consciousness. So, for example, it doesn't take years of therapy to know that you believe, say, that a friend is making a mistake, or that a colleague is aggressive; but such a belief, should you have it, might not just announce itself to you, either. I should stress that, as different as quotidian self-knowledge is from deep self-knowledge, I think it is continuous with deep self-knowledge. Indeed, a central upshot of my paper is that our means of gaining quotidian self-knowledge—a means I call *causal self-interpretation*—is continuous with the kind of analytical interpretation that produces deep self-knowledge.

The plan for the paper is this: in the next section, I seek to describe a route by which one can come to know what one believes. It is the route of *causal self-interpretation*. This is hardly the only way one comes to know about one's beliefs, but it is a routine, and as we shall see, crucial means by which one knows what one believes (that is, by which one gains *doxastic self-knowledge*). In section 3 I explore cases where causal self-interpretation is, moreover, the only route to doxastic self-knowledge available to us. In section 4 I very briefly situate the causal self-interpretation account with respect to some other accounts of doxastic self-knowledge; and finally, with the ground thus cleared, in section 5 I investigate the liberating value of doxastic self-knowledge, when it is gained through causal self-interpretation.

2. Causal self-interpretation

In an important range of cases, we know what we believe only with effort. We must in these cases seek clues of our belief—clues found in characteristic mental events that the beliefs *cause*. The fact is we sometimes undertake a kind of *causal self-interpretation*, in coming to know what we believe. Or so I claim. In highlighting the importance of causal self-interpretation, I oppose much current philosophical thinking about self-knowledge of belief. Current philosophical thinking tends to see self-knowledge of belief as a kind of basic self-knowledge—readily accessible knowledge one has in virtue of being a conceptually competent and rational agent. My opposition to other philosophers is not of interest, of course—what is of interest is this: knowing what one believes has value; and important aspects of this value are revealed when we take account of the role of causal self-interpretation in gaining this knowledge.

But what is this causal self-interpretation that I say is central to self-knowledge? A good place to begin is with ordinary experience, with those

events that unfold on our path to self-knowledge. When we attend to ordinary experience, as many have noted, we find a rich field of discrete elements. Many figures, scenes, and imaged phrases pass through one's mind in any given stretch of wakefulness. And often, almost without noticing as much, one interprets these *imagings* (sometimes incorrectly) as being caused by specific mental states or attitudes.

This sort of self-interpretation is, I believe, a common route to all kinds of quotidian self-knowledge. Here are some examples: in the midst of one's daily activity, one pictures lemons, and where they're to be found at the store, *because* (one thinks) one intends to buy some. One suddenly imagines the house and fields where one grew up, *because* (one thinks) the summer sun just now reminds one of the summer sun there. One hears the words "Point Reyes" in one's inner ear, *because* (one thinks) one wants to see that part of California again. Assigning a cause is part and parcel of identifying the mental attitude that underlies the imaging. That is to say, assigning a cause is a way of identifying one's imaging as the product of an intending, a remembering, or a desiring, to a specific effect. Assigning a cause is knowing one's mind.

I have argued elsewhere that such causal self-interpretation is a central means by which one knows one's desires.⁸ I want to argue here that this sort of causal self-interpretation also has a role in one's knowledge of one's beliefs. My claim that self-interpretation is causal rests heavily on the careful description of our everyday experience, so let's start by considering a case.

Case 1: Owen is talking with an associate, and growing ever more uncomfortable with the conversation. He feels impatient with his associate, is upset at how the conversation proceeds. Only later, reviewing the conversation, is Owen in a position to understand why he felt aggrieved—his associate was giving Owen's position too little consideration; he was dismissive. In coming to these judgments, Owen replays bits of the conversation, recalls his associate's posture and tone of voice. (He sees in his mind's eye various postures, hears in his mind's ear the tone of voice, and words that were spoken.) All these particulars are what Owen recalls in seeking to make sense of his reaction. When Owen later self-attributes emotions and beliefs about the conversation, he seeks to make best sense of why he feels the way he does, and of those particulars of the circumstances that he continues to recall.

Cases like this are routine, but complex. What can we say about Owen's self-knowledge here? There are several points to note before answering. First, when Owen says "how irritating!" or "he was dismissive", Owen *expresses* his irritation and his belief; when Owen goes on to say, "I am irritated, because I believe he was dismissive", Owen doesn't merely express his emotion and his belief. Owen here makes a *self-attribution* of a particular emotion and belief.⁹ And his self-attribution of a belief has both an *expressive* and a *reporting* role. That is, Owen both gives voice to his belief, and simultaneously reports on its presence. (This is important to the value of self-knowledge; I'll have more to say about the dual role of belief attributions below.)

Another point to note: With his self-attribution, Owen also expresses his self-knowledge. We can distinguish aspects of quotidian self-knowledge here. In making the self-attribution, "I am irritated, because I believe he was dismissive", Owen knows (i) that he has certain emotions and beliefs about his associate's demeanor. In addition to knowing his emotions and beliefs, Owen expresses self-knowledge about the relation between these: he knows (ii) that his emotional reaction is driven by his belief about how he was being treated.

How does Owen come to make these self-attributions? This is of vital interest to us, since we seek first an account of how self-knowledge of belief is possible, and second an account of its liberating power. So I want to focus on this question: How specifically does Owen come to self-ascribe the belief that his conversation partner was dismissive? On what grounds does he say "I believe he was dismissive"?¹⁰

Some have argued that one correctly self-attributes belief solely on the basis of one's reasons or evidence for thinking the belief true.¹¹ In Owen's case, the claim is, he has *evidence that the content of his belief is true*; that is to say, Owen can point to facts about his conversation partner's behavior as evidence *that he was dismissive*, and on the basis of this evidence Owen self-attributes the *belief* that he was dismissive. *Rational self-interpretation*, on this view, is the route to self-knowledge of belief.

I think the matter is more complicated. I think our case reveals the fact that there is a *mixed basis* for self-knowledge of belief. Start with this fact: Owen knows that he feels irritation and that he believes he was dismissed, and he knows these both together—that is, he makes an holistic self-interpretation of both belief and emotion. Ascribing both is required to make best sense of the circumstances. Were he to self-ascribe the belief that

his associate was dismissive, and a feeling of elation, that would make little sense. Sense-making is a package deal. Moreover, Owen supposes, whether or not he has reflected much on the fact, that a *belief* about how one is treated can cause *emotions* like anger, disappointment or irritation. Owen deploys this much psychological understanding in self-attributing the belief.¹²

What Owen's case reveals is that there is a mixed basis for self-knowledge of belief. As I noted, the rational self-interpretation theorist argues that one self-attributes belief solely on the basis of one's evidence for thinking the belief true. I would agree that such evidence is indeed part of the basis of self-attribution of belief. That is, part of the basis of Owen's self-attribution of a belief (e.g. the belief *that he was dismissive*) is Owen's evidence that the content of his belief is true. Owen can point to facts about his conversation partner's behavior as evidence *that he was dismissive*. However, as our case makes clear, Owen bases his self-attribution on still more evidence. Another part of the basis of Owen's self-attribution of belief comes with evidence of related mental states of his. He feels irritation; he repeatedly replays certain postures and tones of voice. Both what one imagines over and over again, and one's emotional reactions, can serve as evidence for one's self-attribution of belief. One understands that a belief to a such and such an effect precisely would cause imagings and emotions of such and such kind. In our case, Owen self-ascribes the belief that his partner was dismissive partly out of this understanding; he understands that *this belief* would cause just *these reactions* in him. In assigning the belief as a cause, he identifies his various imagings as the product of this particular belief of his. So it's not wholly on the basis of evidence for the truth of a belief that one self-ascribes a belief. In Owen's case, it is not wholly on the basis of his conversation partner's behavior that Owen makes the self-attribution "I believe he was dismissive." Owen bases his self-attribution of belief partly on such evidence, but partly also on the evidence that his own emotional reactions provide, and the thought that these reactions have a cause (namely this belief).¹³

What we learn from such cases is that there are two evidential bases for a self-ascription of belief. The first is evidence the world provides for the truth of the belief—a reason for thinking the belief true; the second is evidence one's own mental states provide. We can call these two evidential bases for a self-ascription of belief the *reason-giving* and the *causal* bases of self-knowledge.

My case for claiming that self-interpretation is causal (or causal as well as rational) rests entirely on the description of our experience—when we attend to what goes on when we self-ascribe beliefs (as well as other mental states), we discover this about the nature of our interpretive efforts. Reflecting on ordinary cases, I don't think anyone can doubt that causal self-interpretation has a role in self-knowledge. But we might wonder *how much* of a role causal self-interpretation plays, in *contrast* to rational self-interpretation. Perhaps when we make self-ascriptions, it is rational self-interpretation, made on the basis of reasons for one's beliefs, that is really of primary importance? This is a good question, and to begin to frame an answer, I want to explore how and when causal self-interpretation takes on primary importance. Toward this end, I will next investigate cases where one makes a self-ascription of belief based solely on causal evidence, and in direct conflict with one's reason-giving basis. This happens chiefly when one discovers in oneself a belief that one finds irrational in some way—for instance, a belief that is outright biased or prejudicial, or a belief one unreflectively inherits from an authority. What is important about such cases is the way that causal self-ascription provides a crucial route to self-discovery and self-criticism.

3. The role of causal self-interpretation in self-criticism

I'd like to first clarify some terminology. Let's return briefly to our case. In order for Owen's self-ascription to (partly) have a causal basis, he must suppose, whether or not he has reflected much on the fact, that a belief about how one is treated can cause emotions like anger, disappointment or irritation. Thus Owen employs what some call *a third-person perspective* on his own mental states: he self-attributes a belief in part on the basis of what makes causal sense of why he has the self-attributed emotion. Of course Owen does not occupy this vantage point exclusively: Owen also finds evidence in his associate's demeanor for the truth of what he believes about his demeanor. This is just to say, Owen also employs what some would call *a first-person perspective* on his own mental states: he self-attributes in part on the basis of his reasons for thinking his belief true.

We can see, however, that speaking in terms of first- and third-person "vantage points" or "perspectives" is misleading, as it suggests that one can occupy only one at a time, and so to deploy only one set of resources for self-interpretation at a time. In fact, as I think our case makes clear, self-knowledge involves both third-person and first-person perspectives at the

same time. It involves making a self-attribution that makes sense from both perspectives.¹⁴ So I prefer to say that self-knowledge involves both seeing *causal relations* among one's states and seeing *reasons* for the attitudes themselves. One seeks the best self-interpretation one can find in light of both kinds of consideration.

Next, we can note that the two bases for self-interpretation might come apart. Imagine Owen feeling irritation and anger at how the conversation went, and wanting to say, "I believe he was dismissive", but not being able to find any reason-giving basis for self-ascribing the belief. Replaying the conversation, recalling his associate's tone and look, he cannot produce a reason for thinking the belief *true*. If Owen is an apt self-interpreter, if he's any good at knowing himself, this tension in the bases of self-ascription will make him hesitate. (Indeed, he may chose to fall back on a different self-attribution, rather than "I believe he was dismissive", he'll say "I *felt* he was, but I can't say why.")¹⁵

Now we can pose a question concerning how great a role causal self-interpretation has. The question goes something like this: "So the case you have just outlined seems to show that the reason-giving basis for the self-attribution of belief is *required*, doesn't it? Without support from reasons, one cannot so much as make the self-attribution. When it comes to knowing one's beliefs, then, causal self-interpretation alone seems insufficient; a self-interpreter will always need a reason-giving basis for the self-ascription of belief, won't she?"

Answering this question begins with noting that, although the reason-giving basis may routinely be part of a completely satisfactory ascription of belief, it is not required. To see that this is so, imagine another case where the causal and reason-giving bases of self-ascription come apart, but in this case, the causal basis provides overwhelming evidence in favor of an ascription of a belief that the reason-giving basis speaks against. A person who holds a prejudiced belief is a good example:

Case 2: Cora has every reason to think that blue-eyed people are just as intelligent as brown-eyed people. On the basis of these reasons alone, she would self-ascribe belief *that blue-eyed among us are just as intelligent*. And yet, were she able to take account of the causal bases for a self-attribution, such as her negative emotional reactions to blue-eyed people, and her dispositions to treat them as inferior, Cora would have to self-attribute the belief *that blue-eyed among us are inferior*. (The tension this situation

produces would be great in an able self-knower. For others, the tension would be lessened by engaging in certain modes of self-deception.)

What Cora's case illustrates is that the reason-giving basis for self-attribution is not required for self-knowledge. Were Cora to make her self-attribution of belief in part or wholly on her reason-giving basis (that is, on the basis of what she has *reason* for believing), her self-attribution would be incorrect. (She could be stopped from making an attribution altogether, of course, but that is itself another failure of self-knowledge.) Since she doesn't believe what she has reason to believe, she cannot get self-knowledge on reason-giving grounds.¹⁶

In fact, one might make the stronger claim: in cases where one diagnoses irrational belief in oneself, the causal basis of one's self-knowledge is crucial. In Cora's case if she is to know her doxastic state, she must make a self-ascription of belief based on causal evidence, where this is in direct conflict with her reason-giving basis. In such cases, causal self-ascription provides one's only route to self-discovery and self-criticism.

Still, someone might press the line of questioning about the relative importance of finding reason-giving bases for self-knowledge, saying "Well, nonetheless, it is still *more important* in some sense to hear from Cora what she takes it she has reason to believe, and what she rationally commits herself to—and causal self-interpretation will be silent on these matters."

In response, I think we should agree that knowing what Cora takes herself to have reason to believe *is* very important. After all, this will be part of her basis for self-correction, as she tries to rid herself of her irrational belief. In a clear sense, moreover, the belief that she sees reason to have is more reflective of *who she wants to be*. And that too is a very important thing to know, and certainly a part of self-knowledge. And we should also agree that causal self-interpretation doesn't tell one about these matters. What I disagree with is that any of these facts should *lessen* the importance (indeed the critical role) of causal self-interpretation. If Cora harbors a prejudicial belief that she sees no reason for, and if she can only self-attribute on reason-giving grounds a belief in the equality of intelligence of the blue-eyed among us, then she fails to have significant self-knowledge. And she likely fails to be able even to begin the process of becoming who she wants to be.

The point deserves emphasis: in many cases one does not need to consult causal evidence, and can make perfectly good self-ascriptions of belief on reason-giving grounds alone; however in cases where reason-giving

and causal bases support conflicting self-ascriptions (as in Cora's case), one is epistemically at risk of failing to know the full facts about one's beliefs if one confines one's attention to reason-giving bases alone. In specific cases, moreover, the very possibility of self-criticism rests on the availability of causal evidence of one's belief. So, causal self-interpretation is not only a routine means for knowing what one believes, but in cases of irrationality, it is vitally important, and plays a primary role.

To recap, I have tried to suggest that one routinely makes self-attributions of belief on a mixture of causal and reason-giving bases. In addition to considering what one has reason to believe, one also often has recourse to kinds of interpretation usually thought of as third-personal; specifically, one often resorts to inference to the best explanation, seeking to identify the causes of the various and sundry elements in one's stream of consciousness, thereby to know one's mind.¹⁷ Moreover, I have tried to sketch how, in some cases—especially of irrational belief—the very possibility of rational self-criticism rests on the availability of a causal self-interpretation.

I'll have more to say about the value of rational self-criticism soon. But before going on to describe the liberating effects of self-knowledge gained through causal self-interpretation, I want very briefly to situate my account here with respect to two leading alternative accounts of self-knowledge of belief. My main concern is to show that neither of these accounts handles irrational belief. In light of what we've seen here, that's a significant shortcoming in each of the views I will consider.

4. Two alternative theories of self-knowledge of belief

There are several prominent theories of self-knowledge of belief.¹⁸ In this brief section, I do not aim to carry out a full critical assessment of any of the alternative accounts. My aim is only to orient my view with respect to two leading alternatives, so as to highlight the virtue of causal self-interpretation, as it accounts for how we might know about our irrational beliefs.

The Transparency Account. Suppose that I am asked or prompted in some way with a question about my current state of mind: 'Do I believe that p ?' In order to answer, I ask myself *whether p is true*, and consider the reasons for thinking the content p true; if I come thereby to judge that p , I will also thereby come to believe it, and in the same stroke know that I believe it.¹⁹

The so called *transparency* of one's beliefs (and some would have it, other attitudes as well) is a matter of the question "Do I believe that p?" getting answered by addressing instead the question "is p true?" Beliefs are transparent in the sense that their subject matter is the sole concern of the rational agent. On the Transparency account of doxastic self-knowledge, it is this fact about beliefs that accounts for their unique epistemic availability. We have already seen something like the Transparency account at work above (section 2). One knows what one believes by attending to what one has reason to believe.

I'll set aside critical discussion of the substance of the Transparency account: the only point I want to note here is that, on this account, not all one's beliefs are transparent to one, not all are discoverable through consideration of reasons for thinking p true. Restrictions apply: first, theorists interested in transparency confine their account to what some call 'occurrent' beliefs—beliefs formed upon being prompted. Second, the account is restricted to occurrent beliefs formed rationally, upon making the relevant judgment concerning what one has reason to believe. So the Transparency account is restricted in its scope. Beliefs formed irrationally, through biasing mechanisms, or under various kinds of motivational pressure, cannot be known this way.²⁰

The Rational Supervenience Account

Sydney Shoemaker famously argues that I know what I believe insofar as I am rational: self-knowledge supervenes on rationality.²¹ More specifically, if one is rational, then having various first-order beliefs at the same time constitutes having the second-order beliefs that give one self-knowledge. Here is how Shoemaker traces the connection between rationality and self-knowledge: Imagine a rational creature, faced with new evidence from a fresh experience. If the evidence contradicts a current belief, then since the creature is rational, we can expect that a readjustment of attitudes will occur. But first-order beliefs and desires alone will not rationalize such a readjustment; some second-order beliefs are also required. Since being rational requires responding to new evidence with readjustments in one's attitudes, and such readjustments are rational only if made on the basis of beliefs about one's own attitudes, then being rational requires also having these second-order attitudes. And these second-order attitudes just are instances of self-knowledge.

Again, I leave substantive discussion of the account aside, and simply note here that, like the Transparency account, the Rational Supervenience account is restricted in its scope. As Shoemaker says, in speaking about cases of self-deception:

What I have asserted... is a connection between self-knowledge and rationality; that given certain conceptual capacities, rationality necessarily goes with self-knowledge. It is entirely compatible with this that there are failures of rationality that manifest themselves in failures of self-knowledge. And such I assume we have in cases of unconscious belief.²²

So if one is *irrational*, and one's second-order beliefs about one's first-order beliefs are formed through biasing mechanisms, or under various kinds of motivational pressure, then the Rational Supervenience account does not apply to one.²³

As I have been emphasizing, each of these alternative accounts of doxastic self-knowledge explicitly restrict their domain of explanation. On each of these accounts, attention focuses on cases in which the subject is in some sense or other a rational, well-functioning believer. (What it takes to be rational and well-functioning may vary with each account, but that needn't concern us here). As we might say, each account gives a story about self-knowledge subject to *provisos*: the account only applies provided you're rational, or provided your beliefs are formed directly from your rational judgments about what is to be believed, with no intervening pressure from your motivational side. About those cases where believers are self-deceived, motivated to disguise their own beliefs from themselves, or form beliefs on a less than reasonable basis, these accounts are silent.

Now of course it is in just these situations that self-knowledge is important. In situations of self-deception, or motivated or biased belief, self-knowledge can often be the first step to self-correction. So we precisely want an account of how the ordinary person, when less than rational, might come to know what she believes. (Returning to being rational in fact may require some self-knowledge, and it wouldn't be much consolation to be told that we'll have self-knowledge once we return to being rational.) So we need an account of how one knows what one believes when one is less than rational, less than fully responsible for one's attitudes, and when the beliefs themselves are not immediately accessible for being immediately responsive to reasons. And here I think it's clear that causal self-interpretation is what does the work.²⁴

In the next section, I want finally to investigate the liberating effects of doxastic self-knowledge gained through causal self-interpretation.

5. Liberating effects of doxastic self-knowledge gained through causal-self-interpretation

There are no doubt many ways in which self-knowledge can help one effect various goals of self-realization. For instance, one effect has been often noted in the literature on self-knowledge: If one knows something, then one has a justified belief about it;²⁵ and a justified belief about one's own attitudes produces normative pressure on the attitudes that are the object of one's knowledge.²⁶ In this way, justified belief about one's own attitudes and feelings can produce normative pressures to change irrational attitudes. Change can come in a variety of ways: acquisition of new attitudes, or fuller realization of the functional role (both epistemic and practical) of existing attitudes. Having a justified belief about one's own attitudes does not of course *constitute* these changes, as other psychic pressures can interfere. But in the well-functioning person, justified belief about one's attitudes has as part of its functional role to produce normative pressures on the attitudes that are its object. These sorts of normative pressures have been noted by philosophers.

What interests me is how self-knowledge of *belief*, specifically when it is had through *causal self-interpretation*, can produce psychological change of a liberating kind. In brief, the idea that I want to try to flesh out is this: Knowing what one believes, sometimes anyway, involves a distinctive self-critical posture, a posture that one adopts when one seeks the causes of one's beliefs. And adopting and sustaining this self-critical posture produces important kinds of cognitive psychological freedom.

But what kind or kinds of psychological freedom? And what is the nature of this self-critical posture—how is it gained and how is it sustained long enough to permit one to become a better epistemic self? A little reflection reveals interesting ways in which causal self-interpretation permits one both to adopt and sustain a self-critical posture, and thereby achieve interesting kinds of cognitive liberty.

Let's begin with the question of how one comes to adopt the relevant self-critical posture in the first place. This can be quite a feat. The thing to notice is that if one has a baseless or irrational belief, then becoming one's best epistemic self requires changing that belief; but changing a belief can be difficult, *especially* when it is baseless or irrational. The reasons for this are

complex, though one central source of difficulty certainly is this: In a case where one already has a prejudicial belief, merely considering what one has reason to believe is often not sufficient to produce the belief one ought to have. (If change were this easy, the world would be much more free of prejudice than it is.) Rather, considering what is reasonable to believe in such cases can instead result merely in a person *believing that she believes the reasonable thing*, deceiving herself about what she in fact believes. This creates a puzzle about how one can even begin to know about baseless beliefs, how one can even begin to adopt the required self-critical posture. For in such a case, one has to *see oneself as unreasonable* first, in order to rid oneself of unreasonable beliefs. And there are evidently psychological barriers to so seeing oneself. So how can one do this? The answer, I think, is clear enough: In such cases, knowing about the belief one *should* have is not helpful, for knowing this is of a piece with seeing oneself as reasonable. Seeing oneself as *unreasonable* is only possible if one knows about the belief one *does* have—i.e. the baseless belief. And, since rational self-interpretation won't deliver this knowledge, one can *only* know about the belief one does have—the baseless belief—through causal self-interpretation. Causal self-interpretation, then, is the only means one has in such cases to *see oneself as unreasonable*. Seeing oneself as unreasonable is the difficult entry point, the moment in which one adopts or takes up a self-critical posture regarding one's own beliefs. So it is causal self-interpretation that permits one to adopt the self-critical posture that is essential to the process of ridding oneself of irrational or baseless belief.²⁷

Next, let's consider how one maintains this self-critical posture, and the role of causal self-interpretation in its maintenance. One way to characterize the maintenance of a self-critical posture regarding one's beliefs is to say that one *stands apart* from one's beliefs. But *standing apart* is of course a metaphor—more concretely what it amounts to is that one *reflects* upon one's beliefs as elements in one's psychology, as opposed to merely acting upon, or more broadly *living with* one's beliefs as commitments. How can such reflection be psychologically effected? I think a key element is in the manner in which one self-attributes one's belief, specifically, that one formulates one's self-attribution by explicitly *reporting* on the attitude in question. In our first case above, Owen moves from self-attributions that are mere *expressions* of the relevant attitudes ("how irritating! He was being dismissive") to self-attributions that are both expressions and *reports* of the relevant emotions and beliefs ("I felt irritated with him. I believe he was

being dismissive"). What is the value of being able to explicitly formulate one's attitudes in a report? Clearly, one great value is that it permits reflection on the justifiedness of the attitudes. Owen's self-attribution "I believe he was dismissive" naturally invites the question of whether this belief is justified, creating room for self-criticism. If one were restricted merely to *expressing* one's beliefs, one would not be in a position cognitively to turn on oneself, so to say, and question the belief.

Another way to put the point is this: In the moment when one attributes a belief to oneself on causal grounds, one is not wholly occupied with reasons to think the content of the belief true.²⁸ Reasons for thinking the belief true are only part of what one keeps in mind in making the self-attribution. One's self-attribution also must make sense of how one feels, how one is disposed to act, what one is occupied with in the way of mental imagings. Since one is not wholly occupied with reasons for thinking the belief true, one can also entertain the possibility that one holds a belief *against reason*. Again, were one confined to reason-giving bases alone for gaining self-knowledge, one would enjoy no such room for questioning.

So causal self-interpretation creates the space to maintain the relevant self-critical posture. When one gains doxastic self-knowledge through causal self-interpretation, one stands apart from one's belief. In knowing on causal grounds that one has a belief, one has room to question whether one's belief is reasonable. That is a kind of "room for maneuver" that, in one's cognitive life, marks an important liberty.

Finally, I offer a more speculative point. I want to return to the issue of how one comes to adopt the relevant self-critical posture in the first place. As we saw, this is quite a feat; I think there's more to say about the kind of cognitive liberty it produces. Start with a piece of common wisdom: Being able to make and take self-criticism requires *maturity*. But what exactly does this maturity consist in? In the case of self-criticism of one's beliefs, more specifically, what does maturity require? We can start by returning again to the question of what is required to self-ascribe an irrational or baseless belief. First and foremost, it seems, one has to see the belief both as *one's own* and *without basis*. What this in turn requires, I suggest, is a capacity simultaneously to free oneself from both the perspective of the internal critic of one's beliefs and the perspective of the biased believer.

In many cases, it is extremely unpleasant for creatures like us to own up to baseless or irrational belief. As evidence we find familiar cases in which the initiation of self-criticism generates not *change*, but *self-deception*. This

belief I would self-ascribe on causal grounds is seemingly baseless, so I *find reasons* for the belief after the fact (engaging in a kind of *reason-creating rationalizing self-deception*). Alternatively, this belief I would self-ascribe on the basis of reasons is not one I find any other evidence of having; but, determined to see myself as reasonable, I act in ways that produce *causal evidence* of my believing what I do not (engaging in a kind of *evidence-creating rationalizing self-deception*).

We can diagnose these kinds of self-deception: the psychic pressure to have reason for one's beliefs is significant, and can overwhelm a person's earnest efforts to know what she actually believes. We might put our diagnosis more vividly this way: What is often at work in cases of self-deception is a need to see oneself as *wholly reasonable*, or as we might say, to identify entirely with the *internal critic* who delivers rational criticism. In so identifying, one seeks to see oneself as reasonable throughout, and thereby falls victims to one or another kind of rationalizing self-deception.²⁹

What is interesting are cases in which one resists this pressure. Clearly there are cases where, against the psychic pressure to see oneself as reasonable, sometimes, one makes the self-ascription of an irrational belief. And in these cases, instead of identifying wholly with the internal critic and thereby being led into various kinds of self-deception, one also owns up to the irrational or baseless belief. But—equally important—this is not to say that one identifies with the fallible believer, either. Were one to identify with the fallible believer, that would mean simply *persisting* in one's irrational belief. A truly self-critical posture involves both seeing the biased belief as *one's own* and *without basis*. And that requires freeing oneself from the perspective of the internal critic (who is of course wholly reasonable) *and* freeing oneself from the perspective of the biased believer (who is unreasonable). Self-criticism is possible only when one steps back from both perspectives, and adopts a perspective from which both the facts of what one believes and the facts about what one has reason to believe are visible. When one occupies this perspective, one can deliver the bad news to oneself about one's irrational belief, correct one's belief accordingly and get on with one's epistemic life.

If one can truly adopt a self-critical posture, then, one widens one's perspective on one's beliefs, and this marks a kind of maturity. Now, I suggest that this kind of maturity involves the attainment of a kind of psychological liberation, too. It is the liberation one has when one manages to *free oneself from the pressure always to see oneself as reasonable*—at

least long enough to succeed in making genuine self-criticism. Since causal self-interpretation is, again, a central means by which one assumes the relevant self-critical posture, it has a central role in producing this distinctive sort of freedom.

On this last point, I am tempted to add an analogy. I think we might understand the freedom at issue on an analogy with Hanna Segal's idea, namely, that our goal is the enlargement of the ego, or self, against the superego, or the voice of conscience. I find it suggestive to think that one achieves something like this kind of enlargement of the ego when one steps outside the perspectives of the internal critic of irrational beliefs and the fallible owner of them. The idea is this: when one frees oneself from the pressure to identify with the self-critical agency (the voice of epistemic conscience) one takes up the perspective of a person with an irrational belief (one's self); and at that point, whatever criticism of one's beliefs there is to be made comes from *none other than one's self, who is also the owner* of the belief. What one gains in such moments of liberation is a "realistic" perception of one's own epistemic state.³⁰

The analogy is, as I said, speculative. What is certain, I believe, is that causal self-interpretation produces various kinds of cognitive psychological freedom. In knowing on causal grounds that one has a belief, one has room to question whether one's belief is reasonable, and that kind of room for maneuver marks an important cognitive liberty. Causal self-interpretation is a central means by which one assumes a distinctive kind of self-critical posture, one that involves freeing oneself from the pressure always to see oneself as reasonable.

Conclusion

In many important cases, knowing what one believes takes effort—it involves seeking to know one's beliefs *as causes*. And when one gains self-knowledge of belief this way, it can produce change of a liberating kind. Not because discredited beliefs simply go away—clearly nothing so simple describes our epistemic lives. Rather, because coming to know what one believes through causal self-interpretation amounts to an exercise of psychological freedom.³¹

Department of Philosophy
Stanford University

Bibliography

Bilgrami, Akeel. 2005. *Self-Knowledge and Resentment*. Cambridge: Harvard University Press.

Burge, Tyler and Peacocke, Christopher. 1995. "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society*, XCVI: 91-158.

Curley, Edwin. 1994. *The Ethics and Other Works: A Spinoza Reader*. Princeton: Princeton University Press.

Dilman, Ilham. 1984. *Freud and the Mind*. New York: Basil Blackwell Publishing.

Frankfurt, Harry. 1971. Freedom of the Will and the Concept of the Person, *Journal of Philosophy*, 68 (1): 5-20.

Lawlor, Krista. Knowing What One Wants (forthcoming) in *Philosophy and Phenomenological Research*.

McGeer, Victoria. 1996. Is 'Self-Knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation, *Journal of Philosophy*, 93(10): 483-515.

Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

----- . 1994. Interpretation Theory and the First Person *The Philosophical Quarterly*, 44 (175) 154-173.

Segal, Hanna. 2007. The Mind as Conflict and Compromise Formation: Comments on Charles Brenner's Paper, in *Yesterday, Today and Tomorrow*. Ed. Nicola Abel-Hirsch. Routledge.

Shah, Nishi and Velleman, David. 2005. Doxastic Deliberation, *Philosophical Review*, 114 (4): 497-534.

Shoemaker, Sydney. 1996. *The First-Person Perspective and Other Essays*. New York: Cambridge University Press.

Velleman, David. 2001. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Wilson, Timothy. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge: The Belknap Press of Harvard University Press.

¹ See *The Ethics*, Pt.V Prop 3. Curley (1994, p247) translation. In this sense, one's self-mastery increases as one's self-knowledge does.

² For Frankfurt, a volition is a second-order desire that certain of one's first-order desires be effective. See Frankfurt (1971).

³ Moran (2001).

⁴ Epistemic self-knowledge is a matter of having a justified true belief about one's own attitudes, while practical self-knowledge is a matter of endorsing, by seeing reasons for, one's attitudes. For more on this distinction see also McGeer (1996). Some understand the attainment of *self-knowledge* just to be the attainment of greater authenticity or autonomy, and some read Freud as having just this in mind by self-knowledge. See for instance Dilman (1984). My interest is in self-knowledge understood as an epistemic matter, the attainment of self-knowledge is a matter of justified self-ascriptions of attitudes, emotions, wishes, impulses and so on.

⁵ See Velleman (2001). Though Velleman does not specifically speak of the distress caused by acting without self-knowledge, he does stress the role of a fundamental desire to know what one is doing, in the activity of autonomous agents. It is not hard to imagine that such a desire for self-knowledge, when thwarted, produces a distinctive kind of human suffering.

⁶ Not surprisingly, this *effortlessness* is itself difficult to explain, and requires significant theoretical efforts. See for instance Burge, T. and C. Peacocke (1995), Shoemaker (1996).

⁷ Let me also add that I think *basic* and *deep* are both important kinds of self-knowledge, deserving of investigation. But I also think the relative absence of accounts of quotidian self-knowledge needs remedy.

⁸ Knowing What One Wants (forthcoming).

⁹ Owen may just say “I’m irritated because he was dismissive”, dropping the explicit “I believe...” The point remains, though that he is now self-attributing a belief. I use the more explicit formulation to keep in view the fact that the case of interest involves self-attribution of belief.

¹⁰ Note that self-attribution and self-knowledge are distinct. Questions about what makes a self-attribution an authoritative action can be distinguished from what makes a belief about one’s mental states an epistemically authoritative cognition. I blur these distinctions for ease of exposition here, but in other contexts they are worth remembering.

¹¹ For recent discussion of this view, see Shah & Velleman (2005).

¹² Some would call it theory, but whether folk psychology counts as a theory is a good question in my opinion—though this is not the place to defend my opinion.

¹³ And also in what his imagination selectively replays for him as most pertinent about his partner’s behavior. But here we may say that it requires further reflection to ground self-knowledge: Owen may be sensitive enough to the fact that his replaying some features of his conversation over is a clue to how he feels or what he thinks about the conversation. In his mind’s eye he recalls the image of his associate interrupting. He may be reflective enough to ask himself, Why do I keep returning to that? When he answers, “because I believe he was dismissive”, his self-attribution has as part of its epistemic ground this further fact that these are the salient features his consciousness selects to serve up.

¹⁴ Moran’s (1994) paper is very interesting here.

¹⁵ Here Owen doesn’t make a self-attribution of belief, but of a feeling. I would want argue that his self-attribution of feeling also has a causal basis, but I won’t pursue the argument here. See Wilson (2002).

¹⁶ We can now see why with a self-attribution, one both gives voice to one’s belief, and simultaneously reports on its presence. Belief attributions have a dual role. In being made partly on the basis of reasons for thinking the belief true, a self-attribution *expresses* a commitment to the truth (and so “I believe that p” is another way of asserting, “p”); and in also

being partly made one a causal basis—that is evidence for thinking one has the belief—the self-attribution *reports* the fact that one has a belief.

¹⁷ The resulting view is that inference, which has long been held to be irrelevant for knowing one’s own mind, has a significant role in gaining self-knowledge. This holds true as much for beliefs as it does a range of attitudes, such as desires, intentions, fears and hopes. If I have made self-knowledge sounds complicated, well, it is. However, complicated or not, self-interpretive exercises that balance causal and reason-giving factors are ones we routinely engage in. What we routinely try to do is complicated.

¹⁸ Here I’ll confine my attention to those that make self-knowledge an epistemic matter, not a purely practical achievement.

¹⁹ See Moran (2001) and Shah and Velleman (2005).

²⁰ I think cases like the one considered in the previous section show that even in the restricted case of ‘occurrent’ belief, one’s self-attribution of belief may be epistemically grounded in more than reason-giving considerations alone. I make an argument to this effect in “Knowing What One Believes” (manuscript).

²¹ Strictly, being rational and having the relevant concepts of belief and so forth, are sufficient for one to know one’s beliefs.

²² Shoemaker (1996, p.49).

²³ Note that Akeel Bilgrami’s recent work (2005) shares these broad features with Shoemaker’s account. Like the Transparency and Rational Supervenience accounts, Bilgrami’s account is restricted in its scope to the rational subject.

²⁴ By way of further situating my account, let me add that we might expect that defenders of each of the alternatives canvassed above would not object to the claim that causal self-interpretation is the basis of self-knowledge of belief in those cases where the provisos fail—that is, in cases where beliefs are formed or maintained through ‘irrational’ mechanisms. And for my part I agree that it’s important to know what one has reason to think true. Rational constraints on belief are a routine part of what one keeps in mind when one knows what one believes.

²⁵ Naturally not all philosophers agree. Some suppose belief has nothing to do with knowledge, others that justification has nothing to do with knowledge. I won’t take on these views here, but see Brakel, this volume, for more on this matter.

²⁶ Shoemaker (1995).

²⁷ Since causal self-interpretation can be one's only means of ridding oneself of irrational or baseless belief, causal self-interpretation can be one's only means of becoming one's best self.

²⁸ Likewise for desires, fears and other attitudes. One is not in the grip of reasons for thinking the content of the desire desirable, the fear fearful and so on.

²⁹ I don't intend here to suppose that the internal critic is identical with the superego, the activity of which is far from wholly reasonable. By "internal critic" I mean aspects of one's reflective self, and especially one's reflective self as it is concerned with the rationality of one's beliefs.

³⁰ This admittedly stretches Segal's idea, for she had nothing like epistemic self-criticism in mind when she spoke of "perceiving realistic guilt": She writes "...if the ego is an organ of perception of psychic and external realities then it is capable of perceiving the damage done in reality, or in phantasy, and that perception is realistic guilt, a guilt which is proportional to intentions and appropriate to the damage done. Freud said 'Where id was there ego shall be.' I would rather say 'Where superego was there ego shall be.' (Segal, H. "Comments On Charles Brenner's Paper: The Mind as Conflict and Compromise Formation")

³¹ Thanks for helpful comments to Lanier Anderson, Linda Brakel, Peter Rudnytsky, and Tamar Schapiro.