Acta Anal (2016) 31:285–296
DOI 10.1007/s12136-015-0279-3

The Structure of Cognitive Agency

Daniel Breyer[1]

Received: 13 August 2015 / Accepted: 15 December 2015 / Published online: 23 December 2015

Abstract Credit theories of knowledge have to explain the conditions under which beliefs are attributable to cognitive agents. The most promising way to explain these conditions is to offer an account of cognitive agency that is a plausible development of the uncontroversial notion that we are believing subjects. This article develops and defends a Structuralist model of cognitive agency.

Keywords Cognitive agency · Epistemic agency · Cognitive integration · Credit theories of knowledge · Belief ownership

Virtue epistemologists like John Greco and Ernest Sosa have developed detailed credit theories of knowledge, according to which 'knowledge is belief whose success is 'creditable' to the believer' (Sosa 2011: 86). Believers cannot be properly credited with beliefs, however, unless their beliefs can be attributed to them as their own (Breyer 2013b: 2–6). As a result, credit theorists need to explain the conditions under which beliefs are attributable to subjects. The most promising way to explain these conditions is to offer an account of cognitive agency that is a plausible development of the uncontroversial notion that we are believing subjects, and this is exactly what John Greco and I tried to do when we introduced and briefly sketched two novel accounts of cognitive agency—the 'Taking Responsibility model' and the 'Structuralist model' (Breyer and Greco 2008). In subsequent work, I have developed and defended the Taking Responsibility model (Breyer 2010, 2013a, b). In this article, I want to develop and defend the Structuralist model. Although I argue that this is a powerful and plausible model of cognitive agency, my defense is limited, because difficult questions about the relationship between agency and practical identity suggest that it is an open question whether the Structuralist model provides a fully satisfying account of cognitive agency.

✉ Daniel Breyer
dbreyer@ilstu.edu

[1] Illinois State University, Normal, IL, USA

1 The Structuralist Model of Cognitive Agency

Greco and I only suggested the contours of the Structuralist model in our initial sketch (Breyer and Greco 2008: 181–183). So, in this section, I need to develop the Structuralist model more fully, before defending it against objections in Section 2. To start with, however, I want to offer a brief defense of the notion of cognitive agency itself, because some epistemologists are understandably wary of it.

Kieran Setiya has recently expressed concerns about what he calls 'epistemic agency,' the view, as he puts it, 'that we can be active, rather than passive, in relation to our beliefs' (Setiya 2013: 179). One reason epistemic or cognitive agency makes little sense, according to Setiya, is that it is simply false that beliefs are active states like actions. Rather, he points out that beliefs are static states, not dynamic events. As a result, any notion of cognitive agency that requires beliefs to be dynamic, active, and event-like is simply mistaken. I agree with Setiya on this (Breyer 2013b). Fortunately, we need not view cognitive and moral agencies as analogous in the sense that one produces beliefs while the other produces actions. We can explore a different analogy instead—an analogy that highlights the similarities between beliefs and desires, as well as the similarities between the mechanisms that ground belief formation and those that ground desire formation. In developing the Structuralist model, this is precisely the approach I take.

Of course, this approach might not satisfy everyone. After all, the nature of agency is a controversial issue. I acknowledge that and recognize that I cannot hope to resolve those controversies here. What I propose to do, instead, is develop a plausible account of cognitive agency that is analogous to a plausible account of moral agency. I think any plausible account of moral agency will account for the fact that we are beings who are involved in the production of desires and actions in such a way that these products become attributable to us as expressions of who we are. Likewise, I think any plausible account of cognitive agency will account for the fact that we are beings who are involved in the production of beliefs in such a way that these products become attributable to us as expressions of who we are. With this in mind, let us begin developing the Structuralist model of cognitive agency.

The Structuralist model is analogous to Laura Ekstrom's account of moral agency (1993, 2005a, 2005b). According to Ekstrom, agents act autonomously when their actions flow from critically evaluated desires. Ekstrom calls such desires 'preferences.' For Ekstrom, an agent's desire is her own, only if it counts as a preference and integrates with other preferences. By analogy, the Structuralist model of cognitive agency holds that a belief counts as an agent's own, in the sense that it is attributable to her, only to the extent that it is a product of her cognitive character. On this model, cognitive character is nothing more than a causally integrated system of cognitive mechanisms (or dispositions), all of which are aimed at truth (Breyer and Greco 2008: 183). The central idea behind the Structuralist model, then, is that, whenever we have an integrated cognitive character and a commitment to attaining the truth, we have a full-fledged cognitive agency.

The Structuralist model is closely analogous to Ekstrom's account in important ways (Breyer 2013b: 12). According to Ekstrom, a desire is one's own, only if it is (a) an embodiment of one's conception of the good, (b) formed by a process of critical evaluation, (c) part of a supportive system, (d) sensitive to defeating evaluations, and

(e) psychologically comfortable. Likewise, according to the Structuralist model, a belief is one's own, only if it results from a cognitive mechanism (or disposition) that is one's own, where a mechanism is one's own, only if it is (i) attuned to the truth, (ii) manifested when one is motivated to believe the truth, (iii) part of a supportive system, (iv) sensitive to defeating evidence, and (v) psychologically comfortable. The Structuralist model grounds cognitive agency in a commitment to the truth, in the same way that Ekstrom's model grounds moral agency in a commitment to the good. The Structuralist model requires that beliefs that flow from cognitive agency do so in a way that reflects one's commitment to the truth, just as Ekstrom's model requires that desires that flow from moral agency do so in a way that reflects one's commitment to the good. The Structuralist model also requires both cognitive integration and sensitivity to defeating evidence, in the same way that Ekstrom's account requires both preference integration and sensitivity to defeating evaluations.

As we have seen, both the Structuralist model and Ekstrom's model require psychological comfort. This notion needs some unpacking, however. What does it mean for a desire or a belief or a cognitive mechanism to be 'psychologically comfortable'? We might cash this out in different ways, but for the moment, I would like to contrast psychological comfort with alienation, because this will help us understand the Structuralist model more deeply. Consider Sally who is an unfortunate thrill-seeker (Breyer and Greco 2008: 179; Breyer 2013b: 13). While walking across the Brooklyn Bridge, Sally finds herself with the urge to jump into the foaming waters below. What should we say about her urge? We might say that Sally acts compulsively, rather than autonomously, and conclude that she is not in fact free with respect to her urge to jump. Following Ekstrom's model, we would have to offer a better explanation than this, however. This is because we cannot simply say that Sally acts compulsively; we have to explain why her action counts as compulsive. Perhaps Sally's urge fails to embody her commitment to the good, or perhaps it fails to integrates with her preferences, or perhaps Sally simply feels as though her urge is not her own. If either of these possibilities is the case, then Sally is alienated from her urge, because it does not properly express who she is—either because it fails to embody her conception of the good, because it clashes with what she otherwise cares about, or because it simply feels as though it is not hers. Of course, we might describe Sally differently. We might say that she actually owns her urge to jump and that her urge actually expresses who she is, but to do this, we have to maintain that Sally's urge embodies her conception of the good, integrates with her other preferences, and otherwise fits comfortably in her psychologically. If we cannot say this, then we have to say that Sally's urge is not her own and that, in fact, she is alienated from it in a way that impairs her agency.

With this in mind, consider Norman, a clairvoyant who knows the location of the President, but who has no idea why he has these kinds of beliefs about the President (BonJour 1985). According to the Structuralist model, Norman's case is analogous to Sally's. Norman could own his belief that the President is in New York, but only if his beliefs were to arise from a cognitive mechanism that aims at the truth and cooperatively interacts (and so integrates) with his other cognitive faculties, which are themselves attuned to the truth and jointly constitute Norman's cognitive character. What if Norman feels weird about these beliefs, even though they have the right causal history? By analogy with Ekstrom's model, the Structuralist model tells us that Norman's beliefs about the President would not count as products of his cognitive agency, because

Norman's discomfort reveals that the mechanism that produces such beliefs is relevantly alien. Rather than owning either his beliefs about the President or the mechanism that produces such beliefs, Norman is alienated from both, just as Sally would be alienated from her urge to jump, if that urge had failed to integrate with her other preferences, had failed to embody her conception of the good, or had otherwise failed to fit comfortably in her psychologically. Both Norman and Sally suffer from impaired agency whenever they experience alienation; just as Sally's moral agency is impaired, because her desire arises from a fragmented moral character, so too Norman's cognitive agency is impaired, because his belief arises from a fragmented cognitive character.

As I have already suggested, the core idea behind the Structuralist model is that full-fledged cognitive agency results when we combine an integrated cognitive character with a commitment to attaining the truth. In this sense, the Structuralist model of cognitive agency is minimal, but it is nonetheless robust enough to explain why attributing even perceptual beliefs to an agent is appropriate, since a belief needs only arise from an integrated set of cognitive mechanisms, each of which is truth-conducive. What is more, the model is psychologically plausible, since it does not require voluntary control over one's beliefs, meta-level awareness of the factors that justify one's beliefs, or any other implausibly high standards for belief ownership and cognitive agency (Breyer 2013b: 9–10). The Structuralist model is also closely analogous to a plausible account of moral agency, a fact that gives us a good reason to take the model seriously as an account of cognitive *agency*. As a result, the Structuralist model is both attractive and promising.

2 Defending the Structuralist Model

Despite its attractive simplicity and its promising power, the Structuralist model faces potential objections. In this section, I raise and address three objections: (i) the Epistemic Value Objection; (ii) the Uncharacteristic Belief Objection; and (iii) the Sub-Personal Mechanism Objection. I argue that none of these objections undermines the Structuralist model. In Section 3, however, I argue that another objection, which I call (iv) the Adequacy Objection, reveals the model's potential limitations.

Let us consider the Epistemic Value Objection first. As we have seen, the Structuralist model relies heavily on cognitive integration to account for cognitive agency. The Epistemic Value Objection picks up on this and questions the value of cognitive integration in particular. Even if integration can help provide an account of moral agency like Ekstrom's, so the objection goes, why think that *cognitive* integration is valuable? In other words, what distinctively epistemic work does cognitive integration do?

My initial response is that cognitive integration is valuable, just because it provides a plausible account of cognitive agency. Of course, this response has its weaknesses. After all, the Epistemic Value Objection ultimately presses the concern that the Structuralist model of cognitive agency is ad hoc, serving only to solve problems that credit theorists have created for themselves. So let me offer another response.

Neil Levy (2006) has suggested that fragmentation impedes moral agency by impairing one's will, in the sense that it hinders one's ability to do what one values most. I suggest that fragmentation impedes cognitive agency by impairing one's

intellect, in the sense that it hinders one's ability to believe the truth. Let me be clear: my claim is not that coherence is truth-conducive. That is implausible. Rather, my claim is that, under normal circumstances, cognitive fragmentation (i.e., the disharmonious interaction of sub-personal cognitive mechanisms) serves as an *obstacle* to believing the truth. If various cognitive mechanisms are working against each other, the likelihood that one will believe the truth is lowered. This means that cognitive integration contributes, in normal cases, to the objective justification of a belief, in the sense that integration makes it more likely that agents will believe the truth. Of course, in unusual cases (like those of Norman the clairvoyant or Mr. Truetemp or the case you are dreaming up right now), fragmentation does not obstruct one's ability to believe the truth, because in such cases, reliable cognitive mechanisms generate true beliefs *in isolation*. The problem with such isolated beliefs, however, is that they are not subjectively justified, because they fail to be subjects' own in the sense that would allow us to attribute their beliefs to them as cognitive agents. Such beliefs arise 'out of the blue' in the sense that they are not grounded in the cognitive character of a believing subject. The epistemic value of integration, therefore, is twofold. In the first place, cognitive integration increases the objective likelihood that an agent's beliefs will be true, in normal circumstance. The reason for this is that fragmentation is typically an obstacle to true belief, and integration removes this obstacle. And as a bonus, cognitive integration supports an account of cognitive agency and thus provides an explanation for the conditions under which a belief counts as one's own. As a result, cognitive integration is epistemically valuable in both an objective sense and a subjective sense.

The Structuralist model can avoid the Epistemic Value Objection, but objections remain. Let us take the Uncharacteristic Belief Objection next.[1] This objection turns on the commonsense intuition that someone can come to know something in a way that is completely out of character. To make the objection clear, consider Annie the atheist, to whom God reveals Himself. Everything Annie believes supports her disbelief in God. When God reveals Himself to her, however, Annie can surely come to know that God exists. Yet, so the objection goes, because Annie's belief is radically out of character for her, the Structuralist model tells us that Annie cannot own her belief and that she therefore cannot know that God exists. Annie's case is an apparent counterexample to the Structuralist model.

The Uncharacteristic Belief Objection rests on a misunderstanding. The Structuralist model does not hold that an agent's *beliefs* are integrated; it is not a coherence theory. The Structuralist model holds, instead, that cognitive mechanisms (including sub-personal mechanisms) integrate to form a cognitive agent. So it is perfectly fine for Annie to come to believe something that conflicts with her previous beliefs. In such a case, her practical identity might thereby change; she might become, as it were, a different sort of person. Changes in her beliefs do not, however, result in changes in her cognitive character. In other words, cognitive agency is not a function of belief and preference, although moral agency might be. With this in mind, we can say that changes in Annie's beliefs are benign. As a result, the Uncharacteristic Belief Objection fails.

Nonetheless, a worry might persist. As I told the story, Annie has never before experienced God, but we might stipulate that she experiences God for the first time

[1] I thank Allan Hazlett for first raising this objection in conversation.

through a special faculty—her *sensus divinitatis*, which has lain dormant her entire life, until just this moment. In this revised case, Annie seems a lot like Norman the clairvoyant, but (so the objection goes) we nonetheless think that Annie *can* know that God exists based on this unique experience. If this is right, then neither cognitive integration nor belief ownership seems necessary for knowledge, and the Structuralist model of cognitive agency is rendered otiose.

This revised objection certainly seems to have teeth, but it only seems that way because Annie's case is under-described. If we flesh out Annie's case, we can see why the objection ultimately fails. So, consider five possible descriptions of Annie's experience.

(A1) While walking home after work, Annie the atheist experiences the most overwhelming presence in her life and knows, at that moment, that God exists.

(A2) While walking home after work, Annie the atheist experiences the most overwhelming presence in her life and believes without a doubt that God has appeared to her. She has never been more certain of anything in her life.

(A3) While walking home after work, Annie the atheist experiences the most overwhelming presence in her life. She believes that she is experiencing God, she has no conflicting thoughts about her experience, and God has, in fact, appeared to her.

(A4) While walking home after work, Annie the atheist experiences the most overwhelming presence in her life. She believes that she is experiencing God, but she feels uncomfortable about the feeling, because she also reasons that she cannot be possibly experiencing God, even though, in fact, God has appeared to her (despite her doubts). After some reflection, Annie decides that she had an incredibly vivid hallucination.

(A5) While walking home after work, Annie the atheist experiences the most overwhelming presence in her life. She finds herself with the absurd and completely 'out of the blue' belief that she is experiencing God, despite the fact that she also believes that a recent concussion is the source of the belief. After a moment of terror, Annie decides to see a neurologist, despite the fact that God had, in fact, appeared to her.

The Revised Uncharacteristic Belief Objection presents Annie as described in (A1), but (A1) is under-described, while (A2) and (A3) flesh out the description in (A1). In (A2) and (A3), however, we have no reason to think that Annie's cognitive faculties are not integrated, for her belief that God has appeared to her is supported by her other belief-producing mechanisms (at least in the minimal sense that those mechanisms do not interfere with her *sensus divinitatis*) and Annie is psychologically comfortable with her new belief in God. By contrast, in (A4) and (A5), Annie clearly does *not* know that God has appeared to her. In (A4), Annie has defeaters that arise from other cognitive mechanisms, but she only has these defeaters because her cognitive faculties are integrated. In (A5), Annie's belief that she has experienced God comes completely 'out of the blue' and Annie has no idea what to make of it, because her *sensus divinitatis* is just like Norman's clairvoyant faculty.

Clearly, (A1)–(A5) do not exhaust the possibilities. We could describe Annie's case in countless ways. My point is simply that (A1) cannot count against the Structuralist model, because (A1) does not explain how it is that Annie's *sensus divinitatis* is not integrated with her other cognitive faculties but nonetheless produces knowledge (keeping in mind that knowledge requires both objective and subjective justification). The burden is on the critic to describe a case in which Annie's belief is obviously not her own (in the sense that it does not flow from the exercise of her cognitive agency), but nonetheless counts as knowledge. The problem, however, is that if the critic manages to produce a case that she finds compelling, our intuitions will inevitably clash. This is because her case will surely be yet another variation on the Norman case, a case that I would not accept as a counterexample to the Structuralist model (or, for that matter, to a belief ownership condition on knowledge). As a result, I maintain that even the Revised Uncharacteristic Belief Objection fails.

This brings us to the Sub-Personal Mechanism Objection.[2] As I have emphasized, the Structuralist model holds that cognitive mechanisms, including sub-personal mechanisms, integrate to form a cognitive agent. In describing cases of impaired cognitive agency, however, I have argued that agents like Norman the clairvoyant and Annie the atheist are alienated from the *beliefs* that their sub-personal mechanisms produce. This is odd, because it suggests that sub-personal mechanisms, in particular, are not really doing the work they need to do. What seem to matter are not sub-personal mechanisms, but person-level mechanisms (such as conscious perception and conscious reasoning) and the beliefs they produce.

In response to this objection, I want to emphasize that cognitive integration can fail in both sub-personal and person-level ways. To be clear, a sub-personal mechanism is one that operates below the level of conscious awareness, whereas a person-level mechanism is one that either operates at the level of conscious awareness or operates in such a way that, through introspection or reflection, one could become aware of it. Sub-personally, a specific cognitive mechanism can fail to cooperate with other belief-forming mechanisms, and an agent might be completely unaware of this fact. In such cases, I want to argue that we have impaired agency without either awareness or discomfort.

Consider Samuel, who is having trouble identifying his friends when they smile. The reason for this is that various sub-personal mechanisms that underpin Samuel's facial recognition system are not cooperatively interacting with various sub-personal mechanisms that underpin his emotional recognition system. When Samuel looks at his friend Sarah, for instance, he can identify her as long as her face is nearly expressionless, but if she is smiling widely, furrowing her brows, or otherwise expressing basic emotions, he has trouble identifying who Sarah is, even though he can tell that someone is experiencing a basic emotion. Samuel does not realize that this is happening and the mechanisms responsible operate below the level of conscious awareness, but what is key is that, according to the Structuralist model, his cognitive agency is nonetheless impaired, simply because these sub-personal belief-forming mechanisms lack system-wide integration. Of course, Samuel might become aware of what is happening, even though the sub-personal mechanisms involved would remain opaque, and he might start to feel uncomfortable about his facial recognition beliefs, in the same way Norman might start to feel uncomfortable about his beliefs about the location of the President. In

[2] I thank an anonymous referee for raising this objection.

this case, Samuel's discomfort alienates him from his beliefs and impairs his agency, even though he remains unaware of the sub-personal fragmentation responsible for his inability to recognize anything but expressionless faces.

As Samuel's case demonstrates, impaired cognitive agency can result from both *sub-personal disintegration* and *person-level alienation*. What this means is that the Structuralist model should make room for both sub-personal and person-level belief-forming mechanisms. With this in mind, I can clarify my response to the Sub-Personal Mechanism Objection. If the objection is claiming that the Structuralist model should not make room for sub-personal mechanisms at all, then it misses the mark, as we have just seen. But if we read the objection as a request to include person-level mechanisms (like conscious perception and conscious reasoning), then we should simply note that the Structuralist model already makes room for such mechanisms, just as it makes room for awareness and discomfort.

3 The Adequacy of the Structuralist Model

In my view, the real question the Structuralist model faces is whether it presents an adequate account of cognitive agency. Let us call this the Adequacy Objection. My view is that intuitions might conflict on this issue. For some, the Structuralist model will seem sufficiently robust, whereas for others, it will not. In this section, I explain why I think some might find the model lacking and ultimately suggest that, although it is an open question whether the Structuralist model provides a fully satisfying account of cognitive agency, that fact should not count too strongly against the model.

To see why someone might think that the Structuralist model is insufficiently robust, let us start by considering the following account of a cognitive character:

(CC) S's cognitive character is S's stable, reliable, and integrated cognitive pattern when S is motivated to believe the truth.

(CC) It seems to capture the account of cognitive character that we find in the Structuralist model. Now consider two accounts of moral character:

(MC) S's moral character is S's stable, reliable, and integrated pattern of thought and action when S makes moral choices.

(MCI) S's moral character is S's stable, reliable, and integrated pattern of thought and action when X makes moral choices that arise from S's preferences (i.e., S's critically evaluated desires).

We can now consider an important question. Which account is more closely analogous to (CC)—(MC) or (MCI)?

If (CC) was more closely analogous to (MC), then (CC) would seem to be *too thin* to ground credit attribution. This is because (MC) makes no mention of agent-relevant features such as 'concerns and commitments.' In this sense, (MC) seems to present an account of a stable mechanistic system, not an agent. This is a problem because such mechanistic systems are insufficiently robust to account for the fact that we are beings who are involved in the production of beliefs in such a way that these products become

attributable to us as expressions of who we are. In the following passage, Gary Watson raises this objection in somewhat different terms:

> When thought and behavior are exercises of what Dewey calls an agent's moral capacity, they and their results are open to distinctive kinds of evaluation. These evaluations are inescapably evaluations of the agents *because* the conduct in question expresses the agent's own evaluative commitments, her adoption of some ends among others. To adopt some ends among others is to declare what one stands for. (Watson 2004: 270; my emphasis)

Watson is addressing Susan Wolf's worry that 'real self' accounts of responsibility are superficial, insofar as they (i) offer merely causal explanations of behavior and (ii) simply describe qualities of a thing, with the result that 'to blame someone for an outcome is to trace this effect to some fault (imperfection) in the thing' (Watson 2004: 269; see also Wolf 1990: 40–41). Watson argues that evaluations of an agent's character involve more than simply the identification of a person's qualities, let alone a person's normal patterns of thought and action. As he puts it, '[t]he significant relation between behavior and the "real self" is not (just) causal but *executive* and *expressive*' (2004: 270; his emphasis).

Watson's point here is not that one must be responsible for one's character, in the sense that one must have somehow brought it about, in order to own it. Rather, his point is that agent appraisals would be superficial (because they would be merely aetiological) if they did not implicate 'commitments and concerns.' In committing 'oneself to a conception of value', Watson points out that one *takes responsibility* (271). This means that one's actions become expressive of one's 'identity as an agent' (270) and one's practical identity is, at least partly, constituted by 'one's purposes, ends, choices, concerns, cares, attachments, and commitments' (287). Such commitments and concerns make what one does *one's own*. They provide the ownership required to transform aetiological explanations into agent evaluations. In other words, we can evaluate others as agents, only if they have commitments and concerns; otherwise, they are simply mechanistic systems, however complex they may be.

Keeping this in mind, let us return to the Adequacy Objection. The objection, we can now see, is that if (CC) is closely analogous to (MC), it is hard to see how (CC) can provide an adequate account of cognitive character and belief ownership. This is because (CC) would not have the resources to explain how the agent herself is present in the formation of her beliefs, just as (MC) cannot explain how the agent herself is present in the production of her actions, even though those actions flow from what we have stipulated as her stable, reliable, and integrated character. As a result, if the analogy holds, it seems that (CC) cannot ground an adequate account of cognitive agency.

In response, we might argue that (CC) is actually analogous to (MCI), rather than (MC). Recall that (MCI) states that S's moral character is S's stable, reliable, and integrated pattern of thought and action when S makes moral choices that arise from S's preferences (i.e., S's critically evaluated desires). Although (MCI) does not explicitly mention concerns and commitments, it implicitly refers to commitment, insofar as it requires that moral choices arise from one's critically evaluated desires (or preferences). (MCI) is roughly the view of moral character Ekstrom has in mind. On this approach, desires that undergo a process of critical evaluation embody an agent's conception of the good, and in this sense, they reflect an agent's commitments by reflecting what an

agent most values. If (CC) is closely analogous to (MCI), then it would seem to offer the kind of *thick account* of character that credit attribution requires.

Whether (CC) is closely analogous to (MCI) depends on how we read (CC)'s final clause: *when S is motivated to believe the truth*. Does this clause capture the kind of cognitive commitment that could ground credit attribution? The phrase "being motivated to believe the truth" and synonymous phrases such as "trying to believe the truth" and "thinking conscientiously" do not mean "thinking with an explicitly voiced purpose of finding out the truth" or "thinking with this as one's sole purpose." Rather, as John Greco has suggested, they indicate 'the usual state that most people are in as a kind of default mode – the state of trying to form one's beliefs accurately' (2000: 190). As a result, "thinking conscientiously" does not mean "thinking with one's epistemic commitments in mind" or "thinking in a way that results from one's prior epistemic commitments", each of which would imply personal commitment; instead, "thinking conscientiously" simply means "thinking in the way most people do when they are motivated to believe the truth." What is nice about this way of unpacking what it means to be motivated to believe the truth is that it makes room for sub-personal belief-forming mechanisms in a way that the Structuralist model seems to require.

If this is right, however, we might worry that (CC) is not in fact analogous to (MCI), because (CC) does not ground cognitive character in the distinctive kind of personal commitment in which (MCI) grounds moral character. The concern here is that the process of critical evaluation that produces preferences seems *qualitatively different* from the motivation to believe the truth, whereas the former captures the distinctive values of a particular agent (by capturing an agent's conception of the good), the latter captures only the default position that all cognitive agents share. Another way to put this is to say that (CC) points only to the normal, stable, reliable, and integrated cognitive pattern that emerges when someone thinks in the way most people do when they are motivated to believe the truth; it does not point to the *agent herself*—it does not pick out *her* cognitive concerns and commitments and, therefore, it does not identify her as an agent. On (CC), then, it seems that credit attributions would effectively reduce to aetiological explanations.

We might make progress if we insist that (CC) should unpack cognitive character in terms of person-level belief-forming mechanisms, like conscious reasoning, which could plausibly capture individual commitments.[3] When someone consciously reasons through a difficult problem on an exam, for instance, she personally wants to believe the truth: she wants to get the answer right, and getting the answer right matters to her. We might even add that such person-level mechanisms are associated with certain epistemic virtues (such as open-mindedness and conscientiousness) and certain epistemic vices (such as closed-mindedness and negligence) and that these person-level traits provide exactly what we need to make (CC) closely analogous to (MCI), rather than to (MC).

This is an attractive suggestion, but it faces two serious problems. The first is that it seems to suggest that sub-personal belief-forming mechanisms per se are irrelevant to cognitive agency. As we have seen (most clearly in Samuel's case), however, information about sub-personal mechanisms matters to questions about cognitive agency and cognitive character. In particular, we have seen that sub-personal disintegration can impair agency without awareness, without discomfort, and without epistemic vice, because mechanistic malfunction is sufficient to undermine agency. The second

[3] I want to thank an anonymous referee for making this suggestion.

problem is that any plausible account of cognitive agency needs to acknowledge that the reliability of our cognition depends, at least in part, on sub-personal mechanisms.

We might sidestep these problems by explaining how sub-personal mechanisms are related to person-level mechanisms. In fact, this is exactly what I have done while defending the Taking Responsibility model of cognitive agency (Breyer and Greco 2008; Breyer 2010, 2013a, b). According to this model, we do not need fine-grained conscious awareness of the sources of our beliefs, but we do need to recognize ourselves as the source of our beliefs and accept that we are fairly credited with having certain beliefs. As a result, we take responsibility at the personal level for the sub-personal mechanisms that produce our beliefs. By requiring this, the Taking Responsibility model nicely explains the connection between sub-personal and person-level cognitive mechanisms. Notice, however, that if we were to add anything like this requirement to (CC) and the Structuralist model, we would add far too much and thereby construct an *alternative model* of cognitive agency. The trick in developing the Structuralist model is to remain as minimal as possible while emphasizing cognitive integration and cognitive character over even implicit awareness.[4]

What this suggests is that the Structuralist model and (CC) must make room for sub-personal belief-forming mechanisms in a way that does not subsume the sub-personal under the person level. As a result, we cannot read (CC) as capturing personal epistemic commitments by unpacking cognitive character solely in terms of person-level belief-forming mechanisms.

Where does this leave us? We might conclude that the Structuralist model is simply too thin to be adequate, but we need not draw that conclusion yet. This is because it is plausible to claim that (CC) is strictly analogous to neither (MC) nor (MCI). On this alternative reading, (CC) offers a distinctive *epistemic conception* of agency, one that marks a middle way between the extremes of (MC) and (MCI). As we have seen, (MC) seems to provide too little, while (MCI) seems to provide too much. On the one hand, (MC) is too impersonal, providing only a bare-bones description of a mechanistic system, rather than a full-blooded account of agency. On the other hand, (MCI) is too personal, because it emphasizes the importance of individual commitments and concerns to full-blooded agency. By contrast, (CC) emphasizes a general value commitment, the commitment to believing the truth. In this sense, one could argue that (CC) describes more than a mechanistic system, but articulates a conception of agency that avoids entangling itself in questions about practical identity—questions about who we take ourselves to be and what we value. In light of this, it seems as though the adequacy objection has teeth only if we require that accounts of cognitive and epistemic agency address questions about practical identity.

This brings us to an important question. To what extent is agency of any kind entangled with practical identity? For those who follow Harry Frankfurt (1988), for instance, the answer to this question will surely be that agency is intimately entangled with practical identity, but it is not clear, at least not to me, that we have to answer the question in this way. After all, the Structuralist model appears to provide a solid foundation for epistemic credit, at least in so far as we do not press questions about

[4] For further discussion of related issues, see Breyer 2013a, b: 6–11 and 20, where I suggest that the most plausible account of cognitive agency would combine elements of both the Taking Responsibility model and the Structuralist model.

who we take ourselves to be and what we value too hard. And it is not clear to me how hard we *should* press such questions. In light of this, it seems that whether the Structuralist model provides a plausible account of cognitive agency is an open question. This should not count too strongly against the model, however, because we cannot possibly settle this open question without also settling much bigger questions about the nature of agency in general and the relationship between practical identity and agency in particular.[5]

References

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge: Harvard University Press.
Breyer, D. (2010). Reflective luck and belief ownership. *Acta Analytica, 25*(2), 133–154.
Breyer, D. (2013a). Ownership, agency, and defeat. *Acta Analytica, 28*(2), 253–256.
Breyer, D. (2013b). Knowledge, credit, and cognitive agency. *Pacific Philosophical Quarterly, 94*(4), 503–528.
Breyer, D., & Greco, J. (2008). Cognitive integration and the ownership of belief: response to Bernecker. *Philosophy and Phenomenological Research, 76*(1), 173–184.
Ekstrom, L. (1993). A coherence theory of autonomy. *Philosophy and Phenomenological Research, LIII*, 599–616.
Ekstrom, L. (2005a). Autonomy and personal integration. In J. S. Taylor (Ed.), *Personal autonomy: new essays on personal autonomy and its role in contemporary moral philosophy* (pp. 143–161). Cambridge: Cambridge University Press.
Ekstrom, L. (2005b). Alienation, autonomy, and the self. *Midwest Studies in Philosophy, XXIX*, 45–67.
Frankfurt, H. (1988). *The importance of what we care about: philosophical essays*. Cambridge University Press.
Greco, J. (2000). *Putting skeptics in their place.* Cambridge University Press.
Levy, N. (2006). Autonomy and addiction. *Canadian Journal of Philosophy, 36*, 427–448.
Setiya, K. (2013). Epistemic agency: some doubts. *Philosophical Issues, 23*, 179–198.
Sosa, E. (2011). *Knowing full well*. Princeton: Princeton University Press.
Watson, G. (2004). *Agency and answerability.* Oxford: Oxford University Press.
Wolf, S. (1990). *Freedom within reason*. Oxford: Oxford University Press.

[5] I would like to thank John Greco, Allan Hazlett, John Davenport, Todd Stewart, and an anonymous referee for their helpful comments on previous drafts of this article.