BLAMEWORTHINESS, CONTROL, AND CONSCIOUSNESS *OR A CONSCIOUSNESS REQUIREMENT AND AN ARGUMENT FOR IT*

BY

MICHAEL HATCHER

Abstract: I first clarify the idea that blameworthiness requires consciousness as the view that one can be blameworthy only for what is a response to a reason of which one is conscious. Next I develop the following argument: blameworthiness requires exercising control in a way distinctive of persons and doing this, in view of what it is to be a person, requires responding to a reason of which one is conscious. Then I defend this argument from an objection inspired by Arpaly and Schroeder according to which responding to moral reasons suffices for exercising control distinctive of persons.

1. Introduction

Early on 24 May 1987, Ken Parks drove to his mother-in-law's home and stabbed her. He later arrived at a police station with cuts on his hands. During his trial, it became clear he had been sleepwalking. He was acquitted of murder for the reason that 'sleepwalking is a state of automatism in which an individual is unaware of, and has no control of, his or her behavior' (Broughton *et al.* 1994, p. 254). His acquittal seems based on the thought

Pacific Philosophical Quarterly •• (2022) ••–•• DOI: 10.1111/papq.12405

that he is guilty only if blameworthy and that he is not blameworthy because he was not conscious.[1]

Call any view according to which blameworthiness requires consciousness *a consciousness requirement*. Such requirements can take many forms, but the case of Parks illustrates the intuitiveness of the basic idea. Before learning he was sleepwalking, it is easy to blame Parks. Afterwards, it is not so easy. And Shepherd (2015) has argued intuitions favoring consciousness requirements are robust among ordinary folk. One survey he gave had a story in which one person punches another, but the person struck is in the puncher's blind hemifield. People were less likely to blame as compared with a normal case (Ibid., pp. 933–936).

But a number of philosophers reject consciousness requirements. Nomy Arpaly (2003, pp. 159–160) points to people unconsciously motivated by morally bad reasons. And Angela Smith (2005), George Sher (2009), and Randolph Clarke (2014) argue cases of forgetting, as when one leaves the family dog in a hot car, discredit consciousness requirements. Thus, intuitions favoring consciousness requirements are not enough by themselves. Argument is needed.

I have two aims in this paper. My first aim is to make precise a consciousness requirement worth arguing for. The requirement I pinpoint is that one can be blameworthy only for what is, to use language I shall explain, a response to a reason of which one is conscious. I give a counterexample to the consciousness requirement advocated by Neil Levy (2014) and then show my requirement, but not his, has an important resource for explaining intuitions of blameworthiness in the forgetting cases to which opponents of consciousness requirements appeal. I also explain my requirement's implications for when one is blameworthy for behavior motivated by implicit bias and suggest these implications mirror the complexity of our intuitions about such cases.

My second aim is to argue for my consciousness requirement. Notice the rationale for Parks' acquittal connects one's being 'unaware of' one's behavior with one's having 'no *control* of' it. Call any view according to which blameworthiness requires control *a control requirement*. Control requirements can take many forms, but the basic idea is intuitive, and pervasive.[2] Underlying the rationale for Parks' acquittal, then, is the idea that a consciousness requirement flows from a control requirement. This idea turns out to be a good one if the argument I develop is sound, as it has the feature of beginning with a control requirement.

[1]In this paper, *blameworthiness* refers to one's being an appropriate target of resentment, indignation, and guilt. For precedence see Strawson (1962), Wallace (1994), Rosen (2004, p. 297), and Fischer and Ravizza (1998, p. 7).

[2]Fischer and Ravizza write: 'It seems to be a basic presupposition embedded in the way we think about these matters that an agent must in some sense control his behavior in order to be morally responsible' (Fischer and Ravizza 1998, pp. 13–14).

So far, the most developed argument with this feature is Levy's (Ibid.), and it proceeds via claims within cognitive science: that domain-general integration is required for the flexible responsiveness constitutive of *guidance control* (*a la* Fischer and Ravizza 1998), and that consciousness alone allows for such integration.[3] However, Sripada (2015) argues the first of these claims gets the science wrong; Mudrik *et al*. (2011) the second.

My argument, though, has a more conceptual flavor and, for better or worse,[4] it does not depend on the empirical questions Levy's depends on.[5] My argument, in outline, is that blameworthiness requires exercising control in a way distinctive of persons and that doing this, in view of what it is to be a person, requires responding to a reason of which one is conscious.

After developing this argument, I consider an objection inspired by Arpaly (2003) and Arpaly and Schroeder (2014) according to which, to exercise control distinctive of persons, it is enough that one responds to a *moral* reason, even if one represents it unconsciously. For, so the objection goes, only persons can gain the concepts needed to represent moral reasons in *any* fashion, whether consciously or not. I will argue, though, that the only way to sustain this objection requires one to place an *ad hoc* constraint on moral theory.

Here is the plan. Section 2 makes precise my consciousness requirement. Section 3 argues for it. Section 4 replies to the objection inspired by Arpaly and Schroeder. Section 5 explains how my argument does not depend on the empirical claims Levy's depends on and concludes.

2. *A consciousness requirement*

A case from Levy can teach us a number of things about the form a consciousness requirement should take:

> An agent may be morally blameworthy for what he takes to be a theft from an impoverished widow, even if it turns out that the woman is fabulously wealthy and the bag he sneaks off with had in fact been discarded by her. (Levy 2014, p. 36)

[3]Besides Levy's, it is not easy to find a *developed* argument from a control requirement to a consciousness requirement. Sher (2009, Ch. 4) *sketches* one, and argues it fails.

[4]There is disagreement about whether, in philosophy, independence from empirical questions is a virtue or a vice. See Knobe and Nichols 2017, §3. Addressing this disagreement is outside the scope of this paper.

[5]This is not to say my argument is independent from *all* empirical questions, though I am open to the possibility that it is. For I am open to the possibility that it draws only conceptual connections and that conceptual connections are non-empirical. However, it is controversial what it takes for a connection to be conceptual, and controversial whether conceptual connections are thereby non-empirical (see Jackson 1998, Ch. 2). These controversies being beyond the scope of this paper, I will make no attempt to show my argument is independent of empirical questions beyond those Sripada and Mudrik *et al*. raise; for this, see §5.

Call this agent 'Zac'. Levy is right, I take it, that Zac could well be blameworthy.

Given this, Zimmerman (1997, p. 421) and Rosen (2004, p. 307) are wrong to think blameworthiness requires one consciously believe the action one is performing is wrong. Zac could easily lack this belief and still be blameworthy. Consciously believing the bag belongs to a poor woman is plenty.[6]

Via this belief, Zac is conscious of what I shall call a *reason*. My consciousness requirement requires consciousness of a reason, but only in one of the senses of the word.

Normative reasons are facts bearing on what one ought to do.[7] We can imagine Zac not conscious of any normative reasons against taking the bag quite consistently with his being blameworthy, though. There are generally no normative reasons against using discarded bags. Using them only makes the world cleaner. Of course, Zac consciously believes *that the bag belongs to a poor woman*, but *this* is no normative reason against taking it, being a false proposition rather than a fact. If there *is* a normative reason against taking the bag, it could only be *the fact that Zac believes that the bag belongs to a poor woman*. But Zac could easily fail to introspect and become conscious of this fact about his mind.[8] Still, he could well be blameworthy.

Motivating reasons are those things on the basis of which, or for which, one acts.[9] Metaphysically, some hold these things to be *mental states* (e.g., Wedgwood 2006), others *facts* (e.g., Broome 2013), and others *propositions* (e.g., Schroeder 2008). While my requirement and argument for it are formulable without loss in any of these terms,[10] I will, to fix ideas, write as if motivating reasons are propositions. So, to take Parfit's (1997, p. 99) example, if one jumps out the third story window, correctly thinking the building on fire but incorrectly thinking the river below has already thawed, I will

[6]Clarke agrees one could be blameworthy without having 'any belief about the rightness or wrongness of what she does' (Clarke 2017, p. 233); see also FitzPatrick (2017, p. 41). In a similar vein, Arpaly (2003, Ch. 2) argues convincingly that one can be blameworthy even for acting in accord with one's best judgment.

[7]See Parfit (1997, p. 99) for both this kind of reason and the next. I thank an anonymous referee for requesting further clarity in my use of 'reasons'.

[8]See Robichaud (2015, p. 29) for precedence. See also Hatcher (2018).

[9]A good question is what it *is* for one to have acted on the basis of something. Plausibly, it is for certain '[f]acts about causation' to obtain, as Arpaly and Schroeder put it (2015, p. 103). Making those facts precise is no easy task, though. For Arpaly and Schroeder, as for Wedgwood (2006), it is for causation by normativity to obtain – namely, for one's act to be non-deviantly caused by features which make the act have something going for it, normatively speaking, to at least some extent. I will take no stance on this issue, though.

[10]Consider 'one represents a reason' and 'one is conscious of a reason'. To accommodate the mental state view, replace these with things like 'one's representation of such-and-such a content is a reason' and 'one's consciousness of such-and-such a content is a reason'. And to accommodate the fact view, replace them with things like 'the fact that one represents such-and-such a content is a reason' and 'the fact that one is conscious of such-and-such a content is a reason'.

say the propositions *that the building is on fire* and *that the river has thawed* are one's motivating reasons, though, of course, the latter of these is false.

Levy holds that blameworthiness requires consciousness of a motivating reason. According to him, one must be conscious of a content in the sense that one 'is able to effortlessly and easily retrieve it for use in reasoning *and* it is online' (Levy 2014, p. 33, italics his). A content is 'online' when it 'actually guides an agent's behavior' (Ibid., p. 32).[11] This is equivalent to saying the content is a motivating reason.

But blameworthiness does not require consciousness of a motivating reason any more than it requires consciousness of a normative reason. The proposition *that the bag belongs to a poor woman* could, in principle, motivate Zac, if he has a cruel streak, say. But suppose it does not motivate him and instead makes it more difficult for him to take the bag. Suppose, that is, that if it *were* to move him, it would move him to *not* take it. And suppose, further, that Zac is not conscious of *any* of the contents on whose basis he acts. Perhaps, for example, he is entirely motivated by a Freudian desire to impress his recently deceased father who held bold action in high regard. Even so, Zac could well be blameworthy, given his conscious belief that the bag belongs to a poor woman. Levy is wrong, then, to require consciousness of a motivating reason. Zac is a counterexample.[12]

That the bag belongs to a poor woman is not a normative reason, being false, and not a motivating reason, as Zac does nothing on its basis. But it

[11]As an anonymous referee points out, holding that consciousness of a content requires its actually motivating one's behavior, as Levy does, has odd consequences. For example, it implies, implausibly, that one is *not* conscious of the orchid one sees and explicitly notices, so long as one is not doing anything on the basis of an orchid-related content. It is unclear how deep this criticism of Levy is, though. He confesses his terminology is 'stipulative' (Ibid., p. 33, footnote 9). If consciousness in his sense *is* required for blameworthiness, we could perhaps restate his view as requiring not just consciousness but consciousness whose content motivates us.

[12]As an anonymous referee helpfully puts the point, Levy's view cannot capture *culpable indifference with awareness*, i.e., cases where one is blameworthy for not being motivated by something one is conscious of which should have motivated one. In this connection, here is all Levy says as to why the reason one is conscious of must be motivating ('online'):

> It is insufficient that the agent be able effortlessly to retrieve it, because sometimes we may become aware of the contents of our attitudes *by* retrieving them: we see what we are disposed to say about some topic and thereby discover what we think. If the information so retrieved is online only as a consequence of retrieval, it does not count as personally available *until* it is retrieved. We have already seen, on the other hand, that being online is insufficient for personal availability. It is the conjunction of effortless and easy retrievability and being online that is needed. (Levy 2014, p. 33)

Levy seems to be highlighting that in some cases, a content is effortlessly and easily retrievable in a 'deviant' way – for example, only because we can prompt ourselves to 'see what we are disposed to say about some topic' – whereas, in other cases, a content is both effortlessly and easily retrievable as well as online. But this is a reason to consign appropriate blame to the latter kind of case only if *the only way* for a content to be *non-deviantly* effortlessly and easily retrievable is for it to be online, something, so far as I can tell, Levy gives us no reason to believe.

is a reason: before he takes the bag, it is a consideration on the basis of which he *could*, in a certain sense of 'could', refrain from doing so. And it is not alone in being a consideration of this kind. *That it would impress his late father*, *that he is hungry*, and so on, are also among the considerations which could motivate him in this sense, one way or the other. Before he takes the bag, none of these are *actually* motivating reasons, not yet at least. But each is *potentially* a motivating reason, which is why I call reasons so understood *potentially motivating reasons*.[13]

The sense in which *that the bag belongs to a poor woman* 'could' motivate Zac to refrain, and by which it qualifies as a potentially motivating reason to do so, is that it is *presently available* to motivate him in the following sense: there is a possible world in which his mind up until the relevant time is the same as it actually was, in a certain sense of 'same', and it *does* motivate him to refrain at that time. The sense of 'same' I have in mind need not involve sameness down to microphysical detail. Sameness at the 'folk psychological' level of description, which abstracts away from such detail, is plenty – the notion, thus, is entirely friendly to determinism.[14] To say a content is presently available to motivate one, put differently, is to say no alteration to one's mind, in the sense at issue here, is required for it to be right there for one to do something with, for one's faculties, capacities, or mechanisms to operate on it. We can contrast here both *that the bag belongs to the poor woman* and *that it would impress his late father* with *that the woman discarded the bag*. For the last to be right there for Zac to do something with, first an alteration to his mind is required (e.g., his coming to believe it). As it is, only the first two of the three propositions are presently available to motivate Zac.

My consciousness requirement requires consciousness of a *potentially motivating reason* – to which 'reason' henceforth refers.[15]

Now, suppose Zac's taking the bag, entirely unbeknownst to him, signals a sniper to assassinate the mayor. I have maintained Zac could well be blameworthy for taking the bag. A consciousness requirement, though, should imply he is *not* blameworthy for signaling the sniper. But on Davidson's (1963) *coarse-grained* approach to event individuation, Zac's taking the bag is *one and the same event* as his signaling the sniper. For this

[13]This kind of reason, however labelled, has been noticed before; see, for example, Schroeder (2007, p. 14), Setiya (2007, p. 12), and Williams (1995, p. 35).

[14]For precedence, see Wedgwood (2013) and Kenny (1976).

[15]Presumably, if blameworthiness requires consciousness of a potentially motivating reason, it requires consciousness of one which relates in *some* way to what one ought to do. I find tempting the view that it must be a proposition which, if true, would be a normative reason against what one does. But I do not find this view mandatory and I will not, in this paper, defend any particular account of how the reason consciousness of which is required must also relate to what one ought to do.

reason, in this paper, I will understand events to be *fine-grained*; in particular, following Kim (1966), I will, to fix ideas, take an event to be a thing's instantiation of a property at a time.[16] Zac's taking the bag is distinct from his signaling the sniper because the two properties instantiated, here, are distinct. This makes room for a consciousness requirement to allow that he is blameworthy for the former but not the later.

My requirement will do this by holding that what one is blameworthy for must be a *response* to a reason of which one is conscious. An event *e* is a *response* to a reason r, in the sense I have in mind, just in case *e* is done either at least partially *on the basis of* r or at least partially *in spite of* r. In the variant of the case in which Zac takes the bag out of cruelty to the woman, he instantiates the property of taking it on the basis of *that the bag belongs to a poor woman*. But as her supposed poverty actually gives him pause – that content is presently available to motivate him *to refrain*, in particular – he, as it is, instantiates that property in spite of that proposition. Either way, his taking the bag is a response to that reason, in my sense. But there is no reason he is conscious of on the basis of which, or in spite of which, he instantiated the property of signaling the sniper. This is why he could be blameworthy for taking the bag but not for signaling the sniper.[17]

Now, when Zac takes the bag in spite of the woman's apparent poverty, Arpaly would speak of 'a failure to respond' to a reason rather than a

[16]Compared with Sher (2009, pp. 4–5) and Clarke (2014, p. 158). I am friendly, I should say, to the *existence* of both kinds of events. If one wishes, one could use 'events' for coarse-grained events, 'aspects of events' for fine-grained events, and translate 'events' in the main text as 'aspects of events'.

[17]An anonymous referee gives the following case: I punch someone. I am conscious of none of the reasons on the basis of which I do so. But I am conscious of the blue sky. Am I thereby conscious of a reason *in spite of which* I punched the person?

It depends entirely on whether the blue sky could motivate me to refrain from punching them in the sense clarified in the main text. And this depends on the details. Suppose the blue sky could move me to think about blueness but not much else and that my thinking about blueness is quite consistent with punching them then, too. Then the blue sky could not motivate me to not punch them and so fails to be a reason in spite of which I do so. (I would likely be conscious of *other* reasons in spite of which I punch them, though – for example, that there is a person before me.)

Alternatively, suppose I am somehow aware I had best think about the features of random things when around others else Freudian impulses will result in my having punched them. Then the blue sky could motivate not just my thinking about blueness but *my thinking about blueness rather than punching the person* and, on this account, constitute a reason in spite of which I punch them.(Care is required, here, though. To alter the case, suppose my blind hemifield presents me with someone I am about to unconsciously punch, that the blue sky my normal hemifield presents me with could motivate me to think about blueness, and that, *though I have no clue this is the case*, my thinking about blueness would prevent my punching the person, given the psychological mechanisms involved. Suppose I punch. Though I do fail to think about blueness in spite of a reason of which I am conscious, this is not a case in which I punch in spite of a reason of which I am conscious, for the same reason that Zac's is not a case in which he signals the sniper in spite of a reason of which he is conscious, though he does take the bag in spite of a reason of which he is conscious.)

response to one, as I do (Arpaly 2002, p. 231). I like my terminology, though. Taking the bag is something Zac *did with* the proposition that the bag belonged to a poor woman in a sense of 'did' sufficient to render intelligible questions of blameworthiness.[18] Relatedly, some use 'response' as a *success* term, as referring to one's doing the *correct* thing to do with one's reasons.[19] The word is no success term for me: Zac's taking the bag is a response to his reasons even if it is not the correct thing to do with them.

I can now state my consciousness requirement somewhat precisely. Where x is a variable for anything, it is

CONSCIOUSNESS For all x, for all e, x is blameworthy for e only if, for some r, x is conscious of r at the time of e and e is a response x has to r.

Even before filling in that placeholder which is the word 'conscious', the structure of this requirement, by contrast with Levy's, gives us a crucial resource to explain intuitions of blameworthiness in the forgetting cases to which opponents of consciousness requirements appeal.

Consider Sher's (2009) case of Alessandra, who, when picking up her kids from school on a hot day, is 'greeted by a tangled tale of misbehavior, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out' during which the family dog 'languishes, forgotten, in the locked car' (p. 24). Sher submits that, intuitively, Alessandra is blameworthy for forgetting the dog.

In view of cases like this, the distinction between *indirect* blameworthiness, which exists by virtue of blameworthiness for past conduct, and *direct* blameworthiness, which does not, is often brought to bear.[20] Maybe Alessandra is only indirectly blameworthy for forgetting the dog, having failed in the past to develop, say, the virtue of presence of mind. But, while I accept the distinction between indirect and direct blameworthiness – and 'blameworthiness' refers to direct blameworthiness throughout this paper – one might think Alessandra blameworthy for *that day*'s failure, not merely for past failures to develop herself.

And here CONSCIOUSNESS shines. It *does* allow that Alessandra is (directly) blameworthy for something which occurred that day. When she shuts the

[18]As Clarke points out, 'to do' is 'multipurpose' (Clarke 2014, p. 4). And consider Rachels' comments on whether a doctor 'does' something when he lets a patient die:

> … for any purpose of moral assessment, it is a type of action … If a doctor deliberately let a patient die who was suffering from a routinely curable illness, the doctor would certainly be to blame for what he had done, just as he would be to blame if he had needlessly killed the patient. (Rachels 2016, p. 251)

[19]I thank an anonymous referee for asking me to clarify this point.

[20]See, for example, Levy (2014, p. 3) and Fischer and Ravizza (1998, pp. 49–51).

car door, Alessandra is conscious *that the dog is in the car* and *that it is a hot day*. These are reasons to organize her attention in such a way that she ensures the dog is released soon. But either then or shortly thereafter, she omits to organize her attention in this way *in spite of* these reasons of which she is conscious. We can hold this omission is what she is to blame for.[21] Such explanations of blameworthiness in forgetting cases are CONSCIOUSNESS-friendly because 'a response *x* has to r' can refer to an event occurring *in spite of* r. But they are inconsistent with Levy's requirement, according to which one must be conscious of a reason which *motivates* what one is to blame for. So much the worse for that requirement compared with CONSCIOUSNESS.[22]

Let us now fill in what sort of representational state 'conscious' refers to in CONSCIOUSNESS.[23] It cannot be a state none of whose dispositions is manifest – like, for example, my standing belief, before I began writing this paragraph, that Ringo Star was a drummer for the Beatles. For the content of the state to which 'conscious' refers is a *reason*, which, by definition, must be presently available to motivate one. And a representational state with that feature must be *occurrent*, in the sense that at least one of its dispositions is

[21]In my framework, Alessandra's omission is her instantiating the purely negative property of *not organizing one's attention in the relevant way*. Clarke denies such properties exist:

> Must everything lacking spin-up – a spin-down electron, Operation Desert Storm, humility – be genuinely similar to all the others, and must they all share some common causal capacities? It seems to me that the applicability of the predicate 'doesn't have the property of spin-up' to them all requires neither of these things. It requires that they don't have the property of spin-up, not that there's some property all of them have. (Clarke 2014, p. 41)

Clarke is right that a spin-down electron and humility have no 'common causal capacities', though both lack spin-up. The same is true of Alessandra and the number 7, though both lack the property of organizing one's attention in the relevant way. But grounding common causal capacities is not essential to property-hood, any more than 2 and 4 fail to have the property of being even because they have no causal capacities whatsoever. Grounding or constituting genuine similarity is enough, and a spin-down electron and humility *are* genuinely similar with respect to *not being spin-up*, just as Alessandra and 7 *are* genuinely similar with respect to *not organizing one's attention in the relevant way*.

[22]CONSCIOUSNESS's structure is virtuous in other ways, too. Events of belief formation can be responses to reasons of which one is conscious; so, CONSCIOUSNESS allows one could be blameworthy for these in addition to (paradigmatic) actions or omissions. Relatedly, given that forming a belief in this way need not be to 'choose' to form it, CONSCIOUSNESS is not a version of what Sher dubs '*the searchlight view*', according to which one can be blameworthy only for what one consciously 'chooses' to do (Sher 2009, pp. 4–6). Moreover, CONSCIOUSNESS requires consciousness of *a* reason to which *e* is a response, not *all* of them, which is why it allows that Zac is blameworthy even though he is not conscious of *all* the reasons to which he is responding. And while we are at it, it is worth emphasizing CONSCIOUSNESS does not, as formulated, require consciousness of *e*, one's doing *e*, the moral significance of *e*, or that *e* is a response one has to r. As Arpaly (2003, p. 50) wisely advises, we should avoid 'level-confusions' here and not merely, as Alston (1980) warned us, in epistemology.

[23]For a state to be representational is for it to *represent* something, and to represent something is to be, in some sense, *of* or *about* it. For a survey of the notion, see Pitt 2017.

manifest. For its disposition for its content to be presently available must be manifest, at a minimum.[24]

Occurrent states can make their contents presently available in very different ways, though. The sleep-walking Parks occurrently represented where his car was on the road as he drove. I occurrently represent the sentences in this paragraph as I type. These occurrent representational states are very different. Mine is a state of consciousness while Parks' is not. And, of course, the two sorts of occurrent representation can run in parallel. Adapting a case from Arpaly (2003, p. 17), suppose George is conscious that the window near Rachel has a good view of the landscape and sits next to her rather than Jane, while what motivates him more than anything else is actually the content of his unconscious belief that Jane dislikes him.

As I shall understand 'conscious', to say George is conscious of the content about the window but not the content about Jane, though both are presently available to motivate him, is to say the former but not the latter is presently available to motivate him *in his reasoning*, in particular. Ask George about Jane, and *that Jane dislikes him* may *then* be at hand for his reasoning capacities to operate on ('Huh,' he muses to himself, 'perhaps I should sit *next* to her next time; she may grow to like me!'). But this is just to say an alteration to his mind is needed for that content to be right there for his reasoning capacities to operate on, which is to say that while it may be presently available to motivate him in some fashion, it is not (yet, at least) presently available to motivate him *in his reasoning*, in particular.

Of course, one *could* label as 'reasoning' George's deciding to sit next to Rachel because Jane dislikes him – just an automatic, subpersonal, or implicit kind.[25] I shall not use 'reasoning' in that way in this paper, though. Cognitive psychologists are wont to distinguish System 1 from System 2 mental events.[26] System 1 mental events are 'automatic', 'low effort', 'high capacity', 'independent of working memory', and 'shared with animals', while System 2 events are 'controlled', 'high effort', 'low capacity', 'limited by working memory capacity', and 'uniquely developed … in modern humans' (Evans 2008, pp. 257, 260). Recognizing a face is a paradigm example of a System 1 event, and holding back an insult, voting, and so on, are paradigm examples of System 2 events. George's being motivated by the content of his belief about Jane is a System 1 event, his being motivated by the content of his belief about the chair's nearness to the window is a (simultaneous!) System 2 event. The word 'reasoning', throughout this paper, refers to a System 2 event.

This is to be kept in mind when I say that to be conscious of a content is *for it to be presently available to motivate one in one's reasoning*. I take the notion

[24]For discussion of the occurrent/dispositional distinction, see Robert Audi (1994).

[25]I thank an anonymous referee for this point.

[26]Some might reject the idea that there are two systems, holding that the mental events at issue differ along a gradual spectrum. But I suspect the argument of this paper could be recast in terms of a gradual spectrum.

of consciousness I am using here to be broadly similar to, though not identical with, Levy's notion of a content being easily retrievable for use in reasoning.[27] I will take no stand on how consciousness, so understood, relates to phenomenal consciousness.[28]

Each part of CONSCIOUSNESS has now been explained. A good illustration of the view is its implications for when behavior motivated by implicit bias is blameworthy.[29] Consider Frankie. Remembering her coworker loves frappuccinos, out of kindness she surprises him with one. While doing this, though, unbeknownst to her, she exhibits subtle, off-putting signs of discomfort due to an implicit bias of hers which associates black males with violence (her coworker is a black male). Frankie is not conscious of the content of this bias and would be horrified to learn she had exhibited these signs of discomfort.

Whether CONSCIOUSNESS absolves Frankie depends on the details of the case. Frankie, in being conscious of her coworker and his interests, is conscious of reasons which can motivate her to treat him kindly. Let us suppose these reasons motivated her to refrain from *certain* signs of discomfort, D1, toward which she felt quite inexplicably inclined – no one likes another to be uncomfortable around them, after all – but that she had no inkling whatsoever that, nevertheless, she still showed *other* signs of discomfort, D2. Given that her self-knowledge extended to D1 but not D2, the reasons she was conscious of were presently available to motivate her in her reasoning *to not show D1* – but *not* to not show D2. This is why, had she shown D1, she would have done so *in spite of* the reasons to treat her coworker kindly of which she was conscious, while, as it is, her showing D2 is *not* something she did in spite of reasons of which she was conscious. Her showing D2, then, is not a *response* to any reason of which she is conscious at the time, and CONSCIOUSNESS absolves her.[30] She fails to be conscious of a reason to which her showing D2 – the exact, fine-grained event we are thinking through whether she is blameworthy for – is a response.

[27]For Levy, a content is easily retrievable for use in reasoning when any of a 'large range of ordinary cues' would suffice for its being used in reasoning, no special interventions being needed (Levy 2014, p. 34). Now, so far as I can tell, a person *could* use a given content in their reasoning in the sense of 'could' constitutive of present availability though they in fact *wouldn't* in much of the range of circumstances Levy has in mind. It is possible, indeed, to be pretty consistent in reasoning a certain way *in spite of* a reason on whose basis one *could*, in the relevant sense, reason otherwise. Levy's notion and mine, then, appear to come apart. But there is deep similarity in the emphasis that no special interventions are necessary and in the focus on reasoning. Consciousness construed along these general lines has some similarity to what Block (1995) calls *access consciousness*; for this, see Levy (2014, pp. 35–36).

[28]Compare Levy (2014, pp. 28–29).

[29]I thank an anonymous referee for encouraging me to clarify CONSCIOUSNESS by considering its relationship to implicit bias. For argument that behavior motivated by implicit bias is blameworthy, see, for example, Holroyd 2012; for argument to the contrary, see, for example, Levy 2017b.

[30]At least, it absolves her from *direct* blameworthiness. CONSCIOUSNESS is consistent with holding Frankie is *indirectly* blameworthy, having omitted in the past to expose herself to the relevant group or resolve upon certain intentions which reduce the behavioral impact of implicit bias; for relevant discussion, see Holroyd (2012, pp. 286–291).

But if Frankie had the relevant self-knowledge and *still* showed D2, she could well be blameworthy, for all CONSCIOUSNESS says. For then the reasons to act kindly of which she was conscious would have been presently available to motivate her in her reasoning to not show D2. So, CONSCIOUSNESS's implications about when behavior motivated by implicit bias is blameworthy are complex and depend on the extent of the agent's self-knowledge. This complexity seems to mirror the complexity of our intuitions about such cases, in addition to explaining why we take self-knowledge to matter.[31]

3. *An argument for CONSCIOUSNESS*

Replace 'conscious of' in CONSCIOUSNESS with 'occurrently represents' and the result is the more general view that blameworthiness requires responding to a reason one occurrently represents. This view is the control requirement on whose basis I shall argue for CONSCIOUSNESS.

Now, nothing I have said precludes a *very* inclusive understanding of occurrently representing and responding to a reason. Just to illustrate, one is free to – though, of course, one need not – hold that when an automatic door opens for a pedestrian it detects, it opens on the basis of a reason it occurrently represents.[32] One might, then, wonder whether the capacity to occurrently represent and respond to a reason deserves to be called 'control'. I, myself, am fine with using the word in this way. Notice that a natural way to capture one of the contrasts between a door's opening for a pedestrian and its opening, say, because a car crashes through it is to say the door exercises control in the former case but not in the latter. And it is no accident that with respect to the former case, but not the latter, it is at least *intelligible* to imagine the door occurrently represents and responds to a reason.

What really matters, though, is whether the requirement I have in mind has potential to help convince us to accept CONSCIOUSNESS. If the capacity at issue is so lightweight even automatic doors have it, this makes requiring it, whatever we call it, only *more* plausible, after all. In any case, both to capture its lightweight nature and to give permission to read my phrasing as stipulative, I will call the capacity to occurrently represent and respond to a reason *minimal control*.

[31]See Kelly and Roedder (2008, pp. 532–535) and Holroyd (2015, pp. 518–520) for discussion of ways to learn about one's biases and how this bears on what adjustments to our behavior we can be held responsible to make.

[32]Asks an anonymous referee: could a door do something *in spite of* a reason? Maybe! It may have to be rather sophisticated, though. Suppose, for example, that the door opens just in case the cumulative weight of nearby pedestrians surpasses a threshold a randomizing mechanism reselects every thirty seconds. In principle, we could think of the door's not opening for a given pedestrian as something it does *in spite of* a reason, as the weight of the pedestrian could have motivated the door's opening without any alteration in the internal state of the door at the, as it were, 'folk psychological' level of description. The example of automatic doors is inspired by Goldman (1976, p. 791) and Searle (2004, p. 206).

My argument for CONSCIOUSNESS is as follows. Only exercises of minimal control can be blameworthy. But only persons can do something blameworthy. So, there must be a way of exercising minimal control which is both required for blameworthiness and distinctive of persons. But a way of exercising minimal control is distinctive of persons only if the occurrent representation of reasons built into it, in particular, is distinctive of persons. Furthermore, to be a person is to be a rational being. In light of this, occurrent representation of a reason is distinctive of persons only if the reason is presently available to motivate one in one's reasoning – which is to say, given the notion of consciousness we are working with, only if one is conscious of it. So, CONSCIOUSNESS.

More formally

(1) For all *x*, for all *e*, *x* is blameworthy for *e* only if *e* is an exercise of *x*'s minimal control, that is, only if, for some r, *x* occurrently represents r at the time of *e* and *e* is a response *x* has to r.[33]
(2) For all *x*, for all *e*, *x* is blameworthy for *e* only if *x* is a person at the time of *e*.
(3) If (1) and (2), then: For all *x*, for all *e*, *x* is blameworthy for *e* only if *e* is an exercise of *x*'s minimal control in a way distinctive of persons.
(4) For all *x*, for all *e*, *e* is an exercise of *x*'s minimal control in a way distinctive of persons only if, for some r, *x* occurrently represents r in a way distinctive of persons at the time of *e*.
(5) Persons are rational beings.
(6) If (5), then: For all *x*, for all *e*, for all r, *x* occurrently represents r in a way distinctive of persons at the time of *e* only if *x* is conscious of r at the time of *e*.

CONSCIOUSNESS For all *x*, for all *e*, *x* is blameworthy for *e* only if, for some r, *x* is conscious of r at the time of *e* and *e* is a response *x* has to r.

My argument for (1), my control requirement, begins with the intuition that one cannot be (directly) blameworthy simply for being who one is. One must *do* something.[34] Now, some contest this intuition. Adams (1985) and Graham (2017, pp. 167–169), for example, point to objectionable mental states – racism, malice, and so on – and suggest that insofar as these constitute who one is, one *is* blameworthy simply for being who one is. In this

[33]Some, for example, Mendelovici (2018) and Woodward (2015), argue only phenomenal consciousness can be occurrently representational. If they are right, and if phenomenal consciousness entails consciousness in the sense at issue in this paper, nothing more would be needed to connect (1) to CONSCIOUSNESS. Cognitive scientists and philosophers of mind, however, generally assume there are occurrent representational states beyond varieties of consciousness (van Gulick 2018, §4.7). And discussions of consciousness requirements generally play out against the backdrop of this assumption (see Levy 2008, p. 217). My argument is no exception in this regard.

[34]*A la* endnote 18, 'do' includes omissions.

vein is the view that what one is blameworthy for, most fundamentally, is improper intrinsic desire (e.g., malice) or lack of proper intrinsic desire (e.g., indifference).[35] But now see the evil grin of the 'demonic neuroscientist' as he finishes implanting malice or indifference into an unwitting victim, and it is once again intuitively compelling that while one may be blameworthy for what one did, or omitted to do, to become who one is, simply being a certain way cannot *in itself* constitute a ground for appropriate blame.[36] As Clarke, no friend of consciousness requirements, puts it, while an action or omission is assessable for appropriate 'resentment, indignation, and guilt', 'having nonvoluntary attitudes is not' (Clarke 2014, p. 113).

Of course, that one must in some sense *do* something does not, in itself, imply one must exercise minimal control. Perhaps *manifesting* an objectionable mental state is blameworthy and also possible without exercising minimal control. Smith (2005, 2008), for example, suggests forgetting a friend's birthday, which is surely not a response to a reason one occurrently represents, can manifest inadequate concern for them. One way to phrase the suggestion is that what we pay attention to reflects our concerns, and, in some cases, one forgets a friend's birthday because one has inadequate concern for them.[37]

But I see only two ways to forget something due to inadequate concern. On the first, at some point one occurrently represents a reason to organize one's attention so as to promote one's not forgetting something, but, in spite of this reason, one omits to do this. One with adequate concern may well have been moved by that reason. But, of course, here the thing one did which was blameworthy is an exercise of minimal control or series of such exercises – just as (1) would have it.

On the second way to forget something due to inadequate concern, one's inadequate concern immediately[38] causes one to *not* occurrently represent, in the first place, certain reasons to organize one's attention in the relevant way. But here, *if* we have granted simply being a certain way cannot in itself constitute a ground for appropriate blame, we should also grant one is not blameworthy for this. The demonic neuroscientist grinned when he made his victim indifferent to friends. He grins again when the thought of checking their social calendar, or the vague sense that it would be good to connect with people, or so on, simply does not cross the victim's mind (consciously *or* unconsciously) on lazy Saturday afternoons as it often did before. If

[35]I thank an anonymous referee for this proposal. Arpaly and Schroeder (2014) develop an account of blameworthiness in terms of intrinsic desire; but note it is phrased in terms of actions which manifest the relevant desires (p. 170). In earlier work, though, Arpaly says 'one can be condemned for having racist or sadistic desires in the first place' (Arpaly 2003, p. 143).

[36]The moniker 'demonic neuroscientist' comes from Fischer (1999, p. 126).

[37]I thank an anonymous referee for advice here. Notice the idea is *not* that forgetting a friend's birthday entails one has inadequate concern for them. As Talbert (2017, pp. 56–59) points out in this context, things happen. Rather the idea is that, in *some* cases, one's forgetting is explained by inadequacy of concern.

[38]In particular, without the mediation of any exercises of minimal control.

one is not blameworthy for the indifference all by itself, one is not blameworthy for this either. So, it seems holding one could be blameworthy for this second way to forget something due to inadequate concern would require us to retrace our steps and maintain that simply being a certain way *can*, in itself, constitute a ground for appropriate blame.

In lieu of another way of making good on the idea that, without exercising minimal control, one can nevertheless *do* something in the sense required for blameworthiness, (1) seems like the view to beat. In addition, it is worth mentioning that (1) is a stripped-down component of the influential *reasons-responsiveness* approach to systematizing the pervasive idea that blameworthiness require control,[39] a component even some who argue against

[39]As exemplars of this approach, take Fischer and Ravizza (1998) and their notion of *guidance control*. *e* is an exercise of such control only when the way *e* occurs would involve recognition of a certain pattern of reasons regarding *e* and could involve recognition of a reason against *e* on whose basis one refrains from *e* (Ibid., pp. 71–76). As Fischer and Ravizza clarify, *e* occurs in this way only if *e* is done 'for a reason' in 'the *actual sequence*' (Ibid., p. 64). For this last point, see also Mele (2010, p. 104) and McKenna (2013, pp. 154–158). Now, if *e* is done for a reason one recognizes, *ipso facto e* is a response to a reason one occurrently represents. So, guidance control implies minimal control.But minimal control does not imply guidance control, on several counts. Guidance control requires (i) that *e* be done *for* a reason, in particular, (ii) that the occurrent representation be *recognition*, in particular, (iii) that a pattern of reasons would be represented, (iv) that *e* comes about in a way which could involve representing a reason against *e* on whose basis one refrains from *e*, and (v) that the way for *e* to come about have a causal history of the right sort (see Fischer and Ravizza 1998, pp. 89–91). Minimal control requires none of these things. I have already argued, in relation to the correlative feature of Levy's consciousness requirement, that not requiring (i) is a virtue of a requirement on blameworthiness, as *e*'s being done in spite of a reason is plenty. And for critical discussion of (v), see Mele (2010, p. 109).Notice, also, that the fact that something is an exercise of guidance control only if done for a reason in the actual sequence vindicates my using '*e*' – a variable for fine-grained events – when characterizing Fischer and Ravizza's view. For this fact means exercises of guidance control are responses to reasons. So how events and responses to reasons are individuated must align in Fischer and Ravizza's framework. And, as illustrated in Section 2, responses to reasons are fine-grained events.

[40]Clarke, for example, appears committed to (1). Consider his case of Ann, who runs a stop sign because she began thinking about her work (Clarke 2017, pp. 238–239). For Clarke, Ann is blameworthy because she had the capacity to 'maintain attention on' her driving, perhaps due to her capacity to 'notice features of [her] situation and appreciate their normative significance' (pp. 241–242). When Levy objects that Ann cannot 'exercise her power to attend *by a reasoning procedure*' (Levy 2017a, p. 255), Clarke replies that

> … she is capable of realizing the need to attend, and she is able to act rationally in response to that recognition. Thus, she has an ability to attend in response to the recognition of sufficient reason to do so …. To blame her for not stopping isn't to blame her for failing to do what she couldn't have done rationally. (Clarke 2017, p. 250)

Clarke is saying Ann had the capacity to stop 'rationally' – for a reason, I take it – by virtue of the fact, ultimately, that she had the capacity to 'realize the need to attend'. But if it matters to Clarke that Ann has the capacity not merely to stop but to do so *for a reason*, it also must matter, or at least *should*, that she has the capacity not merely to realize the need to attend but to do so *for a reason*. She could not have the capacity to do this, though, without somehow occurrently representing a reason to attend – for example, *that she is driving*. So, on Clarke's framework, Ann could be blameworthy, ultimately, only on account of an exercise of minimal control – namely, her not realizing the need to attend in spite of a reason to do so she occurrently represented.

consciousness requirements appear committed to.[40] And to be fair, it should not be a surprise to see acceptance of (1) together with rejection of consciousness requirements. We can grant (1) for the reason that blameworthiness requires doing something but wonder *why*, to count as doing something in the relevant sense, it is not enough to a respond to a reason one *occurrently represents*, full stop – even if not a reason of which one is *conscious*, in particular. Indeed, one might well think that one motivated by an implicit bias is blameworthy for what they are *doing* even granting that *what* they are doing is responding to a content they occurrently represent yet are *not* conscious of.[41] (2)–(6), that is, have substantial work to do. And whether they can do it should matter to anyone for whom (1) is so much as a theoretical option.[42]

(2) says only a person can do something blameworthy. I shall assume, with many others, that (2) is true.[43]

[40]Clarke, for example, appears committed to (1). Consider his case of Ann, who runs a stop sign because she began thinking about her work (Clarke 2017, pp. 238–239). For Clarke, Ann is blameworthy because she had the capacity to 'maintain attention on' her driving, perhaps due to her capacity to 'notice features of [her] situation and appreciate their normative significance' (pp. 241–242). When Levy objects that Ann cannot 'exercise her power to attend *by a reasoning procedure*' (Levy 2017a, p. 255), Clarke replies that

> … she is capable of realizing the need to attend, and she is able to act rationally in response to that recognition. Thus, she has an ability to attend in response to the recognition of sufficient reason to do so …. To blame her for not stopping isn't to blame her for failing to do what she couldn't have done rationally. (Clarke 2017, p. 250)

Clarke is saying Ann had the capacity to stop 'rationally' – for a reason, I take it – by virtue of the fact, ultimately, that she had the capacity to 'realize the need to attend'. But if it matters to Clarke that Ann has the capacity not merely to stop but to do so *for a reason*, it also must matter, or at least *should*, that she has the capacity not merely to realize the need to attend but to do so *for a reason*. She could not have the capacity to do this, though, without somehow occurrently representing a reason to attend – for example, *that she is driving*. So, on Clarke's framework, Ann could be blameworthy, ultimately, only on account of an exercise of minimal control – namely, her not realizing the need to attend in spite of a reason to do so she occurrently represented.

[41]At the end of Section 2, I explained how CONSCIOUSNESS implies one is not (directly) blameworthy for behavior motivated by implicit bias unless one has some inkling one is engaged in it.

[42]Sher accepts a 'voluntariness condition' on blameworthiness and expresses openness to its being guidance control (Sher 2009, pp. 149–151). As per endnote 39, this would imply (1).

[43]According to Locke, 'person' 'is a forensic term appropriating actions and their merit' (Locke 2009 [1690], p. 376). Stone says '[p]ersons are *conceived* as responsibility bearers' (Stone 1988, p. 530). And as Baker says, '[o]nly persons have moral responsibilities' (Baker 2007, p. 29).

(3) says (1) and (2) together imply there is a way of exercising minimal control which is both required for blameworthiness and distinctive of persons. My argument for (3) begins with the observation that even non-persons can exercise minimal control. Above, I playfully suggested automatic doors might well have this capacity. For a clearer case of non-persons with this capacity, consider chimps.[44] Chimps fight and sometimes kill their competitors. They can detect their competitors' motives and act violently in response. These behaviors qualify straightforwardly as exercises of minimal control.[45] Now suppose that (2) is true, which says only persons can do something blameworthy. It follows that chimps can never do something blameworthy, despite the various ways they exercise minimal control. Now suppose also that (1) is true, which says only exercises of minimal control can be blameworthy. It follows that there is a way of exercising minimal control both required for blameworthiness and distinctive of persons.[46] For if there is no particular way of exercising minimal control required for blameworthiness, or if there is such a way but chimps can engage in it too, nothing could prevent its being the case that, in possible cases, chimps are blameworthy for how they exercise minimal control.

I shall now argue for (4), which says that a way of exercising minimal control is distinctive of persons only if the occurrent representation of the reasons involved is distinctive of persons. Now, a way of exercising minimal control – that is, a way of occurrently representing and responding to reasons – is distinctive of persons only if either the response is distinctive, the reasons themselves are distinctive, or the occurrent representation of them is distinctive.

I shall first argue that if the response is distinctive, this could only be *because*, more fundamentally, either the reasons or the occurrent representation of them is distinctive. *e* counts as a *response* to a reason r, in the first place, only if *e* is done on the basis of r or in spite of r. *e* is fine-grained: a thing's instantiation of a property at a time. Call the property P. *e* is done on the basis of r only if the instantiation of P is motivated by r. And *e* is done in spite of r only if P instantiates and, for some property Q the instantiation

[44]As in Gruen (2017, § 3.1), there is controversy whether chimps are, or should be, *legal persons*. But the issue there is whether they have certain rights our laws should respect. And even if they do, it does not follow they are persons in the sense at issue when we say only persons can be blameworthy. Should one think they are persons in *that* sense, speaking instead of dogs, cats, and so on, would not affect my argument. I thank an anonymous referee for pressing me on this point.Also, when *The Cambridge Declaration on Consciousness* (Low 2012) says chimps and other non-human animals have *consciousness*, *phenomenal* consciousness is meant. It does not follow chimps have consciousness in the sense at issue in this paper.

[45]In this connection, see Glock's (2009) argument that animals can act for reasons.

[46]My talk of the 'way' one exercises minimal control, and so on, is similar to Fischer and Ravizza's talk of 'the way the action comes about', talk they themselves trade in for talk of 'mechanisms' (Fischer and Ravizza 1998, p. 38, footnote 8). All such talk faces the so-called 'generality problem'. Addressing this issue would take us too far afield. But see Fischer (2004, p. 169).

of which r was presently available to motivate, either Q has the form *not-P* or P has the form *not-Q*.[47]

In this way, what could count as a response to r, in the first place, is fixed by which properties r is presently available to motivate the instantiation of. There is something distinctive about a response to r only if there is something distinctive about one of these properties. But then we must ask what makes it the case, in the first place, that a given property's instantiation is something a given proposition is presently available to motivate. Because a proposition not occurrently represented in any way is also not presently available to motivate the instantiation of any properties, the answer is that this particular proposition is occurrently represented in some particular way. Given this, it is unclear what could explain which properties the proposition is presently available to motivate the instantiation of besides features of the proposition itself or else the manner in which it is occurrently represented. But this is just to say that, if the response is distinctive, this could only be because, more fundamentally, either the reasons themselves are distinctive or the occurrent representation of them is distinctive.

Of the remaining two options, the occurrent representation of the reasons and the reasons themselves, it is unclear whether it could be just the reasons themselves which are distinctive. For suppose persons can occurrently represent the purportedly distinctive reasons in a certain way. And suppose non-persons can also occurrently represent things in that very way. Then it is unclear what could explain why non-persons could not also occurrently represent the reasons purported to have been distinctive. Consider an analogy. Had I the same X-ray vision Superman can use to see a certain kind of thing, it would be odd if nevertheless I could not, even in principle, use this power to see it, too. Similarly, had non-persons the same representational capacity we can use to represent a reason of a certain kind, it would be odd if nevertheless they could not, even in principle, use this capacity to represent it, too. So, it seems a way of exercising minimal control is distinctive of persons only if the occurrent representation of reasons, at a minimum, is distinctive of persons. And that is precisely what (4) says.

To confess: the X-ray vision analogy gives only inconclusive support to (4). Perhaps the gap between Kryptonian and human intelligence is great enough Superman can see with X-ray vision certain things even humans with X-ray vision never could, in something like the way humans can look at a chessboard and see the blundering of a queen – something even animals with keen eyesight never could. In this way, I believe, Arpaly and Schroeder

[47]To explain with examples already used in this paper, Zac's taking the bag is done in spite of *that the bag belongs to a poor woman* because P, *taking the bag*, instantiates while this proposition was presently available to motivate Q, *not taking the bag*. Here the Q-property has the form *not-P*. But Alessandra's not organizing her attention in the relevant way is done in spite of *that the dog is in the car* and *that it is a hot day* because P, *not organizing her attention in this way*, instantiates while these propositions were presently available to motivate Q, *organizing her attention in this way*. Here, by contrast, the P-property has the form *not-Q*.

would resist (4) and maintain that persons alone can gain the concepts needed to represent, *in any fashion*, genuinely *moral* reasons – so that what is distinctive about how persons exercise minimal control is the reasons they represent, not the way they represent them. This objection to (4) warrants sustained discussion, and the next section, Section 4, is devoted to it. Before then, though, it would be good to have the rest of my argument on the table.

Let us turn to (5), the idea that persons are *rational beings*. This is an idea often found in philosophy, and it would take me too far afield to attempt a defense of it.[48] Instead I content myself with a few comments on what I take the idea to mean.

To say persons are rational beings is, at least, to say they must have *rational capacities*: capacities for (System 2) reasoning events of various kinds, for example, theoretical, moral, and practical. Presumably, this is why non-human animals, lacking these capacities, are not persons. There is more to the idea that persons are rational beings, though. For it might also be true that persons must be in space and time. But, if so, the sense in which persons are rational beings is deeper than the sense in which they are spatiotemporal beings. The idea is that a being is a person *by virtue of* having rational capacities. More precisely, *only* by virtue of having these capacities could a being be a person, and a being is a person *simply* by virtue of having them. Or, as we can put it, a being is a person *exactly* by virtue of having them.[49]

This clarification of what it means to say persons are rational beings informs what it takes to argue for (6), the last premise of my argument. (6) says if persons are rational beings, then one occurrently represents r in a way distinctive of persons only if one is conscious of r. Explicated, (6)'s antecedent says a being is a person exactly by virtue of having rational capacities. If it is true, what is distinctive of persons is *one and the same* as what is distinctive of rational beings. Hence, (6) is true if

> (6)^{R} For all x, for all e, for all r, x occurrently represents r in a way distinctive of *rational beings* at the time of e only if x is conscious of r at the time of e.

[48]Locke defines a person as 'a thinking intelligent being that has *reason and reflection*, and can consider itself as itself …' (Locke 2009 [1690], p. 370). Kant holds that 'beings without reason … are called *things*; rational beings, by contrast, are called *persons* …' (Kant 2002 [1785], p. 46). And Dillon says 'to be a person is to be a rational being' (Dillon 2007, p. 122).
Despite these precedents, one might object to (5) that infants and the severely intellectually impaired are, nevertheless, persons. One response to this worry is to say that infants and the impaired are rational beings, just ones with undeveloped or blocked rational capacities. Another is to hold that, even if not persons, they are for other reasons not to be excluded from our moral community (see, e.g., Steinbock 1978). But, again, defending (5) would take us too far afield; in this paper, I treat it as an assumption.

[49]According to Sher, a person is no mere reason-responder but instead 'an enduring causal structure whose elements interact in ways that *give rise to*', inter alia, responsiveness to reasons (Sher 2009, p. 121). This is consistent with (5), though. That a being is a person exactly by virtue of having rational capacities does not tell us what persons *are*, metaphysically speaking. Indeed, rational capacities seem to require an 'enduring causal structure' something like Sher describes.

In my argument for $(6)^R$, I will assume, to fix ideas, that representing r is constituted by having a set of dispositions to respond in certain ways to r.[50] This allows us to zero in on the question of what it takes for such a disposition to be distinctive of rational beings. My argument begins with a claim about it takes for such a disposition to feature in an event of *reasoning*, and runs as follows:

(6.i) The manifestation of a disposition *d* to respond in a certain way to r constitutes an event of reasoning in *x* only if r is presently available to motivate *x* in *x*'s reasoning at the time of *d*'s manifestation.
(6.ii) If (6.i), then such a disposition *d* is distinctive of rational beings only if *x* is conscious of r at the time of *d*'s manifestation.
(6.iii) If the consequent of (6.ii) is true, then $(6)^R$ is true.

Here is my argument for (6.i). Suppose *d*'s manifestation is a response to r which constitutes an event of reasoning in *x*. Then *d*'s manifestation is *x*'s being motivated in a specific way by r, or *x*'s doing otherwise in spite of r, *in x's reasoning* – else, while *d*'s manifestation may well be a *response* to r, it would not constitute *reasoning* as opposed to some other kind of event. But *x* could be motivated in a specific way by r in *x*'s reasoning at a given time only if r was presently available to motivate *x* in *x*'s reasoning at that time. And *x* does otherwise in spite of r in *x*'s reasoning only if, again, r was presently available to motivate *x* in *x*'s reasoning at the relevant time. So, if *d*'s manifestation constitutes an event of reasoning in *x*, r is presently available to motivate *x* in *x*'s reasoning at the time, just as (6.i) says.

I turn now to (6.ii), on which if (6.i) is true, *d* is distinctive of rational beings only if *x* is conscious of r at the time of *d*'s manifestation. If it is not the case that either

(R) *d*'s manifestation constitutes an event of reasoning in *x*,

or

(RR) *d*'s manifestation makes it the case that r is presently available to motivate *x* in *x*'s reasoning,[51]

then *d*'s manifestation leaves r entirely out of the reach of rational capacities – capacities to put r to use in *reasoning*. But a disposition whose manifestation leaves r out of the reach of rational capacities is not distinctive of rational beings. For there will be *some* possible world in which a creature which is not a rational being has it.

[50]For one of the ways to make this supposition concrete, see Schwitzgebel (2002).

[51]An example of an (RR)-type disposition is one whose stimulus condition is being specially prompted and whose manifestation makes r presently available to motivate *x* in *x*'s reasoning.

So, *d* is distinctive of rational beings only if either (R) or (RR) is true of *d*. Now suppose (6.i) is true: *d*'s manifestation constitutes an event of reasoning in *x* only if r is presently available to motivate *x* in *x*'s reasoning. Then (R) obtains only if r is presently available in this way. The same is true of (RR), as, by the definition of (RR), *d*'s manifestation makes r presently available in this way at the time. But for r to be presently available in this way is just for *x* to be *conscious* of r, given the notion of consciousness we are working with. This means either (R) or (RR) obtains only if *x* is conscious of r at the time of *d*'s manifestation. So, (6.ii): if (6.i) is true, *d* is distinctive of rational beings only if *x* is conscious of r at the time of *d*'s manifestation.

Consider next (6.iii), which connects what it takes for *d* to be distinctive of rational beings to what it takes to occurrently represent r in a way distinctive of rational beings. Suppose the consequent of (6.ii) is true, that is, that *d* is distinctive of rational beings only if *x* is conscious of r at the time of *d*'s manifestation. It follows, keeping in view of our idea-fixing assumption that representing r is constituted by having a set of dispositions to respond in certain ways to r, that insofar as *x* represents r in a way distinctive of rational beings, each disposition *d* constituting this representation is such that *x* is conscious of r at the time of *d*'s manifestation.[52] And because the *occurrence* of a representation involves the manifestation of at least one of its dispositions, it follows that *x occurrently* represents r in a way distinctive of rational beings only if *x* is conscious of r at the time. So, if *d* is distinctive of rational beings only if *x* is conscious of r at the time of *d*'s manifestation, then *x* occurrently represents r in a way distinctive of rational beings only if *x* is conscious of r at the time. And this is exactly what (6.iii) says.

(6.i)–(6.iii) imply $(6)^{R}$, which implies (6), the last premise of my argument for CONSCIOUSNESS.

4. An objection to (4)

(4) says an exercise of minimal control is distinctive of persons only if the occurrent representation involved, in particular, is distinctive of persons. In this section, I consider the objection that what makes an exercise of minimal control distinctive of persons is not the occurrent representational states they alone can have but instead the *reasons* they alone can occurrently represent. The objection is inspired by Arpaly (2003) and Arpaly and Schroeder (2014).[53]

Arpaly argues animals[54] cannot represent specifically *moral* reasons. The dog who ruins a favorite toy dinosaur

[53]I thank an anonymous referee for helping me understand the nature and force of this objection.
[54]Here and hereafter, I use 'animals' as shorthand for 'animals which are not persons'.

> … does not understand *mine*, *favorite*, or *dinosaur*, not even in the murky, visceral way a small child does. Similarly, the dog's mind presumably cannot grasp – nor can it track, the way even unsophisticated people can – such things as increasing utility, respecting persons, or even friendship. (Arpaly 2003, p. 146)

Representing a moral reason requires a '[m]oral education', that is, 'some help from one's own or other people's reflection' (Ibid., p. 147). Animals lack the rational capacities required for this education.

And it is not just *any* 'help' from 'reflection' that is needed, too. Upon reflection, Ron adopts the wacky view that his sole moral obligation is to ensure things come in twos. He purchases just one ice cream cone. He then feels guilty, not because his friend who forgot their wallet must now go without ice cream, but because his cone is not one of *two*. Unless Ron can grasp what is *actually* right or good about buying two cones, he no more represents a moral reason than the dog. And to do this, he must use *correct* concepts about what is right or good, which, according to Arpaly and Schroeder, are 'the concepts deployed in grasping the correct normative moral theory' (Arpaly and Schroeder 2014, p. 164). Buying two cones

> … must be presented by concepts that would allow [Ron] to trivially deduce that it is necessarily an instance of MAXIMIZING HAPPINESS, or RESPECTING PERSONS, or whatever the correct normative theory distinguishes as the right or good …. (Ibid., p. 167)[55]

And if 'the right or good is irreducibly plural' on the correct theory, the purchase must fall under the guise of at least one of the different kinds of rightness or goodness (Ibid., p. 164).

Persons stand apart from animals because they are capable of gaining correct concepts of the right or good. Moreover, after these concepts are in place, a person can represent moral reasons not just consciously but also *unconsciously*. Here Arpaly draws an analogy with sports. In learning the game, a football player gains concepts and beliefs which allow them to represent certain reasons while playing it. And after, but only after, these are in place, the player can represent and respond to the relevant reasons not just consciously but also unconsciously – as when, for example, a quarterback makes 'a brilliant last-minute pass' (Ibid, p. 146). Similarly in the moral case. When Twain's Huck Finn, raised in the antebellum south, refrains from turning Jim over to slavecatchers despite believing he ought to do so, Huck, says Arpaly, is moved by a moral reason he unconsciously represents – perhaps 'Jim's humanity' (Arpaly 2003, p. 10). But Huck can represent this reason in the first place only because he has 'a background of beliefs' which could only have been 'formed in specifically human ways', including 'the

[55] Arpaly and Schroeder (Ibid., p. 164) use block capitals to refer to concepts, following Fodor (1998). On occasion I will too.

belief that friends should be helped, the belief that property should be respected, and so on' (Arpaly 2003, p. 147).

Thus, so runs the objection to (4), Huck's exercise of minimal control is distinctive of persons *not* because the occurrent representational states involved are distinctive of persons but because he occurrently represents a *moral* reason, something an animal cannot gain the concepts needed to represent, 'the concepts deployed in grasping the correct normative moral theory' being out of their reach. I take this kind of objection to be especially important. For it does not reject any of the ideas which put us on a search for an exercise of minimal control distinctive of rational beings. Instead, taking this search seriously, it purports to show that what we find gives us no reason to accept CONSCIOUSNESS.

To respond. An initial question is how to interpret the idea that representing a moral reason requires using the concepts 'deployed in grasping the correct normative moral theory'. Take MAXIMIZATION *a la* utilitarianism. Even some philosophy majors fail to gain this concept. These same people, though, represent moral reasons. True, Arpaly and Schroeder take it to be a count against a normative theory if, given their account of what it takes to represent a moral reason, the theory implies many people do not represent moral reasons (Arpaly and Schroeder 2014, pp. 223–224). If their account plus utilitarianism would require the grasp of MAXIMIZATION, the upshot would then be that we should reject utilitarianism. But this seems too quick a way to learn utilitarianism is false.[56] Better, I think, to interpret their account as requiring the grasp of not *every* concept constitutive of the correct normative theory but rather some handful of basic concepts of which the theory is a systematization – for example, for utilitarianism, PAIN, PLEASURE, CAUSE, and so on.

The next thing to see is that some normative theories systematize basic concepts some animals *do* have. When chimps fight, striking and scratching one another, they seem to represent and respond to what would CAUSE PAIN. But those are basic concepts of utilitarianism. So, the objection to (4) in view appears to fail if utilitarianism is true. For then 'the concepts deployed in grasping the correct normative moral theory' – once we charitably interpret what those concepts amount to – *are* within the cognitive wheelhouse of animals, after all.

Of course, some theories systematize basic concepts animals do *not* have, for example, Scanlon's contractualism:

> An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement. (Scanlon 1998, p. 153)

[56]See Sliwa (2017, pp. 141–142) for further development of this sort of point.

REASONABLE REJECTION, INFORMED AGREEMENT, and so on, are not within the cognitive wheelhouse of chimps. So, the objection to (4) has legs if contractualism is correct. But if contractualism *is* correct, Arpaly and Schroeder's account of what it takes to represent a moral reason implies, quite implausibly, that only lucky people get to do so. The basic concepts of contractualism require not just an education only a person could have but one many, many people never receive. But even those who cannot think of their actions in terms of principles it would not be reasonable to reject as the basis for informed agreement can still, surely, represent moral reasons.

The problem looming for the objection to (4) does not depend on exactly which concepts utilitarianism or contractualism involve, or even whether it is best to interpret Arpaly and Schroeder as requiring only the basic concepts of the correct theory. My comments on those fronts aim simply to bring out that if the concepts required by the correct theory are simple enough, it is not just persons but also animals which can possess them, while if they are complex enough for it to be clear that animals cannot possess them, the question becomes pressing whether only lucky people could.

Of course, in principle, one could hold that the concepts required by the correct theory *have just enough complexity that animals cannot possess them but not so much that only lucky people could*. If a concept meets this condition, let us say it occupies *the sweet spot*. Concepts in the sweet spot would be ungraspable for chimps but essentially universal among people, from the rural Mongolian who honors their ancestors in the spirit world to the secular, university-educated Norwegian. It would surprise me if *any* concepts occupy the sweet spot, actually. But even if some do, the thing to see is the objection to (4) in view works *only if* the concepts of whatever normative theory turns out to be correct are among them. The view that the concepts of the correct theory occupy the sweet spot allows one to reject (4). But without independent reason to accept that view, it is, in this context, the definition of *ad hoc*.[57]

Perhaps the following argument is independent reason to believe the concepts of the correct theory occupy the sweet spot. We know, as a kind of

[57]As Sliwa writes, the

> … commitment that the concepts that figure in the agent's desires must be those that figure in the correct normative theory has implausible consequences. It presents … a dilemma: either there are implausibly high intellectual demands on who can perform morally admirable actions or there are implausible constraints on what the correct normative theory must look like. (Sliwa 2017, p. 142)

This dilemma inspires my *trilemma*: The concepts required by the correct theory are either (i) simple enough animals possess them, or (ii) complex enough only lucky people do, or (iii) occupants of the sweet spot. The (i)-view fails to get the objection to (4) off the ground, the (ii)-view generates false implications, and endorsing the (iii)-view without independent reason to do so is *ad hoc*.

datum, that animals cannot grasp what is actually right or good while essentially all people can. But to grasp what is actually right or good is to use the concepts of the correct normative theory. Hence, whatever theory this may be, we know its concepts occupy the sweet spot.

But that is a bad argument. Compare: We know, as a kind of datum, that animals cannot grasp what is actually water while essentially all people can. But to grasp what is actually water is to use the concepts of the correct chemical theory of water. Hence, whatever this theory may be, we know its concepts occupy the sweet spot.

The water argument is bad because while water *is* H2O, it is not the *concept* H2O which allows people to grasp what is actually water. WHAT FALLS FROM THE SKY, WHAT QUENCHES THIRST, and the contents of perceptual experience, somehow, work just fine. Two things follow from this. First, the premise that people but not animals can grasp what is actually water has nothing to do with whether the concepts of the correct chemical theory occupy the sweet spot. Second, that premise's implausibility becomes evident; for, plausibly, some animals can represent some of the unsophisticated contents by virtue of which people grasp what is actually water.

Similarly, even if the right or good *is* maximizing happiness, it is not MAXIMIZING HAPPINESS which allows people to grasp what is actually right or good. WHAT WOULD CHEER UP EMILY, WHAT MY SPOUSE REQUESTED, and so on, as well as, plausibly, the contents of some emotions, somehow, work just fine.[58] Two things follow from this. First, the premise that people but not animals can grasp what is actually right or good has nothing to do with whether the concepts of the correct normative theory occupy the sweet spot. Second, that premise's implausibility becomes evident; for, plausibly, some animals can represent some of the unsophisticated contents by virtue of which people grasp what is actually right or good.[59]

[58]One view in moral epistemology is that emotions are to the moral facts what visual experiences are to the physical facts; see, for example, Johnston (2001), Döring (2007), and Milona (2016). Another view, for example, Nussbaum's (2004), is that emotions are moral judgments. Either sort of picture, and perhaps others too, could vindicate the modest view that at least *some* emotions have contents which play *some* role in how we grasp what is right or good. Besides following from straightforward observations about how ordinary people often grasp what is right or good, support for this modest view also derives from the fact that psychopathy – a condition characterized by deficits in emotion – appears to make it harder for one to grasp what is right or good. On this last point, see Blair (1995), Nichols (2008), and McGeer (2008).

[59]None of this is to deny that in certain contexts – a difficult moral dilemma, say – the correct normative theory may be our *only* hope for progress, just as conceiving of water as H_2O may be indispensable in certain scientific or technological contexts.

[60]Another strategy of response is to say both that Huck is not conscious of any reason which motivates him and that his being motivated by the relevant reason is not distinctive of persons. While I find this strategy implausible, see Sliwa (2017, p. 139).

In lieu of some other reason to think the concepts of the correct normative theory occupy the sweet spot, or some other way of maintaining that no animals but essentially all people can represent moral reasons, the objection to (4) remains *ad hoc*.

A final way to revive the objection to (4), though, is to cast it directly in terms of Huck Finn. According to Arpaly, when Huck refrains from turning Jim over to the slavecatchers, no reason of which he is conscious motivates him. But he *is* motivated by a reason and, intuitively, he therein exercises control in a way distinctive of persons. Not being conscious of this reason, it could not be by virtue of the *kind* of representational state involved that this exercise of minimal control is distinctive of persons. It must be by virtue of something else, perhaps the kind of content the relevant state has. However, we ultimately make sense of this, (4) remains false.

My response is that Arpaly is mistaken to think Huck is not conscious of any reason which motivates him.[60] Consider Huck's internal dialogue after he protects Jim:

> Then I thought a minute, and says to myself, hold on, – s'pose you'd a done right and give Jim up; would you felt better than what you do now? No, says I, I'd feel bad – I'd feel just the same way I do now. (Twain 1886, p. 128)

Huck, clearly, had conflicting emotions. Now, his upbringing gave him more vocabulary to *label* the content of the emotion inclining him to 'give Jim up', but it does not follow that he is not *conscious* of the content of the emotion inclining him to protect him. Much we cannot identify or label – the shades of colors represented in visual experience, logical relationships before we take a logic class, and on and on – can nevertheless constitute contents presently available to motivate us in our reasoning.[61] Just so for Huck. The content inclining him to protect Jim, his inability to describe it notwithstanding, played a starring role in the slow, effortful, distinctively human, System 2 reasoning event which was his refraining from turning Jim in.[62] That content was presently available to motivate him in his reasoning and, indeed, it did so. Huck was conscious of it in the sense at issue in this paper.

What I have said about Huck does not depend on the identity of the content of the emotion inclining him to protect Jim. So it is unclear how the identity of its content could explain why no chimp could have this emotion. We could nicely explain this, though, by saying its content is presently available to capacities chimps do not possess, namely, *rational* capacities. No

[61]Such contents would be a subtype of what is sometimes called *nonconceptual* content; for a survey of the notion, see Bermúdez and Cahen (2020).

[62]Indeed, the slow, dramatic internal faceoff between it and the contents of Huck's other emotions is a large part of what makes Twain so compelling a read, here.

[63]See endnote 39 for more about guidance control.

chimp could have the emotion in question because it is a state of *consciousness*. Reviving the objection to (4) directly in terms of Huck Finn fails.

5. *Conclusion*

As mentioned in Section 1, to this point Neil Levy (2014) has given the most developed argument from a control requirement to a consciousness requirement. Levy's argument proceeds via two claims within cognitive science. The first is that the flexible responsiveness involved in guidance control requires *integration* – namely, the availability of the morally salient features of one's situation for assessment 'for consistency and for conflict with' a broad range of one's attitudes (Ibid., p. 113).[63] The second is that due to the 'global neuronal workspace' account of consciousness, consciousness alone suffices for integration (Ibid., p. 49). As to be expected, some question whether Levy has the science correct. Sripada, for example, questions the first claim, arguing that humans have enough 'motivational modules' for guidance control in the absence of integration (Sripada 2015, p. 40). And Mudrik *et al*. (2011) question the second, arguing that unconscious processes can secure integration. One thing exciting about Levy's argument is that cognitive science may vindicate it in the face of these objections.

By the same token, something comforting about my argument, at least for those who find CONSCIOUSNESS attractive, is that it would still stand even if Sripada or Mudrik *et al*. are *right*. Suppose that, by virtue of Parks' motivational modules, his killing his mother-in-law was an exercise of guidance control. Alternatively, suppose that unconscious processes integrated the relevant features of Parks' situation. Then Levy's argument that Parks is not blameworthy no longer goes through. But mine still would. Parks was not conscious of anything at the time. So, the dispositions implicated in his exercises of guidance control, or in the unconscious integration of the relevant features of his situation, do not bear the needed relationship to reasoning. So, no exercise of minimal control distinctive of persons occurred. So, Parks could not be blameworthy.

Let me address a final objection, which runs as follows: (2) says only a *person* can do something blameworthy. Now, we have reason to accept (2) only if we have reason to think the features required for blameworthiness can only be instantiated by persons. But then, when it comes to reason to think blameworthiness requires consciousness, the only thing which matters is whether these features involve consciousness. And if that is right, the concept of the person should be dropped from the argument: it is an irrelevant fifth wheel.

[64]For a theory of the grounding relation, see, for example, Paul Audi (2012). Notice that, *a la* endnote 5, grounding relations can hold even among conceptual connections.

This objection calls for a general comment about the approach taken in this paper. I have treated the restriction of blameworthiness to persons as a datum from which to argue, not something for which argument is needed. Now, it might be that the fact that only persons can be blameworthy is not metaphysically rock-bottom. That is, it might be in virtue of other more fundamental facts that it is a fact in the first place – that something *grounds* the fact that only persons can be blameworthy.[64] And it might also be that whatever grounds this also grounds CONSCIOUSNESS. But even if all of this is right, it is a count against my argument only if the (purported) facts to which its premises refer should be the very same which ground the (purported) fact to which its conclusion refers. And it is simply not true that all arguments should reflect the order of grounding in this way, any more than, because fire causes smoke and not vice versa, an argument from smoke to fire is thereby untoward. And in particular, even if both the fact that only persons can be blameworthy and CONSCIOUSNESS are grounded by other, more fundamental facts, it does not follow that my argument is untoward.

Besides, sorting out the order of grounding vis-à-vis personhood, reasoning, consciousness, blameworthiness, and so on, is a difficult task unto itself.[65] My hope in this paper is accordingly modest. It is that we reflect on what to do with a certain handful of ideas. Here is one way to put it. The premises of my argument and the negation of CONSCIOUSNESS constitute an inconsistent set. Which member of this set should we reject?[66]

Department of Humanities & Languages
Flame University, Pune, India

REFERENCES

Adams, R. (1985). 'Involuntary Sins,' *Philosophical Review* 94, pp. 3–31.

Alston, W. (1980). 'Level-Confusions in Epistemology,' *Midwest Studies in Philosophy* 5(1), pp. 135–150.

Arpaly, N. (2002). 'Moral Worth,' *The Journal of Philosophy* 99(5), pp. 223–245.

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.

Arpaly, N. and Schroeder, T. (2015). 'A Causal Theory of Acting For Reasons,' *American Philosophical Quarterly* 52(2), pp. 103–114.

Arpaly, N. and Schroeder, T. (2014). *In Praise of Desire*. New York: Oxford University Press.

Audi, R. (1994). 'Dispositional Beliefs and Dispositions to Believe,' *Noûs* 28(4), pp. 419–434.

Audi, P. (2012). 'Grounding: Toward a Theory of the In-Virtue-Of Relation,' *Journal of Philosophy* 109(12), pp. 685–711.

[65]For related discussion, see Hatcher (2017, §4.6).

[66]For invaluable feedback on previous drafts of this paper, I thank Mark Schroeder, Janet Levin, Jim Van Cleve, and, especially, Ralph Wedgwood. I also thank audience members of the 2017 meeting of the Long Island Philosophical Society Conference, anonymous referees from a number of journals and, particularly, an extremely helpful anonymous referee for this journal.

Baker, L. (2007). 'Persons and Other Things,' *Journal of Consciousness Studies* 14, pp. 17–36.
Bermúdez, J., and Cahen, A. (2020) 'Nonconceptual Mental Content,' in Zalta, Edward (ed) *The Stanford Encyclopedia of Philosophy*, URL: https://plato.stanford.edu/archives/sum2020/entries/content-nonconceptual/
Blair, R. J. R. (1995). 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath,' *Cognition* 57(1), pp. 1–29.
Block, N. (1995). 'On a Confusion About a Function of Consciousness,' *Brain and Behavioral Sciences* 18(2), pp. 227–247.
Broome, J. (2013). *Rationality Through Reasoning*. Oxford: Wiley-Blackwell.
Broughton, R., Billings, R. et al. (1994). 'Homicidal Somnambulism: A Case Report,' *Sleep* 17 (3), pp. 253–264.
Clarke, R. (2014). *Omissions: Agency, Metaphysics, and Responsibility*. New York: Oxford University Press.
Clarke, R (2017) 'Ignorance, Revision, and Commonsense,' in Robichaud, P. and Wieland, J. W. (eds) *Responsibility: The Epistemic Condition*, 233–51. (New York: Oxford University Press)
Davidson, D. (1963). 'Actions, Reasons, and Causes,' *The Journal of Philosophy* 60, pp. 685–700.
Dillon, R. (2007). 'Arrogance, Self-Respect and Personhood,' *Journal of Consciousness Studies* 14, pp. 101–126.
Döring, S. (2007). 'Seeing What to Do: Affective Perception and Rational Motivation,' *Dialectica* 61(3), pp. 363–394.
Evans, J. (2008). 'Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition,' *Annu Rev Psychol* 59, pp. 255–278.
Fischer, J. (1999). 'Recent Work on Moral Responsibility,' *Ethics* 110(1), pp. 93–139.
Fischer, J. (2004). 'Responsibility and Manipulation,' *The Journal of Ethics* 8(2), pp. 145–177.
Fischer, J. and Ravizza, M. (1998). *Responsibility and Control*. New York: Cambridge University Press.
FitzPatrick, W. (2017) 'Unwitting Wrongdoing, Reasonable Expectations, and Blameworthiness', in Robichaud, P. and Wieland, J. W. (eds) *Responsibility: The Epistemic Condition*, 29–46. (New York: Oxford University Press)
Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.
Glock, H.-J. (2009). 'Can Animals Act for Reasons?' *Inquiry* 52(3), pp. 232–254.
Goldman, A. (1976). 'Discrimination and Perceptual Knowledge,' *The Journal of Philosophy* 73 (20), pp. 771–791.
Graham, P. (2017) 'The Epistemic Condition on Moral Blameworthiness: A Theoretical Epiphenomenon', in Robichaud, P. and Wieland, J. W. (eds). *Responsibility: The Epistemic Condition*, 163–179.(New York: Oxford University Press)
Gruen, L. (2017) 'The Moral Status of Animals', in Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy*, URL: https://plato.stanford.edu/entries/moral-animal/
van Gulick, R. (2018) 'Consciousness', in Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy*, URL: https://plato.stanford.edu/archives/spr2018/entries/consciousness/
Hatcher, M. (2017). *A Deontological Explanation of Accessibilism*. PhD Dissertation. USC.
Hatcher, M. (2018). 'Accessibilism Defined,' *Episteme* 15(1), pp. 1–23.
Holroyd, J. (2012). 'Responsibility for Implicit Bias,' *Journal of Social Philosophy* 43(3), pp. 274–306.
Holroyd, J. (2015). 'Implicit Bias, Awareness and Imperfect Cognitions,' *Conscious Cogn* 33, pp. 511–523.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Oxford University Press.

Johnston, M. (2001). 'The Authority of Affect,' *Philosophy and Phenomenological Research* 63 (1), pp. 181–214.

Kant, I. (2002). *([1785]) Groundwork for the Metaphysics of Morals (Translation by Allen Wood)*. London: Yale University Press.

Kelly, D. and Roedder, E. (2008). 'Racial Cognition and the Ethics of Implicit Bias,' *Philosophy Compass* 3(3), pp. 522–540.

Kenny, A. (1976). *Will, Freedom, and Power*. Oxford: Blackwell Publishers.

Kim, J. (1966). 'On the Psycho-Physical Identity Theory,' *American Philosophical Quarterly* 3 (3), pp. 227–235.

Knobe, J. and Nichols, S. (2017) 'Experimental Philosophy', in Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy*, URL: https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/

Levy, N. (2008). 'Restoring Control: Comments on George Sher,' *Philosophia* 36, pp. 213–221.

Levy, N. (2014). *Consciousness and Moral Responsibility*. Oxford: Oxford UniversityPress.

Levy, N. (2017a) 'Methodological Conservatism and the Epistemic Condition', in Robichaud, P. and Wieland, J. W. (eds) *Responsibility: The Epistemic Condition*, 252–265. (New York: Oxford University Press)

Levy, N. (2017b). 'Implicit Bias and Moral Responsibility: Probing the Data,' *Philosophy and Phenomenological Research* 94(1), pp. 3–26.

Locke, J. (2009) ([1690]) Essay Concerning Human Understanding, in Ariew, R. (ed) *Modern Philosophy: An Anthology of Primary Sources* (Cambridge: Hackett Publishing Company)

Low, P. (2012) '*The Cambridge Declaration on Consciousness*.' Churchill College, Cambridge. URL: http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

McGeer, V. (2008). 'Varieties of Moral Agency: Lessons from Autism (and Psychopathy),' *Moral Psychology* 3, pp. 227–258.

McKenna, M. (2013). 'Reasons-Responsiveness, Agents, and Mechanisms,' in D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility*. Oxford: Oxford University Press.

Mele, A. (2010). 'Moral Responsibility for Actions: Epistemic and Freedom Conditions,' *Philosophical Explorations* 13(2), pp. 101–111.

Mendelovici, A. (2018). *The Phenomenal Basis of Intentionality*. New York: Oxford University Press.

Milona, M. (2016). 'Taking the Perceptual Analogy Seriously,' *Ethical Theory and Moral Practice* 19(4), pp. 897–915.

Mudrik, L., Breska, A., Lamy, D. and Deouell, L. Y. (2011). 'Integration Without Awareness: Expanding the Limits of Unconscious Processing,' *Psychol Sci* 22, pp. 764–770.

Nichols, S. (2008) 'Sentimentalism Naturalized', in Sinnott-Armstrong, W. (ed) *Moral Psychology*, vol. 2: 255–274. (Cambridge: The MIT Press)

Nussbaum, M. (2004) 'Emotions as Judgments of Value and Importance', in Solomon, R. C. (ed.) *Thinking About Feeling: Contemporary Philosophers on Emotion*, 183-199. (New York: Oxford University Press)

Parfit, D. (1997). 'Reasons and Motivation I,' *Proceedings of the Aristotelian Society, Supplementary Volumes* 71, pp. 99–130.

Pitt, D. (2017) 'Mental Representation', in Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy* URL: https://plato.stanford.edu/archives/spr2017/entries/mental-representation/

Rachels, J. (2016) 'Active and Passive Euthanasia,' in Kuhse, H., Schüklenk, U., and Singer, P. *Bioethics: An Anthology*, 248–251. (Oxford: John Wiley & Sons, Inc.)

Robichaud, P. (2015). 'Some Doubts About the Consciousness Requirement for Moral Responsibility,' *Journal of Consciousness Studies* 22(7–8), pp. 26–36.
Rosen, G. (2004). 'Skepticism About Moral Responsibility,' *Philosophical Perspectives* 18(1), pp. 295–313.
Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.
Schroeder, M. (2007). *Slaves of the Passions*. New York: Oxford University Press.
Schroeder, M. (2008). 'Having Reasons,' *Philosophical Studies* 139(1), pp. 57–71.
Schwitzgebel, E. (2002). 'A Phenomenal, Dispositional Account of Belief,' *Noûs* 36, pp. 249–275.
Searle, J. (2004). 'Minds, Brains, and Programs,' in S. Shieber (ed.) *The Turing Test, 201–224*. Cambridge: MIT Press.
Setiya, K. (2007). *Reasons Without Rationalism*. New Jersey: Princeton University Press.
Shepherd, J. (2015). 'Consciousness, Free Will, and Moral Responsibility: Taking the Folk Seriously,' *Philosophical Psychology* 28(7), pp. 929–946.
Sher, G. (2009). *Who Knew? Responsibility Without Awareness*. Oxford: Oxford University Press.
Sliwa, P. (2017) 'On Knowing What's Right and Being Responsible for It', in Robichaud, P. and Wieland, J. W. (eds) *Responsibility: The Epistemic Condition*, 127–45. (New York: Oxford University Press)
Smith, A. (2005). 'Responsibility for Attitudes: Activity and Passivity in Mental Life,' *Ethics* 115 (2), pp. 236–271.
Smith, A. (2008). 'Control, Responsibility, and Moral Assessment,' *Philosophical Studies* 138, pp. 367–392.
Sripada, C. (2015). 'Acting from the Gut: Responsibility without Awareness,' *Journal of Consciousness Studies* 22(7–8), pp. 37–48.
Steinbock, B. (1978). 'Speciesism and the Idea of Equality,' *Philosophy* 53(204), pp. 247–256.
Stone, J. (1988). 'Parfit and the Buddha: Why There Are No People,' *Philosophy and Phenomenological Research* 48, pp. 519–532.
Strawson, P. F. (1962). 'Freedom and Resentment,' *Proceedings of the British Academy* 48, pp. 1–25.
Talbert, M. (2017) 'Akrasia, Awareness, and Blameworthiness', in Robichaud, P. and Wieland, J. W. (eds) *Responsibility: The Epistemic Condition*, 47–63. (New York: Oxford University Press)
Twain, M. (1886). *The Adventures of Huckleberry Finn*. New York: Charles L. Webster and Company.
Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.
Wedgwood, R. (2006). 'The Normative Force of Reasoning,' *Noûs* 40(4), pp. 660–686.
Wedgwood, R. (2013). 'Rational 'Ought' Implies 'Can',' *Philosophical Perspectives* 23, pp. 70–92.
Williams, B. (1995). 'Internal Reasons and the Obscurity of Blame,' in *Making Sense of Humanity*. Cambridge: Cambridge University Press.
Woodward, P. (2015). *The Emergence of Mental Content: An Essay in the Metaphysics of Mind*', Ph.D. Dissertation. Indiana University.
Zimmerman, M. J. (1997). 'Moral Responsibility and Ignorance,' *Ethics* 107(3), pp. 410–426.