Received: 24 January 2023 | Accepted: 21 April 2023

DOI: 10.1111/ejop.12867

ORIGINAL ARTICLE

European Journal of Philosophy WILEY

Mental agency and rational subjectivity

Lucy Campbell[1] | Alexander Greenberg[2]

[1]University of Warwick, Philosophy, Social Sciences Building, University of Warwick, Coventry, UK

[2]University of Southampton, Philosophy, Faculty of Arts and Humanities, Southampton, UK

Correspondence
Alexander Greenberg, University of Southampton, Philosophy, Faculty of Arts and Humanities, Avenue Campus, Highfield Road, Southampton, SO17 1BF, UK.
Email: a.greenberg@soton.ac.uk

Funding information
Leverhulme Trust, Grant/Award Numbers: ECF-2018-075, ECF-2019-406

Abstract

Philosophy is witnessing an "Agential Turn," characterised by the thought that explaining certain distinctive features of human mentality requires conceiving of many mental phenomena as *acts*, and of subjects as their *agents*. We raise a challenge for three central explanatory appeals to mental agency––agentialism about doxastic responsibility, agentialism about doxastic self-knowledge, and an agentialist explanation of the delusion of thought insertion: agentialists either commit themselves to implausibly strong claims about the kind of agency involved in the relevant phenomena, or make appeals to agency which seem explanatorily redundant. The agentialist literature does not contain a clear answer to this Agentialist Dilemma, and we put it forward here as a core challenge for the Agential Turn. But we also accept the fundamental motivation behind the Agential Turn, its critique and rejection of a purely passivist and spectatorial conception of the human mind. We close by urging the recognition of a broader category of *rational subjectivity*, a category which includes states which are neither active nor passive, but nevertheless form part of a subject's rational point of view on the world.

1 | INTRODUCTION[1]

Over the past few decades, philosophy of mind has witnessed an "Agential Turn." Motivated by dissatisfaction with the idea that we are mere spectators or passive subjects of our own psychology, various human mental phenomena are increasingly viewed as *active*, and their subjects as their *agents*.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *European Journal of Philosophy* published by John Wiley & Sons Ltd.

But within these literatures, we also find a stronger type of claim. Many agentialists not only think that many features of our mental lives are active, but view this idea as doing significant explanatory work. Matthew Boyle nicely captures the essence of these stronger sorts of explanatory claim, in relation to belief, in the following passage:

> A point of persistent controversy in recent philosophical discussions of belief concerns whether we can exercise some sort of agential control over what we believe. … [T]he idea that we have some kind of discretion over what we believe has appealed to philosophers working in several areas. This idea has been invoked, for instance, to characterize the basic difference between rational and non-rational cognition, to account for our epistemic responsibility for what we believe, and to explain how we are able, normally, to say what we presently believe without relying on self-observation or inference (Boyle, 2009a, p. 122).

The idea Boyle refers to here is not just the claim that states like belief are "active" in some sense, but that viewing them as active will enable us to understand why they have further features, such as our responsibility for them or our self-knowledge of them.

We agree that a purely passive and spectatorial picture of the mind is untenable, and acknowledge the importance of the idea that a person's agency isn't exhausted by her bodily movements and their effects. We therefore agree that understanding the nature of human mentality in general will necessarily include recognising and understanding the roles our agency has to play within it. What we think is less clear is that mental agency can function as a general explanatory factor of the kind indicated by Boyle above. In this paper, we identify a dilemma for the idea, showing how it emerges for key explanatory appeals to mental agency across three different literatures, considering agentialist accounts of doxastic responsibility, of doxastic self-knowledge, and of the delusion of thought insertion.

In each case, mental agency is appealed to in order to explain why a person ordinarily or ideally relates in a certain way to a specific class of her mental phenomena––why she is *responsible* for and has *self-knowledge* of her beliefs, and why she experiences a *sense of ownership* over her episodes of conscious thinking, a sense which goes missing in thought insertion.[2] In each case, our challenge begins by observing that a bare unqualified appeal to agency is insufficient to explain the relevant *explanandum*. Although not a problem in its own right, the observation invites an initial *clarificatory challenge*: the agentialist must clarify the kind of agency they are appealing to, in such a way that its relevance to the *explanandum* is clear.

Agentialists do have the resources to respond to this question, at least initially, and we distinguish two strategies that can be found across all three literatures. But however it is developed, we argue that agentialism faces what we will refer to as the Agentialist Dilemma, coming up against one or the other of two problems:

> *The Problem of Strength*: the account commits itself to implausibly strong claims about the kind of agency characterising the relevant phenomena.
>
> *The Problem of Redundancy*: the appeal to agency isn't obviously what ends up doing the relevant explanatory work.

It is neither possible nor desirable to survey each and every version of agentialism across our three literatures. Instead, we focus in on certain particular accounts, which are either especially well-developed, or especially alive to the Problems of Strength or Redundancy, or both. Through these case studies, we aim to highlight the pervasiveness and persistence of the Agentialist Dilemma.

We are not offering a knock-down argument against the general agentialist explanatory strategy. For all we will say, it could be that some form of agentialism in each domain does have the resources, in the end, to escape the Agentialist Dilemma. If there is a variety of agentialism that manages to do so, we do not think this has been clearly spelled out in the literatures in question, in part because the Dilemma has not been explicitly posed. By identifying it,

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

showing its breadth of application, and considering in detail how particular agentialist suggestions might try and fail to escape it, we hope to clarify a key condition of adequacy on the kinds of general explanatory appeals to mental agency characterising a core strand of thought within the Agential Turn.

But what if agentialists cannot in the end provide a convincing answer to the Agentialist Dilemma? Must we therefore turn back to a passivist and spectatorial conception of the human mind? In our conclusion, we suggest that this would be an overreaction. We can perhaps appreciate the key insights of the anti-passivist critique motivating the Agential Turn, without overinflating the explanatory potential of mental agency, by elucidating a broader category of *rational subjectivity*, a category which includes states which are neither active nor passive, but nevertheless form part of a subject's rational point of view on the world.

2 | AGENTIALISM ABOUT DOXASTIC RESPONSIBILITY

It often seems appropriate to hold people accountable for what they believe[3]: we advise or persuade one another to believe or not believe this or that; we warn people not to believe everything they read on the internet; we encourage people to trust medical professionals over conspiracy theorists. When people believe (as we see things) badly, it can be appropriate to criticise them,[4] and we have dedicated epistemic concepts to express such criticism: someone thought to believe badly can be thought not only epistemically unfortunate, but *gullible*, *close-minded*, or *superstitious*.

Such observations have been taken to suggest that we are *responsible* for what we believe.[5] But what explains this responsibility? *Agentialism about doxastic responsibility* holds that a person is responsible for her belief insofar as it is a product of her agency.

2.1 | The clarificatory challenge, assimilationism, and expansionism

It might seem obvious that if our beliefs turned out to be exercises of our agency, then we would have a ready explanation of why we are responsible for them. After all, our actions are usually thought of as paradigm cases of things we are responsible for. However, we need to tread carefully. Simply being the agent of some phenomenon does not per se render one responsible for it. It may be unfair to hold a person responsible for taking off her trousers on the train if she is under hypnosis, or pathologically confused, for example, about where she is. If unbeknownst to you, some ne'er-do-well has laced the sugar with arsenic, then you are not responsible for poisoning the guests' tea. Yet these are both cases of "action" in *some* sense. So "action" is not per se responsibility-conferring.

This observation raises an initial clarificatory challenge for agentialism about doxastic responsibility:

> Given that agency is not responsibility-conferring *per se*, why should the claim that believing involves agency be relevant to explaining doxastic responsibility?

A natural reaction to the cases above is to point out that taking off one's trousers on the train whilst (e.g.) hypnotised, or poisoning one's guest unknowingly are not *intentional*, or *voluntary*, actions, or things done *out of choice*. It is "action" in these more robust senses for which we are paradigmatically responsible. And this line of thought suggests a first option for responding to the clarificatory challenge. According to this––what we will call the *assimilationist strategy*––believing is also something one can do intentionally, voluntarily, or out of choice. Appealing to *these kinds* of agency is relevant for explaining doxastic responsibility, because appealing to these kinds of agency is relevant for explaining responsibility more generally. Yet the idea that we might believe intentionally, voluntarily, or out of choice––a position known as doxastic voluntarism––is highly controversial, and many agentialists prefer a second approach to answering the clarificatory challenge. According to those pursuing what we will call the *expansionist*

strategy,[6] believing involves a *sui generis* form of nonvoluntary and nonintentional mental agency, which is nevertheless claimed to be responsibility-conferring.[7]

In the rest of this section we consider certain particular prominent assimilationist and expansionist accounts in the doxastic responsibility literature, and show how they land on one or the other horn of the Agentialist Dilemma.

2.2 | The assimilationist strategy

Although not commonly held, doxastic voluntarism has been defended. The doxastic voluntarism literature typically explicitly concentrates on the concepts of voluntariness or choice rather than agency,[8] and tends not to focus on doxastic responsibility directly, but on adjacent phenomena such as epistemic norms on belief, or epistemic praise and blame. But since this literature standardly treats doxastic responsibility as coming in a package with these other phenomena,[9] we can treat many doxastic voluntarists––especially those who argue that we can choose or decide what we believe in much the same way as we can choose to act––as offering a version of the assimilationist strategy. The view on offer is one on which belief is subject to one of the forms of robust, uncontroversially responsibility-conferring agency found in familiar and undisputed cases of action.

Doxastic voluntarism is often rejected by appealing to common sense. It is claimed to be self-evident that one cannot believe something just by deciding or choosing to do so, even if offered some very attractive practical inducement. To take William Alston's example, even half a billion dollars would seem powerless to move a person to believe, by a pure act of will, that the USA is still a British colony (1988, p. 263).[10]

Some doxastic voluntarists deny that such examples show that beliefs can never be formed voluntarily or by choice. Carl Ginet argues the case by pointing to cases with a certain structure, earlier identified by William James (1956, 2–4). In what we can call "James cases," a proposition and its negation are both "live hypotheses" (James, 1956, 3), and something of practical importance turns on which is true. Consider being faced with whether to believe that one has locked the front door of the house when one is fifty miles away from home, intent on having a good time (Ginet, 2001, p. 64). Ginet thinks that in James cases, one can choose to believe a proposition or its negation.

Does Ginet's suggestion provide a good assimilationist answer to the clarificatory challenge? One way or another, it seems to come up against the Problem of Strength. That is, if it is going to account for all cases of doxastic responsibility, it will have to rely on implausibly strong claims about the kind of agency characterising belief. To start with, it's not particularly plausible that people genuinely do have ability to choose what to believe even in James cases (see, e.g., Buckareff, 2004). However, even if we were to grant this, it does not obviously help. For cases of doxastic responsibility far outstrip James cases. It's standardly supposed that the vast majority of a person's nonalienated beliefs are supposed to be ones she is, or can be, responsible for. These include cases of believing despite clear countervailing evidence, failing to believe what is staring one in the face, and believing well in difficult epistemic circumstances (e.g., when "good evidence" and "bad evidence" might be hard to tell apart). These are not plausibly James cases.

In order to avoid this problem with Ginet-style doxastic voluntarism, the assimilationist would have to make the case that beliefs are voluntary in this much wider range of cases. In the doxastic voluntarism literature, we find different ways of making this move by philosophers giving different accounts how beliefs can be voluntary. Matthias Steup, for instance, claims that I count as making a "doxastic decision" whenever "I take a doxastic attitude *because* I brought an episode of epistemic deliberation to a conclusion with a verdict about what my evidence supports" (2000, p. 34). Say I conclude, after weighing up the evidence presented in court, that the defendant is guilty. Steup claims this is aptly described as a *decision to believe* that the defendant is guilty, because of its similarity with practical decision: both involve coming to a conclusion having weighed up one's reasons (2000, pp. 32–34). A slightly different tack is taken by Brian Weatherson, who argues that beliefs can be *voluntary* even if not *chosen*. Weatherson accepts that believing is not, in his terms, "volitional," that is, the result of choice or decision (2008, p. 543), but claims that one nevertheless believes "voluntarily" when one could have believed otherwise by exercising one's capacity to consider alternative hypotheses (Weatherson, 2008, p. 554). Weatherson also argues by analogy with the practical case,

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

claiming that some actions––those done out of frustration, for example––are voluntary but not chosen (2008, p. 548).[11]

These versions of doxastic voluntarism might avoid the Problem of Strength, as, unlike Ginet, the conceptions of agency they appeal to do not seem implausibly strong. However, in both cases this comes at the cost of running into the other horn of the Agentialist Dilemma, the Problem of Redundancy. For what Steup calls "deciding to believe that p," and Weatherson identifies as believing "voluntarily," seem to be nothing over and above judging that p either in response to an assessment of one's epistemic reasons (Steup), or in the context where one could have considered alternative hypotheses (Weatherson). If either phenomenon is able to explain doxastic responsibility, then this will be so quite independently of whether it is also viewed as a form of agency. Nobody denies the reality of the *phenomena* Steup and Weatherson identify as grounding doxastic responsibility (reasons-responsive judgment; the capacity to consider alternative hypotheses), and there would seem to be nothing to stop a sceptic about doxastic voluntarism helping themselves to these phenomena in explaining doxastic responsibility. The notion of *agency* therefore does not itself seem to do any essential work in the explanation.[12] This is the Problem of Redundancy, a problem which we will return to and discuss in more depth in relation to the expansionist strategy.

2.3 | The expansionist strategy

We turn now to the expansionist strategy for answering the clarificatory challenge. Expansionists will suggest the Problem of Strength in particular can be avoided by viewing the agency involved in responsible belief not in terms of voluntariness or choice, but as *sui generis*. Here we argue that although expansionists do avoid the Problem of Strength, they do so at the expense of facing the Problem of Redundancy.

Standardly, expansionists about doxastic responsibility defend a cluster of related claims. They claim that we are responsible for our doxastic phenomena (judgement and/or belief) because forming, revising, and/or maintaining our judgements and beliefs in response to epistemic reasons is itself a distinctive form of agency. In being active, reason-responsive, and (thereby) subject to responsibility, believing is contrasted with phenomenal states like feeling sick or experiencing the taste of Aperol––these latter being passive, a-rational, and outside the scope of one's responsibility.[13]

What should we make of expansionism? Initially, it appears to come up rather straightforwardly against the Problem of Redundancy, just as Steup's and Weatherson's assimilationist accounts did: if the fact that doxastic phenomena are exercises of the capacity to respond to epistemic reasons can explain doxastic responsibility, then it would seem able to do so whether or not we identify this capacity as an *agential* one. The agentialist element seems dispensable from an explanation of doxastic responsibility in terms of reasons-responsiveness without doing any damage to its explanatory power. The point would seem to be underscored by the presence in the literature of accounts which ground doxastic responsibility in reasons-responsiveness, yet explicitly deny that this responsiveness is a form of agency (Engel, 2009, pp. 215–217; Owens, 2000, pp. 123–126).[14]

Expansionists will object to the foregoing. An initial response is that our argument rests on a merely terminological dispute about whether or not agential terms should be understood broadly, as covering all forms of reasons-responsiveness, including responses to nonpractical purely epistemic reasons. Is it not unhelpfully flat-footed for us to restrict agential terminology to the narrower class of undisputed action, given that the broader definition can help make sense of doxastic responsibility?[15]

Although there is a danger of lapsing into merely terminological disputes in this area, the objection misfires. Our argument is neutral on how to delimit the scope of agential terminology. Our point is that understanding agency broadly, as including a person's responsiveness to purely epistemic reasons, secures no advantage for an explanation of doxastic responsibility purely in terms of reasons-responsiveness. If reasons-responsiveness can explain doxastic responsibility, it can do so whether or not it is thought of as a form of agency. If the debate between those who assert, and those who deny, that responding to epistemic reasons is a form of agency *is* merely terminological, this

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

could only bolster the suggestion that the notion of "agency" in play is not substantial enough to be doing any work over and above that already done by the notion of reasons-responsiveness in an explanation of doxastic responsibility.

A more promising strategy for the expansionist is to deny the possibility of pulling apart the notion of reasons-responsiveness from the notion of agency. If reasons-responsiveness *cannot but* be thought of as a kind of agency, then there is no room to claim that the explanatory work can be done by the notion of reasons-responsiveness *independently* of the notion of agency. If reasons-responsiveness and agency are conceptually inseparable, then the Problem of Redundancy will have been blocked. This kind of move is made by Conor McHugh, who offers two arguments for the claim that responding to reasons necessarily involves agency (2013, pp. 146–149), one intuitive and one more theoretical.

McHugh's intuitive argument rests on the observation that "things with respect to which the subject is essentially passive don't seem to count as done for reasons" (2013, p. 146). Having a headache isn't something one can do for reasons, he points out; nor, he suggests, is travelling in an automatic car while paralyzed––even if one happens to think that there *is* a good reason to go where one is going. McHugh concludes from this that where something *is* done in response to reasons, it cannot be passive. It must be active.

Is this a plausible argument? McHugh has identified at least two cases in which something is passive and not done for reasons. Does this support the claim that something passive––or something that is *not* active––can never count as a response to reasons? Not obviously. Consider *being sad that one's uncle has moved abroad*. One can be sad for reasons ("He's such a great guy, and I'll miss him so much"). But does this mean that one's sadness must be *active* rather than passive? It doesn't seem obvious that it must. (Later, we consider whether such states might be best thought of as *neither* active *nor* passive.) Such cases, we think, cast doubt on McHugh's intuitive argument that responding to a reason must involve agency.

McHugh recognises that not everyone will be swayed by his intuitive argument, and offers a second argument. He starts by asking what it takes for "something one does" "to count as an action" (2013, p. 148). Of course "something one does" must be understood thinly here, so that it is an open question whether what one "does" in believing something might count as an action. We can think of this thin sense of "doing something" as covering anything that might be denoted by an active verb. To make it clear when we are using this thin sense of "doing," we will continue to put it in quotes.

So when does something one "does" count as an action? McHugh gives the following answer:

> It is widely accepted that this involves causation by certain mental states of one's. These must be mental states whose contents are appropriately related to what one does, and which are of the right kind to bring one to do that thing in a way that makes one's doing it active. (McHugh, 2013, p. 148).

McHugh is appealing here to the idea that what makes something one "does" *active* is causation by (or a counterfactual connection to) certain rationalising mental states. Standardly, agency is theorised as dependent on causation by (or counterfactual dependence on) mental states which include desire and/or intention. But McHugh needs to give an account of what makes something one "does" active which can apply to the case of belief, which is properly responsive only to epistemic reasons. Presumably for this reason, McHugh identifies the relevant states in a way that does not make explicit mention of desire or intention:

> Very plausibly, one way in which mental states could play this role would be by having contents, and being of kinds, such that doing what one does makes sense from one's own point of view, and its so making sense is what brings one to do it, in a causally non-deviant way. (McHugh, 2013, p. 148).

McHugh seems to be suggesting here that it is plausibly sufficient for a set of mental states which suitably cause (or counterfactually support) one's "doing" something to render that "doing" *active*, that what one "does" is rendered

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

intelligible in light of the contents of those mental states. But it is unclear how this constitutes an argument for the claim that responding to reasons is necessarily an exercise of agency––rather than simply a re-statement of the claim in other terms. If this is right, then McHugh's 'second argument' does not provide any reason to accept that a phenomenon's being (nondeviantly) caused by states which rationalise it is sufficient for its being a case of agency.

To sum up, in defining the *sui generis* sense of agency they're interested in in terms of reasons-responsiveness, the expansionists considered in this section face the Problem of Redundancy. If the reasons-responsiveness of belief can explain doxastic responsibility, then it can seemingly do so independently of its identification as a form of agency. The only way to block this objection would be to argue that it's simply not possible to pull reasons-responsiveness and agency apart, so that reasons-responsiveness *cannot* explain doxastic responsibility independently of agency. McHugh's work is important in that he seems to recognise this demand. However, we have argued that he does not obviously manage to meet it.

Of course, there may be other things the expansionist may say to try and support the claim that responding to a reason is necessarily an exercise of agency. But if it is to provide a genuinely agentialist account of responsibility, it would need to be clear how the moves made themselves avoid falling into the Problem of Redundancy. That is, any suggested amendments would need to make it impossible for a doxastic agency sceptic to agree with the substance of the explanation of doxastic responsibility, whilst also avoiding falling back into the Problem of Strength.

The initial clarificatory challenge asked why an appeal to doxastic *agency* is relevant to understanding doxastic responsibility, given that simply being the agent of some phenomenon does not necessarily make one responsible for it. Both assimilationist and expansionist strategies for responding to the question come up against the Agentialist Dilemma. Expansionists face the Problem of Redundancy, whereas assimilationists face the Problem of Strength (at least initially––though we also saw how attempts by assimilationists to escape the latter problem can lead to the former).

3 | AGENTIALISM ABOUT DOXASTIC SELF-KNOWLEDGE

Our 'self-knowledge' of much of our mental lives has a distinctively first-personal character. Although theorists diverge over the precise details, this distinctive character is broadly agreed to include an apparent ungroundedness in inference or observation. A core task for an account of self-knowledge is to explain why it takes this form.

Mirroring the agentialist literature on responsibility for the attitudes, the agentialist literature on self-knowledge focusses centrally on doxastic phenomena, on belief and judgement.[16] We follow this lead, concentrating on agentialism about doxastic self-knowledge, such as––to take Gareth Evans's famous example (1982, p. 225)––a person's knowledge that she judges or believes that there will be a third world war. Agentialists about self-knowledge claim that we know about our own states of mind, in at least some cases, by actively bringing them about or sustaining them––by, in Richard Moran's highly influential phrase, *making up our minds* (2001, pp. 56–57, *et passim*).

According to this line of thought, the reason why doxastic self-knowledge is epistemically ungrounded in any of the usual ways is just the same as the reason why an agent's knowledge of her own action is not epistemically grounded in any of the usual ways. Intuitively, where what one knows is something which one actively brings about, there is simply no epistemological role for inference or observation to play. I do not need to be *informed* of the existence of something I have myself *created*.[17]

3.1 | The clarificatory challenge

The strategy of treating doxastic self-knowledge as a special case of agent's knowledge has some initial appeal and has recently been very influential. But more needs to be said about precisely how the idea is supposed to work. For it is clear that we do not have distinctively first-personal "agent's knowledge" of everything truly described as our

"actions." Susie gave away Government secrets to the enemy, thinking she was passing the information to an ally. The shrieking cat alerts you to the fact that you are standing on her tail. You know about your action only by inference and observation; Susie is completely unaware of hers.

Such observations prompt an initial clarificatory challenge for agentialism about doxastic self-knowledge, parallel to the challenge considered above for agentialism about doxastic responsibility:

> Given that agency *per se* does not bring with it distinctively first-personal (practical) knowledge, why should the claim that doxastic phenomena manifest agency be relevant to explaining distinctive first-personal doxastic self-knowledge?[18]

Like the doxastic responsibility literature, the doxastic self-knowledge literature splits up along assimilationist and expansionist lines. This gives us two possible responses to the clarificatory challenge for doxastic self-knowledge.[19]

3.2 | The assimilationist strategy

Although not all of a person's "actions" will be objects of nonobservational agents' knowledge, it *is* widely accepted that a person will (perhaps essentially, but at least typically) have distinctively first-personal practical knowledge of what she does *intentionally*. She who is intentionally (e.g.) stockpiling baked beans *will* know that she is, and furthermore, this knowledge will be ungrounded in inference or observation (Anscombe, 2000). She doesn't need to see which tins she picks from the supermarket shelf, how many she takes, or what she does with them afterwards, in order to know that she is stockpiling baked beans. This suggests an assimilationist strategy for answering the clarificatory challenge: spell out a sense in which doxastic phenomena involve *intentional* mental action, and explain doxastic self-knowledge in terms of the practical knowledge this brings with it.

We focus here on Antonia Peacocke's (2017) version of assimilationism. Peacocke's account is interesting because although she wants to explain doxastic self-knowledge in terms of intentional mental action involved in belief-formation, she is explicitly mindful of what we have termed the Problem of Strength.[20] Unlike the assimilationists about doxastic responsibility considered above, Peacocke explicitly rejects a straightforward form of doxastic voluntarism.

Three key moves underwrite her approach. The first is to initially focus on *judgement*, rather than on belief. Judging is a mental act, on Peacocke's view, whereas believing is not an act of any kind. Self-knowledge of belief is thus to be understood derivatively in terms of self-knowledge of judgement, which is explained in terms of mental agency. The relationship between self-knowledge of judgement and of belief is an important aspect of Peacocke's account, but we set it aside,[21] because mental agency plays its most central role in her account of self-knowledge of judgement.

Peacocke's second move is to identify a sense in which we can judge as an intentional action, despite the falsity of doxastic voluntarism. What is not possible, she thinks, is to "decide at will the precise content of one's judgments" (2017, p. 363), or "without regard for the truth of some proposition p, will yourself to judge that p" (ibid.). Nevertheless, Peacocke suggests that we *can* act intentionally under higher-level descriptions of the activity involved in judgement:

> Judgment can be intentional insofar as you can set out to judge some things, rather than, say, imagine some things. That is what you do when you set out to determine what's true. You can also set out to make a judgment that matches some particular content criterion—e.g. [...] a judgment whether p (2017, p. 362).

We can mark this distinction with the labels "judging-whether" and "judging-that."[22] Judging-that is an event, of forming or expressing a belief that p; judging-whether, by contrast, is an "extended mental task" (2017, p. 365).

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Peacocke thinks that to claim that we can intentionally judge-*that p* would be to commit to a problematic form of doxastic voluntarism on which one can choose at will which proposition to believe. Whilst one cannot decide or set out to judge-*that* there will be a third world war, or do this intentionally, one *can* decide or set out to judge *whether* there will be a third world war. One can set one's mind to the task of answering this question, and pursue this task intentionally––for example in order to help one decide whether or not to start trying for a baby. When one intentionally engages in the task of judging-*whether p* (e.g., whether there will be a third world war), one will have practical knowledge of what one is doing, *qua* intentional action.

Peacocke's third move is to explain how practical knowledge that one is judging-*whether p* can also deliver distinctively first-personal doxastic self-knowledge that one is judging-*that p*, even though only the former is done intentionally. The idea is that when one is intentionally engaged in the "extended mental task" of judging-whether *p*, one has nonobservational practical knowledge that *judging* is what one is up to, in relation to the question whether *p*. If, in this context, a relevant content (*p* or not-*p*) is tokened in one's consciousness, one's practical knowledge that *judging* is what one is up to, and that the topic of one's judgement task is *whether p*, will ensure that this content (*p* or not-*p*) will have been "antecedently conceptualized" (2017, p. 363) *as* the content of a judgement. That is, it will be conceptualised as a judgement-*that p*. On Peacocke's view, one's nonobservational practical knowledge that one is (intentionally) judging-whether *p* combines with consciousness of *p* (or not *p*), as the content one is tokening in the context of this task, to deliver nonobservational doxastic self-knowledge that one is (nonintentionally) judging-that *p* (or not *p*).

Does Peacocke's assimilationism satisfactorily answer the clarificatory challenge? We argue that it does not––in two stages. In the first stage we show that tokening *p* as part of an intentional (and practically known) task of judging-whether *p* is (*contra* Peacocke) insufficient for conceptualising it as a judgment-that *p*. Peacocke's account thus needs supplementation. In the second stage we distinguish two ways in which one might attempt to supplement it, arguing that one faces the Problem of Strength and the other the Problem of Redundancy. Peacocke's assimilationism thus comes up against the Agentialist Dilemma.

To see the problem, consider the following example of how a given instantiation of the task of judging-whether there will be a third world war might go:

Third World War

Step 1: You set out to judge whether or not there will be a third world war (out of a pure desire for knowledge, or with some further aim).
Step 2: You do some research, talk it through with experts, consider the evidence, etc.
Step 3: You judge, worriedly (1) "<u>There will be a third world war</u>".
Step 4: With increasing horror, you think: (2) "Good grief, surely <u>there won't be a third world war!</u>"
Step 5: Meticulously, you review your evidence, hoping to find something that will change your mind …
Step 6: But nothing undermines your initial conclusion, which you reiterate, now resigned: (3) "There's no escaping it, <u>there will be a third world war</u>"

There are three conscious tokenings of relevant thought-contents here:

1. *That there will be a third world war*
2. *That there won't be a third world war*
3. *That there will be a third world war*

As we are envisioning *Third World War*, tokenings (1) and (3) express judgements; (2) does not. This is not a case of epistemic indecisiveness, of judging that *p*, and then not-*p*, and then *p* again. Obviously such indecisiveness is

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

possible, but it's also quite possible to judge that *p*, then––while continuing to accept this verdict––token the content not-*p* in a non-judgmental mode, before finally reaffirming one's continued judgement by tokening *p*. We stipulate that *Third World War* has this latter structure: in (2), not-*p* is tokened as an expression of horror prompted by your judgement that *p* (in 1). This horror prompts you to review your evidence, not because you now doubt your initial judgement, but rather in the desperate hope that you might *find* some reason to doubt it.

There needn't be anything wrong with your self-knowledge in *Third World War*. So we can assume that you conceptualise the conscious tokenings in (1) and (3) *as* judgements that there will be a third world war, and the tokening in (2) as an expression of horror at the prospect that there will be one, and not also as a judgment that there won't be one (after all, there's nothing horrifying about the idea that there *won't* be a Third World War).

Third World War shows that, rather generally, consciously tokening a content *p* (or not-*p*) as part of an ongoing intentional task of judging-whether *p* is insufficient for conceptualising that content *as* a judgement-that *p* (or not-*p*).

So Peacocke is left with a question: what is it about the tokenings in (1) and (3) which explains why *they* are conceptualised as judgements, whereas (2) is not, even though all three occur as part of one's intentional, practically known, action of judging-whether there will be a third world war? There only really seem to be two options, but each option leads Peacocke onto one of the horns of the Agentialist Dilemma.

The first option is to suggest that your capacity for practical knowledge can itself discriminate between tokening a content *qua judgement*, and tokening a content *qua expression of horror*. But this is in effect to claim that one has *practical* knowledge of *judging-that* there will be a third world war in (1) and (3), and of expressing horror in (2). One has practical knowledge only of what one does intentionally. So the suggestion runs into the Problem of Strength: it entails that a person *judges-that p* (and not only *whether* p) intentionally. As we have seen, this kind of doxastic voluntarism is precisely what Peacocke's account is designed to avoid.

The second option is to claim that tokenings can be conscious not just in relation to their content (as Peacocke initially seems to suggest), but also in relation to their attitudinal component: (1) and (3) appear in consciousness as *judgements*, whereas (2) appears as an *expression of horror*. This avoids the Problem of Strength, since it avoids the suggestion that judging-that p is known *practically* (and so the implication that it is done intentionally). But it clearly faces the Problem of Redundancy. For if a judgement's status *as* a judgement is simply given in consciousness, then this will be so whether or not it occurs as part of an extended intentional mental action of judging-whether, such as that described in *Third World War*. And the phenomenon of mental agency is left without any work to do.

3.3 | The expansionist strategy

In the agentialist literature on self-knowledge, assimilationist approaches like Peacocke's are the minority position. The more prevalent approach is an expansionist one, the central idea of which is that believing involves a *sui generis* form of agency, distinct from intentional action *per se*. Here we concentrate on an especially influential version of the view, inspired by the work of Richard Moran (2001, Chapter 4, 2003, pp. 402–406), as further developed by Matthew Boyle.[23]

In addition to rejecting assimilationism in favour of expansionism, Boyle differs from Peacocke in another important way. Peacocke, recall, thinks that beliefs are subject to agency only derivatively, because they are precipitated by judgements, where only the latter are genuine mental acts. Boyle holds that our beliefs are *direct* expressions of our agency. He does not deny that beliefs are states rather than events. But he suggests a sense in which belief is both stative *and* active: believing is "actively being a certain way" (2010, p. 19); in Aristotelian terms, believing is an "*energeia*"––the stative actualisation of a capacity (2010, pp. 19–21).

Not all *energeiai* are "active." Some––pains and sensations, for example––are the actualisations of *passive* capacities, and so forms of patiency (Boyle, 2009b, pp. 157–161). So Boyle must do more than categorise believing as an *energeia* to elucidate the *sui generis* form of agency involved in belief. He argues that believing is the *energeia* of an *active* capacity, by characterising the genus *rational agency* in such a way that believing comes out as a species (with

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

intentional action being a distinct species). Rational agency is embodied in a "general relationship [which] holds between endorsement and actuality" (2009a, p. 141), according to which, "[one's] present endorsement of X-ing is the ground of [one's] present X-ing, in virtue of a capacity [one] possess[es] to be through the former the source of the latter" (ibid.). When this condition is met, on Boyle's classification, "[one is] the *agent* of [one's] X-ing, and X-ing is [one's] *act*." (ibid.).

Intentionally φ-ing fits this schema insofar as it is φ-ing "in actualization of a capacity to do what [one] represent[s] as to-be-done" (ibid.). But Boyle argues that the same is true of belief, insofar as one believes "in actualization of a capacity to believe what [one] represent[s] as to-be-believed" (ibid., 142). In this way, for Boyle, intentional action (both bodily and mental) is one species of rational agency (manifested in *processes* or *events*), and belief is a distinct species (manifested in [stative] *energeiai*).

So far we have been setting out Boyle's version of what we can call agentialism about *belief*. This is the claim that believing is an exercise of a person's agential capacities, and for the purposes of this paper we do not have any objection to this idea (although we will make some comments relevant to the position in our conclusion). Our target is the stronger claim that thinking of beliefs as exercises of one's agency *explains doxastic self-knowledge*. It is this stronger explanatory claim that we have been referring to as "agentialism about doxastic self-knowledge."[24] Let's now see how Boyle defends the stronger claim.

At the centre of Boyle's account is a notion of *endorsement*.[25] It is by continuing to endorse *p* as "to be believed" that one sustains the actuality of one's belief that *p*, and it is because of this endorsement-dependence that believing fits Boyle's schema for rational agency. Importantly, Boyle views this endorsement as *self-conscious*,[26] and this would seem to be an ineliminable aspect of the account. It is only clear how the dependence of one's belief on one's endorsement of a proposition as-to-be-believed would generate doxastic self-knowledge if this endorsement is self-conscious. Boyle's response to the clarificatory challenge is thus as follows: the agency manifested in believing is relevant to understanding doxastic self-knowledge, despite not being intentional agency, because at its core is an essentially self-conscious element, the endorsement of a proposition as to-be-believed.

We will grant, for the sake of argument, that doxastic self-knowledge can be explained by the dependence of belief on self-conscious endorsement of a proposition as to-be-believed. The problem for Boyle is that acknowledging this leads his expansionist account straight into the Problem of Redundancy. For if the idea that believing that *p* involves self-consciously endorsing *p* as to-be-believed can explain the distinctively first-personal character of doxastic self-knowledge, then it can do so whether or not this endorsement is viewed as an expression of the subject's agency. The capacity of the account to explain doxastic self-knowledge in terms of the endorsement-dependence of belief would appear to be independent of Boyle's characterisation of belief as a species of rational agency.

Unsurprisingly, the natural responses to our argument here mirror those which came up in relation to our objection to McHugh's expansionist account of doxastic responsibility in Section 2.3. Initially, the Boylean expansionist might dismiss our objection as resting on a merely terminological dispute over whether to classify all forms of endorsement-dependence as exercises of agency, or by contrast, to reserve agential terms for phenomena which are dependent on specifically *practical* forms of endorsement (i.e., those contained in intention or practical reason).

But the objection here misconstrues our argument against Boyle in the same way that the parallel objection in Section 2.3 misconstrued our argument against expansionists about doxastic responsibility. We are not objecting to classifying belief as "active." Our claim is only that doing so makes no difference to one's ability to explain doxastic self-knowledge in terms of the self-conscious endorsement of a proposition as to-be-believed. If understanding the relevant kind of endorsement as an agential phenomenon were to give Boyle's account an advantage over a non-agentialist, but still endorsement-based, account of doxastic self-knowledge, the difference between an agentialist endorsement-based account and a non-agentialist endorsement-based account would have to be substantive. But if the dispute between an agentialist endorsement-based account and a nonagentialist endorsement-based account is purely terminological, then the difference is not substantive. Therefore, our objection is supported, not undermined, by the suggestion that the difference between Boyle's agentialism and a pure endorsement-based account is merely terminological.

At this point, the Boylean expansionist is likely to deny that we can conceptually separate endorsement-dependence from agency in the way that our argument requires. Again, this response parallels one we considered in Section 2.3. The objection is that we cannot (as we have for the sake of argument) agree that the idea of *endorsement-dependence* explains doxastic self-knowledge, whilst denying that the idea of *doxastic agency* explains doxastic self-knowledge, because these "two" ideas are really one and the same. In effect, the claim here is that the identification of the genus "rational agency" as endorsement-dependence (rather than dependence on *practical* endorsement) is a conceptual truth.[27]

But it at least unclear that this is a conceptual truth. If it were, then anyone who thinks that believing is dependent on endorsing a proposition as to-be-believed, but does not think that believing is a form of agency, would have to be conceptually confused––about belief, and/or endorsement, and/or agency. Yet such a view seems to be rather commonly held, and it is hard to charge its proponents not simply with being mistaken, but with incoherence.

It is important to see that the endorsement-dependence we find in belief contrasts starkly with the endorsement-dependence we find in intentional action on exactly this point. The endorsement-dependence internal to intentional action is the dependence of what one does on one's practical reasoning and/or intention, where practical reasoning and intention are the relevant forms of (practical) endorsement. The endorsement-dependence in intentional action is thus identical with the execution or carrying out of intention, which is itself identical with intentional action. Denying that the endorsement-dependence involved in believing amounts to a form of agency only commits one to denying that believing is an exercise of agency. By contrast, denying that the endorsement-dependence involved in intentional action amounts to a form of agency commits one to denying that intentional action is an exercise of agency. A person who denied that the endorsement-dependence involved in executing an intention was a form of agency really would be conceptually confused (to say the least). Someone who denies that the endorsement-dependence involved in believing manifests agency does not seem anywhere near as incoherent, even if they are in fact mistaken. The current response on Boyle's behalf would seem to view these positions as equally confused. But they are not. So the response should be rejected.

Our initial clarificatory challenge asked: given that (rather generally) there are forms of agency which are not self-known, how exactly is the idea of doxastic agency supposed to help us understand doxastic self-knowledge? Peacocke offered an assimilationist answer to this question, appealing to intentional mental action in judgment. We argued that on its own her account failed, and that the need to supplement it led her onto the horns of the Agentialist Dilemma. Boyle's expansionist approach posited a *sui generis* form of mental agency, but we suggested that it faces the problem of Redundancy.

We have not considered all of the possible ways in which our argument might be responded to––whether by Peacocke, Boyle, or distinct kinds of assimilationism or expansionism. We do hope to have made it clear, however, how the Agentialist Dilemma poses an initial challenge to these two very different agentialist accounts of doxastic self-knowledge. As we found in relation to agentialism about doxastic responsibility, escaping the Agentialist Dilemma is not straightforward.

4 | THE AGENTIALIST EXPLANATION OF THOUGHT INSERTION

In the delusion of "thought insertion," a person experiences certain episodes of conscious thinking as implanted in their mind––often by someone else. The following clinical quotations are commonly cited as typical:

> Thoughts come into my head like "Kill God." It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts." (Quoted in Frith, 1992, p. 66).

> I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his.... He treats my mind like a screen and flashes his thoughts into it like you flash a picture (Mellor, 1970, p. 17).

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

How should we understand such cases? The standard first move is to claim that the delusion involves lacking a sense of "ownership" or "mine-ness" in relation to one's own thoughts. The subject is aware of the relevant thoughts and describes them as "in" their mind. Yet they deny that the thoughts are "their own." The delusion of thought insertion arises, it is claimed, when the sense of ownership ordinarily accompanying one's episodes of conscious thinking goes missing. Without this sense of ownership, a person may come to view their thoughts as "belonging" to someone else––in the clinical vignettes above, to this chap Chris, and to Eammon Andrews––who is then envisaged as having inserted (sometimes "beamed" or "flashed") the thoughts into the subject's mind.[28]

The second standard move is to explain why the sense of ownership goes missing in these cases. A central hypothesis is that people lack of a sense of ownership in thought insertion because they lack a sense of agency over their thinking in these cases, a sense of agency which in ordinary consciousness is present. Call this the *agentialist explanation of thought insertion*.

4.1 | The clarificatory challenge

The agentialist explanation of thought insertion has a slightly different structure to agentialism about doxastic responsibility and about doxastic self-knowledge. For one thing, the purported *explanans* is not agency per se, but the "sense" of agency, and the *explanandum* not a normative or epistemic relation to some mental phenomenon, but its attendant phenomenology. In addition, where the agentialisms considered in Sections 2 and 3 appeal to the *presence* of mental agency to explain the *presence* of responsibility and self-knowledge in *ordinary* cases of belief, the agentialist explanation of thought insertion appeals to the *absence* of (the sense of) agency to explain the *absence* of (the sense of) ownership in *pathological* cases of conscious thinking. We reflect these differences by terming the account considered here the "agentialist explanation" of thought insertion rather than "agentialism about" thought insertion.[29]

Despite these differences, the agentialist explanation of thought insertion comes up against the very same kind of problem as the versions of agentialism considered above: an initial clarificatory challenge arises, which can be given either an assimilationist or an expansionist answer. Yet neither approach obviously escapes the Agentialist Dilemma between the Problem of Strength and the Problem of Redundancy.

The clarificatory challenge arises for the agentialist explanation of thought insertion when we observe that there are various examples of apparently lacking a sense of agency without lacking a sense of ownership over one's conscious episodes. Sensations are one obvious example, but more pertinently to a discussion of thought insertion, the same seems true of certain episodes of thinking. Busy at the office, you suddenly find yourself accosted by the thought *that you left the oven on*. It "pops into your head." You seem to be on the receiving end of it rather than being its agent (compare concluding that you've left the oven on after effortful memory-work). But there is no lack of a sense of ownership over such thoughts. Prima facie, there even seem to be broadly pathological forms of thinking which lack a sense of agency, without lacking a sense of ownership. A person with OCD, for example, finds her obsessive thoughts uncontrollable––not under her agential control––but may still experience them as *hers*. Indeed, plausibly part of what is distinctively troubling and frustrating about OCD is the feeling of lacking control over an aspect of *one's own* mind.[30]

These and similar observations give rise to the initial clarificatory challenge for the agentialist explanation of thought insertion, as follows:

> Given that we can lack a sense of agency over conscious thinking without lacking a sense of ownership over it, why should appealing to the lack of a sense of agency be relevant to explaining the lack of a sense of ownership in delusions of thought insertion?

4.2 | The assimilationist strategy

Chris Frith defends an agentialist explanation of thought insertion, which can be viewed as a version of assimilationism. Frith, following Irwin Feinberg (1978), applies the standard cognitive-psychological model for action to ordinary conscious thinking. The resulting account affords an explanation of thought insertion which mirrors the standard account of pathologies of ownership in overt action, such as alien hand syndrome.

On Frith's model (1992, pp. 80–83), action (bodily movement; thinking) is initiated by a motor instruction, which at the same time sends an "efferent copy" of itself to a central monitoring mechanism (1992, p. 81). This mechanism also receives feedback from the action itself (through perception or proprioception in the case of bodily action, and introspection in the case of thinking). If the contents of the feedback and the efferent copy mismatch, then central monitoring feeds this back to the motor system, which issues further motor instructions (and further efferent copies). This "comparator mechanism" functions to keep actions on track.

Frith hypothesises that the experience of thought insertion stems from a breakdown in the comparator mechanism, where the efferent copy is either not sent, or not properly used by central monitoring. The person retains introspective knowledge of her thought, but because there is no efferent copy for this feedback to match up to, she experiences this thought as alien, and attributes it to someone else.

The comparator model itself describes subpersonal mechanisms. But if it is to explain the *experience* of thought insertion, these goings-on need to be manifested somehow in consciousness. This is where Frith suggests a form of assimilationism:

> Thinking, like all our actions, is normally accompanied by a sense of effort and deliberate choice as we move from one thought to the next. If we found ourselves thinking without any awareness of the sense of effort that reflects central monitoring, we might well experience these thoughts as alien and, thus, being inserted into our minds (Frith, 1992, p. 82).

The assimilationist suggestion in this passage is that normal conscious thinking is accompanied by the very same sense of agency that we find in undisputed cases of overt action––for Frith: "a sense of effort and deliberate choice."

Frith's assimilationism pretty clearly comes up against the Problem of Strength. Consider again our cases from above, the OCD sufferer, and the person who has the thought "I've left the oven on" pop into her head unbidden. Frith would seem to be committed to the claim that in these cases, either the subject is *thinking with a sense of effort and deliberate choice*, or that they *do not experience these thoughts as their own*. Yet neither claim seems true.

Having the thought "I've left the oven on!" pop into one's head while busy thinking about other things is clearly not experienced as chosen or effortful. One experiences such thoughts as *imposing* on one, as *interrupting* one's active thinking. Yet such thoughts are in no sense experienced as anything other than one's own. Again, if the OCD sufferer experiences a sense of effort in relation to her obsessive thoughts, this can only be the effort of trying to block or ignore them, rather than any sense of effortfully thinking them. It is true that OCD sufferers sometimes represent their ownership of obsessive thoughts in fractured or ambivalent ways.[31] But what is not plausible is to think that a person's obsessive thinking in OCD are experienced as *not her thinking*. In OCD, one is surely alienated from the *contents* of one's thoughts, in the sense that one does not uncomplicatedly endorse these contents. But being "alienated" from the contents of one's thoughts in this sense is not the same as, and does not suffice for, experiencing the *thinking* of those thoughts as not one's own thinking. So despite the complexities of the phenomenology of obsessive thinking in OCD, the clear lack of "a sense of conscious effort" does not plausibly undermine the sense of ownership.

Frith's identification of the sense of agency with a sense of "effort and deliberate choice" therefore clearly faces the Problem of Strength. We think that other assimilationist approaches will face similar problems. For this reason,

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the expansionist strategy is more likely to appeal to agentialists seeking to understand thought insertion. We turn to this now.

4.3 | The expansionist strategy

Expansionists in the thought insertion literature grant that we don't have the same kind of sense of agency over thinking as we do over undisputed cases of voluntary or intentional mental and overt action. Instead, they ground the sense of ownership in a *sui generis* sense of agency, suggested to be present in ordinary episodes of thinking, but missing in delusions of thought insertion.

John Campbell develops an account of thought insertion which draws on aspects of Frith's comparator model for conscious thought, whilst dropping Frith's assimilationism. We can read Campbell as an expansionist. On Campbell's account, the motor instructions which produce conscious thoughts are entirely subpersonal, and unaccompanied by any Frithian sense of effort or deliberate choice. They are caused directly by one's background mental states and external stimuli (1999, pp. 616–617), and importantly, do not arise from any intervening conscious intention to think those very thoughts (1999, p. 216). But Campbell agrees with Frith that the comparator model can underwrite an agentialist explanation of thought insertion:

> You have knowledge of the content of the thought only through introspection. The content of the efferent copy is not itself conscious. But it is match at the monitor between the thought of which you have introspective knowledge and the efferent copy that is responsible for the sense of being the agent of that thought. (J. Campbell, 1999, p. 618).

When the comparator mechanism breaks down, however, the subject "find[s] that he is introspectively aware of a thought without having the sense of being the agent of that thought" (1999, p. 618).

Whether Campbell provides a convincing expansionist response to the clarificatory challenge is hard to assess, because there is a crucial gap in his account. Expansionist accounts must firstly unpack the *sui generis* notion of agency they are appealing to, and secondly show how that notion of agency does the relevant explanatory work. Campbell doesn't seem to have carried out the first step, because although he rejects Frith's characterisation of the relevant "sense of agency" as a sense of effort, choice, or intention, he says nothing to positively characterise it. For this reason, there would appear to be nothing in Campbell's account to differentiate it from the view of someone who agrees with him about the sub-personal mechanisms underlying thought, agrees in rejecting Frith's assimilationism, but thinks that the subpersonal goings-on (normally) generate a sense of thought-ownership directly, rather than via the production of a sense of agency.

Campbell does claim that the sub-personal mechanisms underlying thinking––and our awareness of thinking––resemble those underlying bodily agency. But describing a subject as having a "sense" of agency must involve more than this. It must involve such a sense being describable at the personal level in agential terms. Frith's claim that a match detected by the comparator is manifested in consciousness by a sense of effort and choice does amount to such a description, notwithstanding the fact that it leads immediately to the Problem of Strength. But in rejecting this characterisation, and failing to put anything else in its stead, Campbell gives us no sense of the phenomenology generated by the comparator mechanism as distinctively *agential*. In order to answer the clarificatory challenge, Campbell would need to say more about the "sense of agency" putatively generated by the comparator mechanism. And to *convincingly* answer the clarificatory challenge, he would need to unpack this sense of agency in a way that didn't fall victim to the Agentialist Dilemma. Campbell's denial of Frith's claim that thinking is normally accompanied by a sense of effort and choice thus only provides an incomplete version of the expansionist strategy.

A number of other expansionists provide materials which might be thought to plug this gap. G. Lynn Stephens and George Graham, for example, claim that "a person's sense that he is the thinker or agent of his *mental* activity—of his conscious thoughts and feelings— … depends on his conviction that his occurrent mental episodes express his

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

intentional states", e.g., his beliefs and desires (2000, p. 165). Hanna Pickard, drawing on Richard Moran's work on self-knowledge, claims that we are agents in relation to our nondelusional thoughts because such thoughts reflect our ability to make up our minds through deliberation about reasons (2010, pp. 64–67). Patrizia Pedrini claims that we are agents in relation to our nondelusional thoughts because we are "willing to take responsibility" for them, where "taking responsibility" is understood as a readiness either to give reasons for them or to explain them by reference to nonrational causes (2015, pp. 224–227). Each of these accounts claims that the sense of ownership over one's thoughts goes missing in those suffering delusions of thought insertion because the relevant sense of agency is lacking; that is, because such subjects can't explain their inserted-seeming thoughts' occurrence as expressions of their beliefs and desires (Stephens & Graham, 2000, p. 162); or, because these thoughts "seem to be manifestations of mental states that they do not [...] endorse" (Pickard, 2010, p. 67); or, because they do not "feel any disposition to take responsibility" for these thoughts (Pedrini, 2015, p. 227).

These expansionist accounts all improve on Campbell's by providing a positive personal-level gloss on the *sui generis* "sense of agency" present in normal "own-seeming" thinking. However, they would all seem to face the familiar Problem of Redundancy, for reasons familiar from our discussions of doxastic responsibility and self-knowledge. If the sense of ownership which someone suffering delusions of thought insertion lacks over her thoughts can be explained by those thoughts being at odds with her intentional states, her assessment of reasons, or her sense of responsibility, then this is so whether or not these underlying phenomena are also taken to generate a *sui generis* "sense of agency." That is, the identification of these *explanantia* as generating a sense of agency over one's thinking turns out to be wholly independent of their capacity to explain the sense of *ownership* one is claimed to ordinarily have over one's own conscious thinking.

Unsurprisingly given this, and in parallel to some of the discussion in earlier sections, attention to the literature on thought insertion shows that a number of theorists do in fact provide explanations of thought-insertion in terms of a lack of *the sense of ownership* of conscious thinking––understood in terms like those employed by Stephens and Graham, Pickard, and Pedrini––whilst denying that a lack of a sense of agency has any role to play in the explanation. For example, Lisa Bortolotti and Matthew Broome (2009), and Jordi Fernández (2010) claim that what goes missing in thought insertion is "endorsement" of one's thoughts. They unpack this "endorsement" in the same terms as we find in Pickard's agentialist account––i.e., endorsed thoughts are those which reflect one's capacity to make up one's mind through deliberation about reasons––but they deny that this generates a sense of agency (Bortolotti & Broome, 2009, pp. 212, 219–22; Fernández, 2010, pp. 77–79). And there is no clear reason that we can see why someone who denies that the sense of ownership is to be understood as grounded in a *sui generis* sense of agency could not similarly accept the underlying substance of what Stephens and Graham, or Pedrini, say, without being any explanatorily worse off.

The agentialist explanation of thought insertion is slightly different in structure to agentialism about doxastic responsibility and self-knowledge. Yet here too we find the very same difficulties. Assimilationists in the literature are committed to holding that the sense of thought-ownership is a sense of agency of a strong and familiar kind–– for Frith, a sense of deliberate effort and choice. And it is simply implausible that one's thinking is own-seeming only if it is accompanied by a sense of agency in this sense: the Problem of Strength. Expansionists posit a weaker, *sui generis*, sense of agency in conscious thinking. Campbell doesn't say enough to positively characterise this phenomenon, although Pickard, Pedrini, and Stephens and Graham all try to do so. We have been neutral on whether any of their respective suggestions can successfully underwrite the presence of a sense of thought-ownership––so that when these phenomena go missing, this would also explain the loss of a sense of ownership commonly suggested to explain the delusion of thought insertion. But say they can. They nevertheless run into the Problem of Redundancy. For the underlying phenomena they use to characterise this *sui generis* sense of agency can then be appealed to directly in an explanation of the sense of thought-ownership in normal thought, and its absence in thought insertion. And the idea of a sense of agency is left looking like an idle cog.

As in the discussions above, we do not claim here that nothing more can be done to avoid the Agentialist Dilemma. But it would be a requirement on providing a successful agentialist explanation of thought insertion to

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

clarify how it does so. And absent such an explanation, we should question the helpfulness of the appeal to a "sense of agency" for understanding thought ownership and thought insertion.

5 | CONCLUSION

We have explored three different literatures, and have argued that attempts to provide a general explanation of some a feature of our mentality by appealing to mental agency all seem to face a common dilemma. When trying to explain why an appeal to agency should be relevant to the *explanandum*––that is, when trying to answer our initial clarificatory challenge––agentialist explanations *either* commit themselves to implausibly strong claims about the kind of agency characterising the relevant mental phenomena (the Problem of Strength), or fail to make clear why the relevant explanatory work is done by the appeal to agency––rather than by some distinct phenomenon (the Problem of Redundancy).

The pervasiveness and stubbornness of the Agentialist Dilemma suggests that it represents a general challenge for explanatory appeals to mental agency. We do not rule out the possibility that there might be some agentialist explanations of the phenomena we have discussed which avoid both horns of the dilemma. But we do hope to have made a case for thinking that any fully developed agentialist account will need to clarify how it avoids the dilemma.

We want to end by considering what would follow if, in the end, the Agentialist Dilemma cannot be overcome. Would that mean that the Agential Turn was all along a mistake?

We certainly do not think that we should return to the purely passivistic and spectatorial conception of the human mind which agentialists are reacting against. We agree with agentialists that such a picture cannot accommodate important distinctions, such as that between broadly rational and broadly a-rational aspects of mind, or between ordinary and alienated belief or conscious thought. It might be suggested that accepting this point is already sufficient to motivate agentialism––even if the Agentialist Dilemma lacks a solution. The inadequacies of the passive picture of the mind can make agentialist accounts of mental phenomena like belief and conscious thought seem inevitable.

This suggestion is not unreasonable. But nor, we think, is it irresistible. The fact that there are reasons to reject the purely passive spectatorial picture of the mind only supports appealing to mental agency as a general explanatory factor on the assumption, as yet unargued-for, that the active-passive distinction is exhaustive.

The claim that the active-passive distinction is exhaustive is not obvious, and it certainly cannot be assumed without argument. We suggested earlier that certain emotions––like sadness over one's uncle's having moved abroad––plausibly represent states of mind which are both broadly rational, and nonactive. We now want to build on this thought to make an initial case for acknowledging a category of *rational subjectivity*, which is not helpfully thought of in terms of either agency or patiency.

Consider, first, someone who in later life regrets having not become a professional musician (cf., Setiya, 2017, p. 84). Her regret is mild, neither overwhelming her nor haunting her, but nor is it something she can completely dismiss. She experiences it as an ongoing sad recognition that she missed an opportunity for another kind of life, which might have been fulfilling (not that her own life is particularly unfulfilling––and of course who really knows). Her regret is, importantly, responsive to reasons; it's not a sadness that comes out of nowhere. Yet it needn't structure her choices, or impact how she values her actual life and its components or twists and turns, and she wouldn't describe it as of any great importance to her, although she is not alienated from it.

Or, think of someone's grief at the death of his dog. He does not feel "lumped with" this grief or experience it as an imposition. But nor does he wallow in it, or foster it. He neither particularly welcomes it nor resists it. He simply sits with it. For all this, it is rationally intelligible. It is a rational response to the fact that he has lost his dog; he has lost, as he might put it (perhaps acknowledging that this isn't *quite* the right thing to say), "someone important to me," "someone" who was a significant part of his life.

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Are these states active or are they passive? It is not clear that the question is a good one, at least if it presupposes that they must be one or the other. The claim is not that there are *no* cases of regret, or of grief, or of other mental phenomena, including other emotions, which can helpfully be thought of as either active or passive. The point is rather that there may well be cases which are not helpfully described in *either* way. Importantly, the fact that such cases of regret and grief are responses to reasons––as they clearly are––does not obviously force us to view their subjects as their *agents*.

Positively, it may be worth taking seriously the idea of mental states and phenomena which embody a form of rational subjectivity which is neither essentially passive nor active. Mental phenomena which form part of one's rational subjectivity will be those mental states and attitudes which contribute to one's rational point of view or perspective on the world. The regret and grief described above are clearly elements of their subjects' viewpoints or perspectives in this way. Attitudes, emotions, and thoughts involving this form of rational subjectivity are ones which subjects wholeheartedly identify with, or inhabit, and are not at all alienated from. These are mental states and attitudes that are subject to reasons-seeking "why" questions ("Why do you regret *that*?"; "Why feel so bad about *a dog*?").[32] Yet the subjects in our examples are not properly thought of as having "made up their minds" to feel or to foster their regret or their grief (although it might be possible to do this in other cases).

This kind of picture looks like it may be able to contribute to an accurate characterisation of the distinctive features of our mentality we have seen agentialists try to explain. According to this alternative, one can be seen as responsible for attitudes and thoughts which form part of one's rational viewpoint or perspective, at least insofar as one recognises the applicability of reasons-seeking "why" questions in relation to them. In recognising such attitudes and thoughts for what they are, one's self-knowledge of them is bound up essentially with one's understanding of how they relate to these reasons, and with their status as part of one's point of view on the world (both cognitive and affective). And because such attitudes and thoughts are part of what constitutes this first-personal perspective, they will be experienced as "one's own" in a way that is distinctive, by contrast not only with alien-seeming mental states and events, but also with sensations which one may experience as responsive to purely causal, a-rational, factors.

We have not been able to fully describe or defend this category of rational subjectivity. What we have been starting to try to make plausible is that our ordinary ways of thinking do acknowledge psychological states and events which fall into such a category. If this is right, then recognising that the purely spectatorial and passivistic conception of the mind fails to do justice to the subject's rational involvement with aspects of her own mentality does not entail viewing those aspects as in all cases active.

What we have been suggesting, albeit briefly, is that if the Agentialist Dilemma cannot be avoided, then all is not lost. For those who, like us, agree with the Agential Turn's critique of the passivistic and spectatorial picture of human mentality, but wonder whether the notion of mental agency can carry the full burden of rejecting it, recognising a broader domain of rational subjectivity may well provide the resources to shoulder some of this weight.

DATA AVAILABILITY STATEMENT

Data sharing not applicable – the article describes entirely theoretical research.

ORCID

Alexander Greenberg https://orcid.org/0000-0002-2501-878X

ENDNOTES

[1] For discussion and feedback on earlier drafts, we thank Rachel Achs, Sophie Archer, Anita Avramides, Tom Crowther, Sandy Diehl, Naomi Eilan, Anil Gomes, Sophie Keeling, Hemdat Lerman, Conor McHugh, Oded Na'aman, Ram Neta, David Owens, Eylem Özaltun, Johannes Roessler, Antonia Peacocke, Matt Soteriou, Barney Walker, Jonathan Way, Daniel Whiting, Simon Wimmer, and anonymous reviewers for this and other journals. We also thank audiences in

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Oxford, Warwick, Tel Aviv, and Dortmund. This work was supported by two Early Career Fellowships funded by the Leverhulme Trust (grant numbers: ECF-2018-075 (Campbell); ECF-2019- 406 (Greenberg)).

[2] While these are not the only aspects of the mind mental agency has been claimed to explain, they are among the most common discussed in the agentialist literature. For example, Antonia Peacocke, in a recent survey article on mental action, cites them in a section on "what mental action explains" (2021, sec. 5).

[3] This is also true of other attitudes––for example, intentions, desires, emotions, etc.––but our focus on belief follows the agentialist literature we are discussing.

[4] Though sometimes there are extenuating circumstances: "She was brought up in a cult"; "He's experiencing a psychotic episode."

[5] Although we assume for the purposes of discussion that "responsibility" is the right concept to capture these features of belief, this starting point may be questioned. The authors part company here: Campbell remains agnostic about the usefulness of the notion of "doxastic responsibility"; Greenberg has provided his own (non-agentialist) account of it (2020).

[6] Expansionists are so-called because they suggest that the category of agency is more expansive than some people (mistakenly, in their eyes) regard it as being.

[7] Some agentialists draw on both strategies, for example, Miriam McCormick's account of doxastic responsibility has elements of assimilationism (2018, pp. 635–43) and expansionism (2011, pp. 172–76).

[8] Although this is not always recognised, voluntariness per se does not entail agency (Anscombe, 2000, 89–90; Hyman, 2015, 88–91).

[9] Much of the doxastic voluntarism literature is structured around responding to William Alston (1988), who treated epistemic norms and doxastic responsibility as a package––the "deontological conception of justification"––which he claimed implies doxastic voluntarism.

[10] Of course, one can, as Alston accepts (1988, p. 275), decide to *do* things to influence one's beliefs, for example, gather evidence or make an appointment with the hypnotist. But these are not *mental* actions in the way that believing at will would be, and so aren't relevant to our discussion.

[11] Someone else who appears to be an assimilationist about doxastic responsibility is Ernest Sosa, who claims that judgement is an intentional action, a claim he suggests explains responsibility for belief and judgement. Specifically, he claims judging that *p* is intentionally affirming *p* in the endeavor to affirm aptly (Sosa, 2015, p. 66). However, to our knowledge Sosa does not discuss the challenge of avoiding doxastic voluntarism in any detail, so we do not focus on him. The closest he comes is when he discusses the degree of freedom we have over our judgements: "we enjoy broad freedom both for choice and for judgment. We need only recall how often we freely conclude deliberation or pondering. We then decide whether to accept that the balance of reasons sufficiently favors either side over the other. Very often, on issues both weighty and trivial, it is up to the agent which way to turn" (Sosa, 2015, p. 208). This seems to us to be more or less equivalent to Steup's assimilationism.

[12] Weatherson himself may be happy to accept this, as he says he is "not going to die in a ditch over just what we call voluntary" (2008, p. 554).

[13] For versions of this view see: (Boyle, 2009a, pp. 123–124, 135–136; Hieronymi, 2008; McCormick, 2011; McDowell, 1998, p. 434; McHugh, 2013; Pettit & Smith, 1996; Toribio, 2011). Additional elements, which do not affect our arguments, are sometimes added to the basic reasons-responsiveness model. Some add an "ownership condition" over the reasons-responsive mechanisms (McCormick, 2011, pp. 173–75; McHugh, 2013, pp. 142–143); others add the idea that beliefs are subject to one's skill at noticing "situations that demand the deployment of [one's] rational abilities" (Toribio, 2011, p. 361). Note the similarities between expansionism and Steup's (assimilationist) view. We categorise Steup as an assimilationist because he explicitly claims to be defending doxastic voluntarism, but depending on how we divide up the views, we could categorise him as an expansionist, as his view has a similar structure to those we regard as expansionists, and suffers from the same problem.

[14] Kieran Setiya has appealed to the fact that sceptics can accept belief's reasons-responsiveness in his argument for doxastic agency scepticism (Setiya, 2013, pp. 181–82; for a response, see Neta, 2019a). Our argument differs from Setiya's because he is arguing for scepticism about doxastic agency as such, whereas our challenge is about its explanatory potential.

[15] Conor McHugh levels a related but slightly different charge at the anti-agentialist who "insist[s] on reserving the term 'action' for the intentional or voluntary" (McHugh, 2013, p. 149). In response to *this* point, we agree that not all 'action' is intentional or voluntary. Recall that this is in part why the clarificatory challenge arises for agentialism about doxastic responsibility.

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[16] Insofar as agentialists seek to explain non-doxastic self-knowledge, the focus is restricted to other broadly rational attitudes (intention, desire, etc.) about which we can also be said to 'make up our minds'. Self-knowledge of phenomenal states must be given a different account. See (Boyle, 2009b) for a defense of this approach.

[17] Agentialists about doxastic self-knowledge also commonly seek to explain the transparency of belief, the fact that the self-directed question *whether one believes that p* can be answered by simply considering *whether p*. Both Antonia Peacocke and Matthew Boyle, whose accounts we discuss in detail, offer accounts of transparency. To keep things focused, we abstract away from this (although see footnote 21).

[18] The argument of this section substantially develops a line of thought in (L. Campbell, 2018, pp. 338–339).

[19] With some accounts it's not clear on which side of the assimilationist-expansionist divide they fall. Matthew Soteriou provides a complex and nuanced account of certain kinds of doxastic self-knowledge as grounded in the role of judging in certain conscious activities such as calculating and suppositional reasoning (Soteriou, 2013, Chapters 10–15). Whether this is assimilationism or expansionism depends on whether Soteriou holds that such activities (insofar as they explain doxastic self-knowledge) are necessarily exercises of *intentional* agency. If he does view them in this way then his account has much in common with Antonia Peacocke's, which we consider below. We leave Soteriou's own account to one side not only for this reason, but also because he does not seem to be offering a *general account* of doxastic self-knowledge in terms of mental agency, but rather investigating how best to think of the nature of one's self-knowledge in those specific cases in which one *is* plausibly engaging in mental action.

[20] Other versions of assimilationism about self-knowledge focus less on how to avoid our Problem of Strength (see, for example, C. Peacocke, 2009).

[21] Because of this, and because we abstract away from her discussion of the transparency of belief (see footnote 17), our presentation of Peacocke's account is necessarily simplified. We do not think that this affects our argument.

[22] We depart a little from Peacocke's presentation here, in order to keep our discussion self-contained.

[23] For other differently developed expansionist accounts of doxastic self-knowledge, see: (Burge, 1996; O'Brien, 2007, pp. 88–94, 114–124). Brie Gertler gives a helpful critical survey of expansionist accounts of doxastic self-knowledge (2022). In other work, Gertler defends a position which grants the importance of the idea that belief might in some sense be 'active', but denies that this activity is what explains doxastic self-knowledge (2016).

[24] In conversation it has been suggested to us that Boyle should be interpreted as holding only the weaker agentialist conception of belief (which would also imply that *what a person knows* in doxastic self-knowledge is something "active"). It is possible that Boyle has moved away from an agentialist account in more recent work on self-knowledge, whose focus is nonpositional self-consciousness (2019). If so then this fact is not germane to our interests in this section, which do not concern Boyle ad hominem, but a version of agentialism about doxastic self-knowledge. We have chosen to engage with Boyle's work because he is standardly viewed as a key proponent of the stronger position, which he does defend and develop in his earlier work.

[25] This notion is also central to Boyle's (2011) explanation of transparency of belief. See again footnote 17.

[26] See, for example, Boyle's discussion of the essential self-consciousness of 'making up one's mind' (2009b, p. 154).

[27] Plausibly, nothing short of a conceptual truth will do here, given the opacity of explanation contexts.

[28] One might object even to this first standard move. First, "the sense of ownership" suggests a single identifiable qualitative feature accompanying all non-alien-seeming conscious thought. But the fact that there is such a thing as one's thoughts' "seeming alien" does not entail that there is such a thing as their "seeming mine," at least if this "seeming" is understood as a distinctive positive phenomenology (see Parrott, 2017, p. 50). Secondly, one might question the coherence of being introspectively aware of an episode of thinking without being aware that it is *oneself* who is doing the thinking, and reject the project of aiming to give a coherent interpretation of thought insertion testimony (see Roessler, 2013), as those who posit a sense of ownership aim to do. We set these issues aside in our discussion.

[29] Underlying the agentialist explanation of thought insertion is, of course, a further agential*ism*: agentialism about the sense of thought-ownership. However, the literature focusses on the explanatory role of this agentialism in understanding thought insertion, and we follow that lead here.

[30] For versions of this point, see: (Bortolotti & Broome, 2009, pp. 219–220; Fernández, 2010, p. 76; Parrott, 2017, pp. 42–43, 46–47; Roessler, 2013, p. 665).

[31] In one sufferer's testimony, "OCD" is represented as an "it" which "throws things" at the subject, "is at war" and "argues" with them. This might be taken to suggest a lack a sense of ownership over the sufferer's obsessive thoughts. Yet in the same testimony––almost in the same breath––"OCD" is identified with "irrational me," and when the subject acknowledges their obsessive thoughts as irrational, they are represented as "my thoughts." For the full testimony, see

(https://theocdstories.com/stories/in-my-mind-there-was-a-war/). Thanks to Sophie Keeling for directing us to this enlightening collection of first-person testimonies from OCD sufferers – and for helpful discussion of this topic.

[32] Cf. Ram Neta's discussion of what he calls "rationally determinable conditions" a category which includes actions, but also beliefs, emotions, and preferences (Neta, 2019b, p. 182). Neta highlights how these all are subject to reasons-seeking why questions (181), and that we enjoy an epistemic privilege with respect to the reasons on which such conditions are based (183–4).

REFERENCES

Alston, W. P. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives*, *2*, 257–299.

Anscombe, G. E. M. (2000). *Intention*. Cambridge, MA: Harvard University Press.

Bortolotti, L., & Broome, M. (2009). A role for ownership and authorship in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences*, *8*(2), 205–224.

Boyle, M. (2009a). Active belief. *Canadian Journal of Philosophy Supplementary Volume 35: Belief and Agency*, *35*, 119–147.

Boyle, M. (2009b). Two kinds of self-knowledge. *Philosophy and Phenomenological Research*, *78*(1), 133–164.

Boyle, M. (2010). Making up your mind' and the activity of reason. *Philosophers' Imprint*, *11*(17), 1–24.

Boyle, M. (2011). Transparent self-knowledge. *Aristotelian Society Supplementary*, *85*(1), 223–241.

Boyle, M. (2019). Transparency and reflection. *Canadian Journal of Philosophy*, *49*(7), 1012–1039.

Buckareff, A. A. (2004). Acceptance and deciding to believe. *Journal of Philosophical Research*, *29*, 173–190.

Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society*, *96*, 91–116.

Campbell, J. (1999). Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*, *82*(4), 609–625.

Campbell, L. (2018). Self-knowledge, belief, ability (and agency?). *Philosophical Explorations*, *21*(3), 333–349.

Engel, P. (2009). Epistemic responsibility without epistemic agency. *Philosophical Explorations*, *12*(2), 205–219.

Evans, G. (1982). In J. McDowell (Ed.), *The varieties of reference*. Oxford: Clarendon Press.

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, *4*(4), 636–640.

Fernández, J. (2010). Thought insertion and self-knowledge. *Mind & Language*, *25*(1), 66–88.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum Associates.

Gertler, B. (2016). Self-Knowledge and Rational Agency: A Defense of Empiricism. *Philosophy and Phenomenological Research*, *96*(1), 91–109.

Gertler, B. (2022). Agency and self-knowledge. In L. Ferrero (Ed.), *The Routledge handbook of philosophy of agency* (pp. 253–262). Abingdon and New York: Routledge.

Ginet, C. (2001). Deciding to believe. In M. Steup (Ed.), *Knowledge, truth, and duty: Essays on epistemic justification, responsibility, and virtue* (pp. 63–76). Oxford: Oxford University Press.

Greenberg, A. (2020). Epistemic responsibility and criminal negligence. *Criminal Law and Philosophy*, *14*(1), 91–111.

Hieronymi, P. (2008). Responsibility for believing. *Synthese*, *161*(3), 357–373.

Hyman, J. (2015). *Action, knowledge, and will*. Oxford: Oxford University Press.

James, W. (1956). *The will to believe and other essays in popular philosophy*. New York, NY: Dover Publications.

McCormick, M. S. (2011). Taking control of belief. *Philosophical Explorations*, *14*(2), 169–183.

McCormick, M. S. (2018). Responding to skepticism about doxastic agency. *Erkenntnis*, *83*(4), 627–645.

McDowell, J. (1998). Lecture I: Sellars on perceptual experience. *The Journal of Philosophy*, *95*(9), 431–450.

McHugh, C. (2013). Epistemic responsibility and doxastic agency. *Philosophical Issues*, *23*(1), 132–157.

Mellor, C. S. (1970). First rank symptoms of schizophrenia. *British Journal of Psychiatry*, *117*(536), 15–23.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, NJ: Princeton University Press.

Moran, R. (2003). Responses to O'Brien and shoemaker. *European Journal of Philosophy*, *11*(3), 402–419.

Neta, R. (2019a). Agency and the basing relation. In J. A. Carter & P. Bondy (Eds.), *Well-founded belief: New essays on the epistemic basing relation* (pp. 190–200). Abingdon and New York: Routledge.

Neta, R. (2019b). The basing relation. *The Philosophical Review*, *128*(2), 179–217.

O'Brien, L. (2007). *Self-knowing agents*. Oxford: Clarendon Press.

Owens, D. (2000). *Reason without freedom: The problem of epistemic normativity*. London and New York: Routledge.

Parrott, M. (2017). Subjective misidentification and thought insertion. *Mind & Language*, *32*(1), 39–64.

Peacocke, A. (2017). Embedded mental action in self-attribution of belief. *Philosophical Studies*, *174*(2), 353–377.

Peacocke, A. (2021). Mental action. *Philosophy Compass*, *16*(6), 1–15.

Peacocke, C. (2009). Mental action and self-awareness (I). In B. P. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 358–376). Chichester: Wiley.

Pedrini, P. (2015). Rescuing the "loss-of-agency" account of thought insertion. *Philosophy, Psychiatry, and Psychology*, *22*(3), 221–233.

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Pettit, P., & Smith, M. (1996). Freedom in belief and desire. *Journal of Philosophy*, *93*(9), 429–449.
Pickard, H. (2010). Schizophrenia and the epistemology of self-knowledge. *European Journal of Analytic Philosophy*, *6*(1), 55–74.
Roessler, J. (2013). Thought insertion, self-awareness, and rationality. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 658–672). Oxford: Oxford University Press.
Setiya, K. (2013). Epistemic agency: Some doubts. *Philosophical Issues*, *23*(1), 179–198.
Setiya, K. (2017). *Midlife: A philosophical guide*. Princeton, NJ: Princeton University Press.
Sosa, E. (2015). *Judgment and agency*. Oxford: Oxford University Press.
Soteriou, M. (2013). *The Mind's construction: The ontology of mind and mental action*. Oxford: Oxford University Press.
Stephens, G. L., & Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. Cambridge, MA: The MIT Press.
Steup, M. (2000). Doxastic voluntarism and epistemic deontology. *Acta Analytica*, *15*(24), 25–56.
Toribio, J. (2011). What we do when we judge. *Dialectica*, *65*(3), 345–367.
Weatherson, B. (2008). Deontology and Descartes's demon. *Journal of Philosophy*, *105*(9), 540–569.

How to cite this article: Campbell, L., & Greenberg, A. (2023). Mental agency and rational subjectivity. *European Journal of Philosophy*, 1–22. https://doi.org/10.1111/ejop.12867

14680378, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ejop.12867 by Test, Wiley Online Library on [20/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License