Alex Madva

Equal Rights for Zombies?

Phenomenal Consciousness and Responsible Agency

Abstract: *Intuitively, moral responsibility requires conscious awareness of what one is doing, and why one is doing it, but what kind of awareness is at issue? Neil Levy argues that phenomenal consciousness — the qualitative feel of conscious sensations — is entirely unnecessary for moral responsibility. He claims that only access consciousness — the state in which information (e.g. from perception or memory) is available to an array of mental systems (e.g. such that an agent can deliberate and act upon that information) — is relevant to moral responsibility. I argue that numerous ethical, epistemic, and neuroscientific considerations entail that the capacity for phenomenal consciousness is necessary for moral responsibility. I focus in particular on considerations inspired by P.F. Strawson, who puts a range of qualitative moral emotions — the reactive attitudes — front and centre in the analysis of moral responsibility.*

1. Introduction

Intuitively, moral responsibility requires conscious awareness of what one is doing, and why one is doing it, but what kind of awareness is at issue? In *Consciousness and Moral Responsibility* (2014a), Neil Levy argues that phenomenal consciousness — the qualitative feel of conscious sensations — is not necessary for moral responsibility. He claims that access consciousness — the state in which information (e.g. from perception or memory) is available to an array of mental systems (e.g. such that an agent can deliberate, report, and act upon

Correspondence:
Alex Madva, California State Polytechnic University, Pomona, CA, USA.
Email: alexmadva@gmail.com

Journal of Consciousness Studies, **26**, No. 5–6, 2019, pp. 117–40

that information) — is the only type of consciousness necessary. However, Levy's argument against the necessity of phenomenal consciousness begs the question against the broad class of views inspired by Peter Strawson's 'Freedom and Resentment' (1962), which puts a range of qualitative moral emotions — the reactive attitudes — front and centre in the analysis of moral responsibility. Roughly, I'll suggest that, on a Strawsonian view, being a morally responsible agent requires the capacity to have a range of reactive attitudes, and that, in turn, the reactive attitudes are constitutively related to various affect-laden, phenomenally conscious experiences. Although the reactive attitudes are, of course, complex psychological and social phenomena, they have a qualitative dimension which makes it the case that phenomenal consciousness is necessary for moral responsibility, or so a Strawsonian should argue.

Part of my interest in focusing on Levy, who is not alone in arguing for the normative insignificance of phenomenal consciousness (*cf.* Lee, 2014; forthcoming), is methodological and metaphilosophical. Levy (2017, pp. 4–5; *cf.* 2014a, pp. ix, 31, 122–4, 135) promises to deliver a relatively intuition-independent method for assessing our moral responsibility in particular 'hard cases', such as those cases made salient by research on implicit bias (2017; 2014a, pp. 97–103), situationism (2014a, pp. 131–4), and sleepwalking (2014a, pp. 71–9). He aims to produce an independently plausible and empirically informed account of the necessary conditions for moral responsibility and then, in 'mechanical' fashion (2017, p. 4), to simply see whether these conditions are satisfied in the hard cases or not. While I recognize the appeal of bypassing untrustworthy intuitions when we deliberate about hard cases, I'll argue that Levy is not, as it turns out, in the casuistical clear. He is wading through the same reflective-equilibrial morass as the rest of us, because the generation of his ostensibly independently plausible theory in fact depends on a range of contested (indeed, notoriously controversial) intuitions. So the appearance of a common vantage point from which to adjudicate hard cases is illusory. Indeed, although I will not argue the point here, I suspect that many of the controversial intuitions implicated in this debate are also partly behind the intractability in debates about responsibility in various hard cases (Madva, 2018). Thus, while my narrower aim in what follows is to articulate some Strawson-inspired intuitions and arguments that speak against Levy's premises, my broader aim is to showcase the difficulty of avoiding circularity or 'bottoming out' in appeals to intuition in debates about the normative

significance of consciousness. Levy's *modus ponens* is a Strawsonian's *modus tollens*: the former argues that phenomenal consciousness doesn't make a difference, so it doesn't matter; the latter will argue that phenomenal consciousness matters, so it makes a difference. In these metaphilosophical respects, I am actually of like mind with, for example, Geoff Lee, who also argues for the normative insignificance of phenomenal consciousness, yet nevertheless shares my concern that 'it could be harder than it looks to find a non-question-begging starting point' in these debates (forthcoming, p. 15 of preprint).

Levy's argument appeals to philosophical zombies, who are functionally identical to 'normal'[1] human beings but lack phenomenal consciousness. The conceivability of such entities is controversial, and so, accordingly, is their relevance for philosophical enquiry into topics like moral responsibility. Indeed, one might wonder, given such divergent starting points, presuppositions, and methods, why Strawsonians should care about adjudicating the status of such fictional entities at all.[2] I have a few responses here. First, even if zombies *per se* are not intrinsically worthwhile objects of metaphysical or moral reflection, phenomenal consciousness *is* (see also, for example, Siewert, 1998, §9.4). Take, for example, Jeanine Weekes Schroer's (2015) and Meena Krishnamurthy's (2017) arguments that racial oppression has a distinctive, qualitative 'what it's like' character. If they are right, then it is hard for white people to understand the oppression of people of colour in part because they have not, and perhaps cannot, experience what it's like to be oppressed by virtue of being racialized and socially positioned as black, or to experience what it's like for individuals of visibly Asian or Latin American descent to be assumed to be a foreigner, and so on. Perhaps American civilians cannot know what it's like for wartime soldiers to undergo PTSD, or what it's like for Iraqi citizens to hear drones regularly buzzing over their heads. Such examples highlight the potentially profound social-epistemic and moral importance of qualitative

1 The very idea of 'normal' human beings is obscure and arguably ableist, hence the scare quotes, but I won't address these problematic implications here. I briefly discuss neuroatypical persons in §5.

2 Thanks to an anonymous referee for urging me to motivate Strawsonians to reflect on zombies. Strawson himself would likely have denied both the conceivability and possibility of zombies (see e.g. 1991, §3.5). Thanks to Haoying Liu for bringing this passage to my attention.

conscious experiences. In this light, zombies may represent a useful, catch-all conceptual device for bringing into broader relief the salience of phenomenal consciousness, and facilitating finer-grained investigation into the relevance of various experiences to numerous epistemic, moral, and political projects, including questions of the sympathy and solidarity we owe to those whose experiences differ dramatically from our own, and the general epistemic value of diversity (Gasdaglis, Kim and Madva, in preparation).

Second, as robots become increasingly autonomous, Strawsonians will, like everyone else, have decisions to make about what stance to take towards them (see §5, especially note #13). Should we praise, blame, and hold autonomous robots responsible? Should we treat them as mere tools? Should we develop a variant or entirely novel set of reactive attitudes for engaging with them? Nation states have, for example, begun to grant citizenship rights to robots (even before they recognize the full and equal rights of all their human citizens; Hart, 2018). Although these actions have, so far, clearly been publicity stunts, evidence suggests that we are apt to experience intense sympathetic and other reactions even to highly unsophisticated artificial entities (e.g. Vedantam, 2017), which underscores the risk of our over-attributing mental life and responsible agency to beings who are not genuine participants in the moral community. The arguments to follow, using phenomenally unconscious zombies as conceptual test dummies, imply that certain sorts of (borderline behaviourist) methods for tackling these vexed questions are flawed.

2. Levy's Argument

Let us assume for the moment that zombies are conceivable. (I will return to the controversy about their conceivability in §§3–4, but if it turns out that zombies are inconceivable, that result would suit the broader aims of this essay relatively well. Also note that Levy's argument will only ask us to conceive of zombies who are *functionally*, rather than *physically*, identical to us; *prima facie*, the former seems easier to conceive than the latter.) Levy argues that:

> [S]ince zombies are functional duplicates of us, there is nothing we can do that they can't. They are able to perform morally significant actions just as we are. They are able to do so after due deliberation. They are able to exercise control over their actions. Indeed, they seem capable of fulfilling almost any proposed sufficient conditions of moral responsibility. Since this seems… to be the case, it also seems as though it

> cannot be phenomenal consciousness that is required for moral responsibility. (2014a, p. 28; see also 2014b)

Levy concludes that access consciousness is the only type of consciousness necessary for moral responsibility. The part of his argument of interest to me here seems to be the following:

(1) Zombies lack phenomenal consciousness.
(2) Zombies can do everything we can do.
(3) We can do whatever is sufficient for moral responsibility [e.g. we can deliberate over reasons for action, exert self-control over our inclinations, etc.].
(4) Therefore, zombies can do whatever is sufficient for moral responsibility. (2 and 3)
(5) Therefore, phenomenal consciousness is not necessary for moral responsibility. (1 and 4)

A few points of clarification about this argument are in order. First, strictly speaking, Levy is only arguing here that the debate about the necessity of consciousness for moral responsibility is a debate specifically about access consciousness: 'what is at issue in debates over moral responsibility is whether agents must have a certain kind of access to a certain kind of content in order to be morally responsible' (2014a, p. 28n5). The rest of Levy's book is dedicated to defending the necessity of access consciousness to moral responsibility, against those who argue that consciousness is not universally necessary for moral responsibility (e.g. Adams, 1985; Smith, 2005). I will ignore this complication going forward. Second, premise (3) assumes that adults are sometimes morally responsible: that we actually have certain properties or capacities that constitute sufficient conditions for moral responsibility for at least some of our actions. Levy remains neutral about what such sufficient conditions might be. Of course, global sceptics about moral responsibility would not accept (3). In fact, Levy (2011) himself *is* such a sceptic, although for reasons related to luck rather than consciousness. The fundamental question here is what the conditions for moral responsibility consist in, rather than whether those conditions are ever actually met. Third, following Levy, this reconstruction suggests that the conditions for moral responsibility consist entirely in 'what we can do'. Given this locution, is Levy presupposing that the conditions for moral responsibility consist solely in capacities for action? Almost certainly not: he argues that access consciousness is necessary for moral responsibility, and access consciousness is not an action. Accessing mental contents

need not be a voluntary choice. However, if Levy is making this assumption, his argument is open to objections straightforwardly analogous to those I raise in what follows (e.g. Reader, 2007). I'll also say more about this 'what we can do' locution in the next section.

3. Phenomenal Consciousness and Moral Responsibility

Levy's argument seems to be question-begging. All the work is imported into premise (2), that zombies can do everything that we can do. Numerous, relatively independent lines of reasoning from ethics, epistemology, and perhaps even neuroscience suggest that phenomenal consciousness is necessary for moral responsibility, and for a variety of other moral and epistemic achievements. Since the distinctively *epistemic* significance of phenomenal consciousness has already received philosophical attention (e.g. Lee, forthcoming; 2014; Pitt, 2004; Smithies, 2012; 2013), I focus in what follows on *moral* considerations inspired by Strawson. I will circle back to epistemic and empirical considerations in the concluding section (§5). On a Strawsonian approach, moral responsibility essentially involves the capacity to feel a range of moral emotions — the reactive attitudes — and to be concerned with the emotional experiences and reactions of others. The reactive attitudes are (partly) affect-laden experiences. They do not consist solely in cold, cognitive moral judgments, but constitutively involve experiences with distinctive qualitative characters, e.g. 'what it's like' to feel smouldering resentment when someone expresses ill will toward you, or to suffer the sting of another's blame, or to feel the glow of another's praise.

Why take this Strawsonian tack?[3] Strawson's organizing aim was to respond to 'pessimists' who believed that the truth of determinism (according to which all events — including all human actions — are exhaustively caused by prior events) would make moral responsibility senseless. Strawson took exception with the pessimists' framing of the debate (a framing which he claimed was also shared by many of their opponents), accusing them of trying to 'step outside' human moral life to evaluate, as if in one go, the normative legitimacy of all of its associated practices and lived experiences. Instead, Strawson invited us to investigate these questions 'from the inside-out': what are the

[3] Thanks to an anonymous referee for recommending that I motivate the Strawsonian approach.

particular sorts of contexts and ways in which we find it appropriate to hold each other responsible? What are the concrete conditions in which we find it appropriate to suspend such practices? With some plausible gestures towards answering these questions in view, Strawson then concluded that abstract physical and metaphysical concerns about the causal structure of the universe could not undermine our practices of responsibility. I am not concerned here with the success of Strawsonian approaches as responses to the determinist threat (see, for example, Fischer and Ravizza, 2000; Russell, 1992), but rather with what Strawson 'found' when he approached these questions via his alternative method, namely, a variety of reactive attitudes. Rather than attempting to ground the normative legitimacy of our practices of holding each other responsible in abstract theories about the causes of human behaviour, Strawsonians argue that the legitimacy of these practices depend on, or even consist in, our everyday understanding and lived experiences of reactions like praise, gratitude, resentment, and blame (this is of course not to say that Strawsonians must deny the possibility of *rationally revising* our everyday understandings, lived experiences, and moral practices, but that such revisions are necessarily piecemeal and holistic; we revise some localized judgments and feelings in light of other localized judgments and feelings).

Thus, if you are a Strawsonian of almost any stripe, then you are likely committed to (a) affirming that phenomenal consciousness is essential to moral responsibility and therefore (b) denying that zombies, despite their putative capacities for deliberation and self-control, are capable of moral responsibility. Even if zombies have perfect cognitive access to their perceptions, memories, and reasons for action, *ex hypothesi* they do not actually feel gratitude, resentment, etc. Nor, for that matter, do they even feel pleasure or pain. Zombies seem to respond appropriately to others' expressions of approval and disapproval, but they do not actually care about the quality of others' wills. They lack qualitatively good or ill wills for others to care about. When a zombie shouts, 'I'll never forgive you in a million years!', she is not really feeling rage or resentment. Such utterances are 'full of sound and fury, signifying nothing' (Shakespeare 5.5, e.g. 2003). It is unclear, therefore, that such affectless utterances could constitute genuine acts of blame. There is, it seems, much we can do that they can't. Strictly speaking, zombies cannot perform morally significant actions at all — or so a Strawsonian would argue.

There are several distinctive ways in which lacking phenomenal consciousness might render zombies ineligible for moral responsibility. A first pass at taxonomizing these ways might distinguish between, on the one hand, zombies *qua targets* of others' reactive attitudes or moral concern, where the questions include whether it would be appropriate or possible to judge or hold zombies responsible; and, on the other hand, zombies *qua agents* who express (or who in some purely cognitive and unfeeling sense 'have') the reactive attitudes towards others, where the questions include whether it would be appropriate or possible for zombies to hold us responsible.

Thinking of zombies *qua* targets, it strikes me as inappropriate, absurd, and bordering on psychologically impossible to wholeheartedly praise or blame an entity (that I know with certainty to be) incapable of experiencing anything pleasurable or painful in response, let alone experiencing richer emotional responses like guilt, pride, or resentment. My scepticism here is specifically about appropriate, genuine, and 'wholehearted' praise or blame. I can, of course, easily imagine saying, in a congratulatory tone of voice, 'good job, zombie! I'm so proud of you', perhaps in the hopes of encouraging similarly prosocial behaviour from the zombie in the future. (In fact, I often succumb to praise- or blame-like outbursts towards much less sophisticated non-conscious tools and machines; §5.) However, if I know with apodictic certainty (and I will say more about the importance of certainty in §4) that the target of my expression is in-principle incapable of feeling anything in response, then it seems that my praise-like action would either not be appropriate *qua* praise, or it would not be wholehearted praise, which is to say that I would not seriously understand myself to be engaging in praise, or it would not really count as praise at all. So it seems to me. But if it did somehow count as a genuine act of praise, or if I could wholeheartedly praise an entity that I knew felt nothing in response, it would nevertheless remain difficult to see how this praise could be *deserved*. How could an individual be owed praise despite being incapable of feeling what it's like to be praised in any meaningful sense?

For zombies *qua* agents, it strikes me as equally absurd to suppose that they could have moral standing to, say, express gratitude or resentment toward others. They are, after all, incapable of experiencing these moral emotions, which makes it difficult to understand how they could be properly described as 'expressing' them at all, before we even raise the question whether it could be appropriate to do so. Admittedly, zombies' assertions that such-and-such behaviour

was praiseworthy or blameworthy would be just as reliable at tracking praiseworthy and blameworthy behaviour as are our expressions of praise and blame. There is, then, a limited sense in which these utterances would be appropriate; namely, in that they would be likely (precisely as likely as our own utterances) to be responses to normatively significant actions. So much for a purely reliabilist account of being an agent with standing to praise or blame morally significant actions! If the individuals who are ostensibly expressing gratitude or resentment cannot feel gratitude or resentment, and cannot care about the quality of others' wills, and do not even have qualitatively good or ill wills for others to care about, then they lack the standing to hold others, or themselves, morally responsible. Or so it seems to me.

Of course, Strawsonian intuitions about zombies' ineligibility *qua* targets of the reactive attitudes and *qua* agents who express or harbour reactive attitudes are clearly related: it seems particularly repugnant to think that zombies could properly 'dish it' if they cannot 'take it', i.e. that they could legitimately occupy the stance of a participant in our practices of holding (and being held) responsible if they are in-principle incapable of being on the experiential receiving end of others' reactive attitudes. In what follows, I will urge that, from a Strawsonian perspective, being a genuine participant in the moral community constitutively involves both the capacities to be an agent and a target of the reactive attitudes, to qualitatively care about the quality of good or ill will, and to have a qualitatively good or ill will for others to care about. I will introduce two thought experiments specifically to support the case that phenomenal consciousness is necessary for these Strawsonian dimensions of moral responsibility.

That Levy is so quick to dismiss phenomenal consciousness is striking because he elsewhere writes that, 'Like most theorists of moral responsibility, I am concerned with a notion of responsibility that is constitutively linked to the appropriateness of the reactive attitudes…' (Levy, 2017).[4] Again, it is difficult to see how, say,

[4] He continues: 'I am skeptical that any conception of moral responsibility that divorces it from the reactive attitudes concerns anything that is genuinely similar enough to the kind of moral responsibility at issue here to perspicuously be referred to by the same label' (Levy, 2017). Levy would likely have to deny that the reactive attitudes *constitutively* involve phenomenal consciousness. For an argument to this effect about reactive attitudes and the moral responsibility of groups, see Tollefsen (2003, pp. 231ff.). See §5 (especially note #13) for further discussion. I take the thought experiments introduced in this section to support the constitutive claim.

scolding a zombie could be appropriate — or even intelligible as a genuine act of blame — given zombies' in-principle incapacity to experience any affective responses, like shame or indignation, as a result of being scolded; and it is, for the same reasons, equally difficult to see how it could be appropriate for a zombie to scold us. Since Levy countenances a conceptual link between the reactive attitudes and moral responsibility, and since the reactive attitudes plausibly involve phenomenal consciousness, how could Levy overlook the potential significance of phenomenal consciousness to moral responsibility?

Perhaps Levy is hostile to phenomenal consciousness on anti-dualist grounds? He may take a deflationary or eliminativist view of phenomenal consciousness because he takes it to fit uneasily within scientific theories of the world. But concerns about dualism are a red herring here. The question is whether phenomenal consciousness matters for moral responsibility, regardless what further investigation reveals about the underlying nature of phenomenal consciousness.[5] In this paper, I intend to make as few assumptions as possible about the metaphysics of phenomenal consciousness. For example, perhaps it is (empirically or conceptually) necessary that any entity as functionally complex, embodied, and environmentally and socially situated as a 'normal' person is capable of phenomenal consciousness, in which case: so much the worse for the possibility of zombies, and so much the better for the necessity of phenomenal consciousness to moral responsibility.[6]

5 For gestures toward naturalistic reductions of phenomenal consciousness and affective valence, see e.g. Carruthers (2017) and references therein.

6 Lee (2014; forthcoming) argues that the mental states of zombies would be equally epistemically and morally significant as our own, on grounds of reductive materialism, and the claim that there are no natural 'joints' or 'deep divides' between human consciousness and the quasi-conscious states of zombies. Engaging fully with Lee's naturalistic arguments would take this essay too far afield, but two points bear remarking. First, as I mentioned in the introduction, Lee agrees with me about the pervasive circularity in these debates. The thrust of his arguments is typically to highlight that defenders of the significance of consciousness will be unable to cite independent (mutually agreed upon) reasons for their views and so must treat the significance of consciousness as somehow primitive or non-reducible, which Lee couples with naturalistic protests that such primitivism is implausible (he ultimately claims only to have shifted the burden of proof — Lee, 2014, p. 243). Second, I actually share Lee's scepticism about sharp natural joints 'around' human phenomenal consciousness, but thereafter we part ways: he infers that it is materially possible for indefinitely many 'nearby' cognitive systems to instantiate all the normatively significant cognitive functions without being conscious, whereas I come to doubt that systems

Another possibility is that Levy is thinking of phenomenal consciousness in unduly narrow terms, as merely perceptual (almost epiphenomenal) properties. He writes:

> An agent is phenomenally conscious of something (a taste, a sensation, a sound) when their mental state has… a qualitative character: the apparently ineffable qualities we feel when we perceive colors, or taste wine, or hear the soft pattering of rain… Why does the redness of a ripe tomato look like *that* and not, say, like the blue of a late afternoon sky (or, for that matter, like the ringing of a church bell)? (2014a, p. 27)

It may, then, simply not have occurred to Levy that phenomenal consciousness also includes a host of morally relevant affective experiences (although see Levy, 2014b).

What if we were to ask, by comparison, why does the sting of blame feel like *that* instead of like the glow of praise? Consider, in this vein, two further thought experiments. First:

> INVERSION: Vera has 'inverted' moral qualia. She is functionally identical to 'normal' people and, *qua* target of the reactive attitudes, she responds in typical ways to praise and blame, but she actually feels the glow of praise (i.e. a pleasant, positively valenced affective state) when she is blamed, and she feels the sting of blame (an unpleasant, negatively valenced state) when she is praised. *Qua* agent, Vera outwardly seems to praise and blame others in typical ways, but her acts of blame *feel to her* like acts of praise and her acts of praise feel like blame.

In so far as this case is conceivable (and I will discuss reasons to question its conceivability in what follows), it brings into sharp relief the moral relevance of affective phenomenal consciousness. While we would feel sympathy toward Vera's tragic moral-psychological plight, and should try to help uncross her emotional wires if possible, we

could approximate these functions without being, *at least to some degree*, phenomenally conscious. It is increasingly clear that accounts of consciousness and mentality must be *graded* in numerous ways (see also Madva, 2018), and Lee also notes that 'the presence of consciousness presumably depends on the presence of a number of continuously variable physical magnitudes, meaning that the location of any sharp boundary for consciousness will be highly arbitrary' (forthcoming, p. 12 of preprint). Good riddance to sharp boundaries, but assuming that we 'normal' waking human adults are *far* from the multifarious indeterminate frontiers between consciousness and non-consciousness, I find Lee's grounds for confidence about the possibility of zombies (who are putatively *near* to us with respect to cognitive function yet *radically remote* from us with respect to consciousness) obscure.

would not view her as morally responsible in the same way or to the same extent as a person who feels a more 'normal' range of emotional responses. My intuition is that her experience of others' blame as if it were praise significantly diminishes her responsibility for acting in blameworthy ways. Similarly, how could Vera be *just as responsible as we are* for expressing gratitude despite the fact that doing so arouses in her the painful, upsetting, or otherwise distracting experiences of indignation? Minimally, this case suggests that the sheer fact that, at a certain level of description, she functions like we do should not immediately settle whether she meets any conditions for moral responsibility one might reasonably propose. Of course, I have not asserted that Vera bears no responsibility whatsoever for how she acts (Vera is, after all, *not* a zombie completely devoid of phenomenal consciousness), but only that her radically divergent phenomenal experiences affect the nature or extent of her responsibility, and, in turn, the range of reactive attitudes it would be appropriate to take toward her. Intuitively, it would seem appropriate to resent her less for her ethical failures and admire her more for her ethical successes, and, more generally, to adopt a sympathetic stance as she perseveres through her predicament. In other words, Vera's case minimally suggests that phenomenal consciousness makes a difference to moral responsibility. Levy's second premise, that nothing we can do depends on phenomenal consciousness, seems at worst false and at best non-obvious in Vera's case. It is not clear that she can wholeheartedly engage in practices of praise and blame in the same ways that we can.

Vera's case may be difficult to genuinely conceive, however. It seems more difficult to wrap our heads around than the standard case of inverted colour experience because here *valence* has been inverted, and valence is, intuitively speaking, more closely tied to action (*ceteris paribus*, we pursue what we like and avoid what we dislike) than are phenomenal experiences of colour.[7] Indeed, how could someone with inverted moral qualia even develop into a functioning moral agent? A virtue-ethicist or sentimentalist might emphasize that Vera cannot cultivate virtue because she is incapable of taking the right sort of pleasure in doing the right thing. Even Kant, though often cited for downplaying the importance of feeling to moral agency, argued that the capacities for various moral feelings were preconditions for being

7 Thanks to Peter Ross for discussion about this point.

susceptible to the moral law (Denis and Wilson, 2016; Gasdaglis, 2019).

This brings us to a key distinction — which Levy may simply overlook — between the conditions on being a morally responsible agent *at all* and the conditions on being morally responsible *for some particular action or omission* (see, for example, Wallace, 1994, p. 84). Plausibly, both phenomenal and access consciousness are necessary for responsible agency in general. Both the capacity to access the contents of one's mind and the capacity to qualitatively experience at least some range of feelings and reactive attitudes are likely necessary for being capable of moral responsibility.[8] On this line, zombies would not even be candidates for moral responsibility, and the candidacy of individuals like Vera for full responsibility would be significantly compromised. That said, phenomenal and access consciousness arguably play different roles in moral responsibility. An important project for future research is articulating the various roles that different sorts of consciousness play in constituting responsible agency. Different moral theories will likely spell this out in different ways.

However, Levy might be on firmer ground were he to argue that phenomenal consciousness is not necessary for moral responsibility *in all particular cases*. It seems unlikely that, for every action, there exists some particular feeling that one must experience in order to merit praise or blame for that action. I suspect it depends on the action in question. To see how particular feelings might matter in at least some particular cases (and to thereby cast further doubt on Levy's claim that nothing that we can do depends on phenomenal consciousness), consider the following individual who, in comparison to zombies, more plausibly satisfies the general background conditions of responsible agency:

> BOUTS OF INZOMBIA: Zed is an otherwise 'normal' person who suffers from temporary bouts of 'inzombia': brief periods where everything goes 'dark inside', but he continues to act fully like

[8] Thus, even theorists who deny that awareness is necessary for moral responsibility in *particular cases* (Adams, 1985; Smith, 2005) might accept a different necessity claim: that capacities for phenomenal and access consciousness are preconditions for being the sort of individual to whom moral responsibility could ever be appropriately assigned. However, I agree with these theorists (against Levy) that access consciousness is not necessary for moral responsibility in all particular cases.

himself. During such bouts, Zed would vehemently deny that he was in a zombie state if you asked him. Afterwards, he retains propositional memories of what happened, but the memories are entirely 'numb' and non-episodic: they lack any perceptual or affective character.

ZED & FRIENDS: now imagine that you are very close to Zed and have an intense, long-awaited, moral-emotional exchange with him. Zed's eyes well with tears and his voice quivers as he says, 'I can't tell you how much I admire you and how grateful I am to count you as a friend' (or 'After all these years, I'm finally ready to forgive you', or 'Words can never express how deeply sorry I am'). Later, however, you discover that Zed was in a zombie state during the interaction.[9]

Once you learn that Zed was 'blacked out' in this way, will you think the interaction retains the full moral significance it seemed to have? Might you feel cheated in any way? Might you perhaps want a 'do-over' of the conversation to make sure that Zed *really felt* the feelings you thought he was expressing? Knowing that Zed did not experience the affective states associated with the relevant reactive attitudes influences our moral assessment of his behaviour, and our sense of which reactive attitudes it would, in turn, be appropriate to take towards him. If Zed didn't feel deeply sorry when he apologized, does he still deserve gratitude for apologizing? Is forgiveness still an appropriate response? Can we even say that there was an apology if he literally felt nothing when he gave it? It strikes me as at least open to doubt whether Zed can be properly described as giving an apology if everything was dark inside when he made the apologetic-sounding remarks. I have, therefore, no confidence that Zed can 'do everything we can do' while in a zombie state; whatever he's doing when he says 'I'm sorry', he is not obviously apologizing — or so a Strawsonian would argue. More generally, the cases of Zed and Vera undermine

[9] Note also that we can trivially modify Zed's social location in these examples to generate correlative intuitions about the importance of phenomenal consciousness to being the target or agent of the reactive attitudes, i.e. regarding which reactive attitudes it is appropriate to hold towards Zed and which reactive attitudes Zed can appropriately hold towards others, or, for that matter, towards himself. To this end, we can even imagine ourselves in the positions of Zed or Vera, or in the positions of those close to them, such as a friend who learns that Zed, say, finally got up the 'courage' to propose to his fiancée while in a zombie state, and so on.

the intuition that merely functionally equivalent cognitive-behavioural processes suffice for moral responsibility.

Most of the 'folk' seem to agree that individuals who systematically lacked the capacity for these qualitative experiences could not be morally responsible (Shepherd, 2015). When asked to imagine humanoid machines that acted just like human beings but lacked all conscious sensations (including pain, emotion, etc.), participants tended to agree that such individuals were conceivable, but that they would lack free will and moral responsibility for acting badly. By contrast, they believed that humanoid machines with phenomenal consciousness would bear moral responsibility. In fact, there are rapidly expanding experimental-philosophical literatures on consciousness, responsibility, and their interconnections, which I cannot fully explore here (Goodwin, 2015).[10] One clear upshot from this research, however, is that the capacity for phenomenal consciousness plays a powerful role in the folk's judgments about how individuals ought to be treated. We feel as though individuals entirely lacking phenomenal consciousness have no 'skin in the game' of morality. There is nothing 'at stake' for zombies, which prevents them from being genuine participants in the moral community.

4. Sympathy for the Zombie?

Of course, if actually faced with zombies functionally identical to 'normal' human beings, it would be extremely difficult to withhold our ordinary affect-laden reactions toward them. It would, in fact, be extremely difficult to believe (and perhaps even to conceive) that they were zombies. They would seem to be ordinary participants in the moral community, and they would presumably act outraged, distraught, or at least perplexed by the suggestion that they lacked inner mental lives. I suspect that we would (or at least should; Antony, 1996; Tanney, 2004) sooner doubt whichever scientist or authority figure told us they lacked phenomenal consciousness than we would withhold our sympathy or resentment towards them (this is effectively the plot of countless tales of science fiction and fantasy: a non-human

[10] See Sytsma and Machery (2012) for studies ostensibly suggesting that participants sometimes judge that individuals with sophisticated cognitive capacities but impoverished experiential capacities deserve significant moral consideration, but I agree with Jack and Robbins (2012, p. 402) that Sytsma and Machery's cases do not involve the total absence of phenomenal consciousness.

entity — robot, extraterrestrial alien, animal, plant, or even an ecosystem or planet — displays evidence of feeling, intelligence, and reactive attitudes, such that the perceptive, empathic protagonists of the story fight to defend the interests and rights of the entity, while the villains show the entity a callous disregard).[11] In other words, in so far as we would feel wholeheartedly compelled to praise or resent a zombie, I predict that to *just that extent* we would also find it difficult to wrap our heads around the idea that she was really a zombie. Seriously imagining ourselves in an interpersonal relationship with such an individual naturally involves imagining that we care about the quality of her will and that she cares about ours, and thereby contributes to the difficulty of conceiving that such functionally identical individuals could entirely lack phenomenal consciousness.

But what if we stipulate that we could establish, in some way we all agree to be conclusive (*cf.* Putnam, 1963), that certain individuals really were zombies? In that case, I think we would — and should — shift from the 'participant' stance to the 'objective' stance toward these individuals. We would — and should — cease to think of them as genuine participants in the moral community, whom we might wholeheartedly resent, and shift towards seeing them more as 'objects of social policy… to be managed' (Strawson, 1962). It might remain useful to keep praising and blaming them, to keep them in line; the traditional consequentialist defence of moral responsibility might still apply. We might also *let ourselves* become emotionally engaged with zombies, much as we become emotionally engaged with fictional characters (*cf.* Gendler, 2013, §5.3), but such engagement would be exclusively for our sake, not the zombies', i.e. not owed to them by the requirements of morality.

5. What the Zombie Didn't Know

My intuition is that affectless, non-conscious agents, like zombies, would not only fail to be morally responsible: they would lack the moral status we recognize in far less cognitively sophisticated

[11] These sorts of stories are, moreover, not just the stuff of science fiction: human history is littered with examples of colonialists, war-mongering demagogues, and medical professionals making unjust efforts to deny that various human beings were capable of feeling the full range of moral emotions.

individuals.[12] To repurpose Bentham's (1789) famous exhortation about non-human animals, the question is not *merely* 'Can they reason? nor, Can they talk? But [*also*], Can they suffer?' The central question here is not, as is sometimes debated about other animals, whether the capacity for phenomenal consciousness is sufficient for moral status, or whether some more exalted cognitive capacities are also necessary. The question here revolves around entities that ostensibly have higher-order, rational capacities but lack experiential capacities (including 'lower-order' capacities for pain and pleasure and 'higher-order' moral-emotional capacities, e.g. to react to praise by feeling valued, or, for that matter, to react to praise by feeling embarrassed). Are the experiential capacities *necessary* for the rational capacities, or necessary for the rational capacities to bear the moral significance often associated with them?[13]

The conceivability of zombies might seem to entail the conceivability of possessing rational capacities without possessing experiential capacities, but this inference would be too quick. Since zombies are entirely unacquainted with affective experience, there is reason to doubt that they can properly be said to understand others' feelings, or their moral significance (and how could a zombie be blameworthy for hurting my feelings if it doesn't understand what feelings are?). Can zombies, per Levy's stipulation, actually exercise self-control over their actions? What would be the nature or force of the 'inclinations' they'd have to resist? (And how could a zombie be praiseworthy for

[12] It might be the case that zombies have some modicum of moral status (i.e. patiency, standing, considerability, etc.). For example, if zombies are living organisms, and if, as some argue, all living organisms deserve some moral consideration, then zombies deserve some moral consideration. It might, then, be about as intrinsically wrong to behead a zombie as it would be to chop down a tree. We would, I hope, nevertheless prohibit zombie 'abuse' because, as Kant says about the cruel treatment of animals, doing so would cultivate vicious traits. See also Siewert (1998, p. 364n4).

[13] Two other types of entity with ostensible possession of rational but not experiential capacities are autonomous robots and group agents. See Sparrow (2007, pp. 71–2) for insightful treatment of robots' potential for moral responsibility. Regarding group responsibility, I believe I can remain neutral about several of the core debates here, such as whether group responsibility is reducible to the responsibility of its members. For example, one could maintain that a group's responsibility depends in part on the experiential capacities of its members without taking a stand on whether the group's responsibility reduces entirely to features of its individual members. The weaker dependency claim would entail that a group comprised solely of zombies could not bear responsibility in the same way as regular groups, which strikes me as plausible. Alternatively, one might argue that groups can legitimately experience at least some moral emotions (Schmid, 2014; *cf.* Sosa, 2009; Tollefsen, 2003).

overcoming a temptation if it doesn't actually feel tempted or understand what it's like to feel and resist temptations?) Finally, can zombies even understand what it means to be a 'reason', moral or otherwise, if they don't know what it's like to take something as a reason? These questions — whether zombies can understand the moral significance of others' or their own mental states, or what it means to be a reason — represent but a few of the considerations revolving around the epistemic significance of phenomenal consciousness. In this vein, Declan Smithies has argued that phenomenal consciousness grounds all epistemic rationality.[14] Take introspective knowledge, for example:

> I know by introspection whether I am feeling pain or pleasure and whether I am visually experiencing red or green… the phenomenal character of my experience explains how I know these things. There is a phenomenal difference between feeling pain and pleasure and… between visually experiencing red and green. Intuitively, it is because of these phenomenal differences that I can know by introspection whether I feel pain or pleasure and whether I am visually experiencing red or green. (Smithies, 2013, p. 734)

The capacities to access, integrate, and act upon information thus seem relevant to, but insufficient for, normative statuses like self-knowledge and moral responsibility. Computers, for example, far outstrip human minds in terms of the capacities to access and integrate (certain sorts of) information, but we don't typically think that computers are simply thereby 'conscious' of this information in epistemically or morally relevant senses (e.g. such that it would be appropriate to praise them for accessing data). As far as human minds go, our rational capacities to access and integrate content are thoroughly dependent on affect. In fact, Levy elsewhere writes at length about how neuroscience suggests that 'affect is indispensable to rationality' (Levy, 2009, p. 76; *cf.* 2007, pp. 80–1, 112, 116–20, 187–95, 293–308). Citing Antonio Damasio's research, he contests the assumption that 'reason' and 'emotion' are inherently opposed:

> [R]ather than emotions crowding out reasoning, they might partially *constitute* it… even when we have time to deliberate, we cannot dispense with affect. It makes options salient for us, helping thereby to solve the problem of combinatorial explosion which faces any pure calculating machine. (Levy, 2009, p. 76)

[14] See Smithies (2012; 2013) and the references therein.

Levy (2015, pp. 66–8) further elaborated this view in this journal, in response to Sripada (2015), describing 'valenced signals' as having the dual functional roles of trimming the 'search space' for action options and 'biasing deliberation between options that remain'. Levy notes that a given 'signal may be more or less strong, and either positive or negative. These valenced signals might be experienced as "gut feelings", hunches, intuitions, or affects. As a consequence, the person will feel better disposed toward some options than others' (Levy, 2015, p. 66).

This brings us to neuroscientific and other empirical considerations, although I can only scratch the surface of this rich and admittedly thorny area.[15] For one thing, it is increasingly clear that valence is fundamental to an even wider array of cognitive processes than Levy suggests, making it plausible that phenomenal consciousness is essential to the most basic forms of agency. Godfrey-Smith (2016, pp. 793–4) takes seriously, for example, that 'the internal processing of valence', rather than highfalutin higher-order processes, played a decisive role in the evolution of cognition and subjectivity. And some researchers have argued that 'micro-valences' of affective value are intrinsic to basic perceptual processing, such that everything we see is 'either slightly preferred or anti-preferred' (Lebrecht *et al.*, 2012, p. 2; see also Caplette *et al.*, 2014). These theorists place qualitative (albeit perhaps very *dimly felt*) experience at the very heart of cognition (Duncan and Barrett, 2007; see also Madva, 2018; Madva and Brownstein, 2018). One might also consider individuals whose metacognitive and executive dispositions (i.e. those higher-order dispositions related to access consciousness that Levy makes central to responsibility) are neuroatypical, but whose affective dispositions are relatively neurotypical, and who are clearly morally responsible agents (Pickard, 2015; Stout, 2017; Richman, 2018). Or one might consider the extreme challenges facing individuals who are congenitally insensitive or indifferent to pain, and extrapolate from there to the prospects for a hypothetical agent altogether insensitive to affect and valence (e.g. Carruthers, 2017, §4). However, cases like insensitivity (or indifference) to pain illustrate just how thorny these matters are, because these individuals lack both phenomenal *and* access consciousness of pain (or of pain's unpleasant, negative valence).

15 Thanks to an anonymous referee for pressing me to delve further into empirical considerations.

And while the evidence clearly ties qualitative affective experiences to certain computational functions, for Levy to acknowledge that valenced signals *might* be experienced in certain ways is clearly not to say that they are *necessarily* experienced, and I am not trying to saddle him with such claims. The question, however, arises: would these signals retain their full moral and epistemic significance if they were completely divorced from phenomenal consciousness, i.e. if they were housed in a cognitive system that was in-principle incapable of feeling praised, blamed, pleasured, or pained? Consider the question whether zombies can be accurately described as exercising self-control by overcoming temptations. While I (think I) can imagine that zombies have *functional analogues* to the affective-motivational pushes and pulls of reasons, inclinations, intuitions, and states of feeling 'better disposed toward some actions than others', I cannot imagine that these analogues figure in genuine exercises of rational — or otherwise normatively significant — capacities in so far as they're not even potentially felt. To cast this point in a more Strawsonian idiom, I cannot stably occupy the participant stance toward entities whom, per stipulation, completely lack phenomenal consciousness. I drift inevitably toward doubting that they are truly unconscious, or toward perceiving them as purely hydraulic systems, who no more deserve credit for what they do than ketchup deserves credit for impelling itself out of the bottle, as gravity 'struggles' to overcome friction.

Again, contemporary computer programs (e.g. search engines, news feeds) excel precisely at narrowing down the search space to the most 'relevant' options, biasing our deliberation among remaining options, revising future suggestions in light of our prior behaviour and feedback, etc., but no one seriously thinks that these programs merit full-throated gratitude or resentment *simply by virtue* of fulfilling or failing to fulfil these specific functions. We praise their designers, we get frustrated at their glitches, we occasionally slip into brief emotional outbursts at the devices themselves, and we give innumerable 'thumbs up or down' for the sake of training up all the algorithms of modern life, but none of this amounts to treating the hardware or software as full-blooded participants in the moral community. Yet as mountains of science fiction and fantasy can attest, once we start envisioning that these systems actually feel pleased by our positive feedback, crest-fallen by our censure, or heartbroken by our indifference, the conceptual space for holding them responsible and being held responsible by them suddenly seems to reopen. Thus, even if we were to stipulate

that Levy and others have accurately identified some or all of the functional roles of affective experience, and adequately explained how the processes that fulfil these roles are essential to epistemically rational cognition and ethically praiseworthy action, then all we would thereby grant is that fulfilling these roles is necessary for various normative accomplishments. It would not follow that fulfilling these roles is *sufficient* for those accomplishments, and it is frankly difficult to envision ways of making the case for sufficiency without begging the core question or bottoming out in appeals to brute intuition. While Levy argues that phenomenal consciousness cannot matter because it does not make a difference, any strategy along these lines will have little force against someone who thinks that phenomenal consciousness must make a difference because it obviously matters.

Of course, the contrary, Strawson-inspired position to which I have tried to give voice is not above appeals to intuition, either. Thus, while the narrower aim of this paper has been to cast doubt on the possibility of morally responsible zombies, the broader aim is to cast doubt on the possibility of non-question-begging or intuition-independent ways of adjudicating such debates. Acknowledging as much may, however, ultimately be in the spirit of Strawson's overarching approach to responsibility, which sought to show the irrelevance or incoherence of trying to settle such questions from a presuppositionless Archimedean point external to our practices, and urged us instead to wade through the mud of taking seriously our ground-level intuitions, emotions, and lived experiences as they are before presuming to determine how (and whether) they ought to be.

Acknowledgments

For extensive guidance on early drafts and for ongoing discussion, I am especially indebted to Katie Gasdaglis. For extremely helpful and thorough commentary on a presentation at the January 2018 meeting of the Eastern Division of the American Philosophical Association, thanks to Haoying Liu. I also benefited there from questions from Louise Antony, Robert C. Hughes, Michelle Moody-Adams, and others. I also thank several anonymous referees; Charles Michael Brent; the audience at the June 2016 meeting of the European Philosophical Society for the Study of Emotions, especially Thomas Szanto and Carme Isern Mas; and the Cal Poly Pomona students and faculty who participated in a departmental 'brown bag' presentation in November 2017, especially David Adams, Cory Aragon, Michael Cholbi, Peter Ross, Marmar Tavasol, and Dale Turner.

References

Adams, R.M. (1985) Involuntary sins, *The Philosophical Review*, **94** (1), pp. 3–31.

Antony, L. (1996) Equal rights for swamp-persons, *Mind & Language*, **11** (1), pp. 70–75.

Bentham, J. (1789) *An Introduction to the Principles of Morals and Legislation*, Oxford: Clarendon Press.

Caplette, L., West, G., Gomot, M., Gosselin, F. & Wicker, B. (2014) Affective and contextual values modulate spatial frequency use in object recognition, *Frontiers in Psychology*, **5**, art 512.

Carruthers, P. (2017) Valence and value, *Philosophy and Phenomenological Research*, **93**, pp. 658–680.

Denis, L. & Wilson, E. (2016) Kant and Hume on morality, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2016 Edition)*, [Online], http://plato.stanford.edu/archives/fall2016/entries/kant-hume-morality/.

Duncan, S. & Barrett, L.F. (2007) Affect is a form of cognition: A neurobiological analysis, *Cognition and Emotion*, **21** (6), pp. 1184–1211.

Fischer, J.M. & Ravizza, M. (2000) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Gasdaglis, K.L. (2019) Moral regret and moral feeling(s), *Inquiry*, March, pp. 1–29.

Gasdaglis, K.L., Kim, B.H. & Madva, A. (in preparation) *The Irreplaceable Epistemic Value of Diversity*.

Gendler, T.S. (2013) Imagination, in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy (Fall 2013 Edition)*, [Online], http://plato.stanford.edu/archives/fall2013/entries/imagination/.

Godfrey-Smith, P. (2016) Individuality, subjectivity, and minimal cognition, *Biology & Philosophy*, **31** (6), pp. 775–796.

Goodwin, G.P. (2015) Experimental approaches to moral standing, *Philosophy Compass*, **10** (12), pp. 914–926.

Hart, R. (2018). Saudi Arabia's robot citizen is eroding human rights, [Online], https://qz.com/1205017/saudi-arabias-robot-citizen-is-eroding-human-rights/ [4 July 2018].

Jack, A.I. & Robbins, P. (2012) The phenomenal stance revisited, *Review of Philosophy and Psychology*, **3** (3), pp. 383–403.

Krishnamurthy, M. (2017) White moral blindness, presented at the *Philosophy Speaker Series*, Eastern Michigan University, January 2012, [Online], https://vimeo.com/201868460.

Lebrecht, S., Bar, M., Barrett, L.F. & Tarr, M.J. (2012) Micro-valences: Perceiving affective valence in everyday objects, *Frontiers in Psychology*, **3**, art. 107.

Lee, G. (2014) Materialism and the epistemic significance of consciousness, in Kriegel, U. (ed.) *Current Controversies in Philosophy of Mind*, pp. 222–245, London: Routledge.

Lee, G. (forthcoming) Alien subjectivity and the importance of consciousness, in Pautz, A. & Stoljar, D. (eds.) *Themes from Block*, Cambridge, MA: MIT Press.

Levy, N. (2007) *Neuroethics: Challenges for the 21st Century*, Cambridge: Cambridge University Press.

Levy, N. (2009) Neuroethics: Ethics and the sciences of the mind, *Philosophy Compass*, **4** (1), pp. 69–81.

Levy, N. (2011) *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, Oxford: Oxford University Press.
Levy, N. (2014a) *Consciousness and Moral Responsibility*, Oxford: Oxford University Press.
Levy, N. (2014b) The value of consciousness, *Journal of Consciousness Studies*, **21** (1–2), pp. 127–138.
Levy, N. (2015) Defending the consciousness thesis: A response to Robichaud, Sripada and Caruso, *Journal of Consciousness Studies*, **22** (7–8), pp. 61–76.
Levy, N. (2017) Implicit bias and moral responsibility: Probing the data, *Philosophy and Phenomenological Research*, **94** (1), pp. 3–26.
Madva, A. (2018) Implicit bias, moods, and moral responsibility, *Pacific Philosophical Quarterly*, **99** (S1), pp. 53–78.
Madva, A. & Brownstein, M. (2018) Stereotypes, prejudice, and the taxonomy of the implicit social mind, *Noûs*, **52** (3), pp. 611–644.
Pickard, H. (2015) Psychopathology and the ability to do otherwise, *Philosophy and Phenomenological Research*, **90** (1), pp. 135–163.
Pitt, D. (2004) The phenomenology of cognition or 'what is it like to think that P?', *Philosophy and Phenomenological Research*, **69** (1), pp. 1–36.
Putnam, H. (1963) Brains and behavior, in Butler, R.J. (ed.) *Analytical Philosophy: Second Series*, Oxford: Blackwell.
Reader, S. (2007) The other side of agency, *Philosophy*, **82** (04), pp. 579–604.
Richman, K.A. (2018) Autism and moral responsibility: Executive function, reasons responsiveness, and reasons blockage, *Neuroethics*, **11** (1), pp. 23–33.
Russell, P. (1992) Strawson's way of naturalizing responsibility, *Ethics*, **102** (2), pp. 287–302.
Schmid, H.B. (2014) The feeling of being a group: Corporate emotions and collective consciousness, in von Scheve, C. & Salmela, M. (eds.) *Collective Emotions: Perspectives from Psychology, Philosophy, and Sociology*, pp. 3–22, Oxford: Oxford University Press.
Schroer, J.W. (2015) Giving them something they can feel: On the strategy of scientizing the phenomenology of race and racism, *Knowledge Cultures*, **3** (1), pp. 91–110.
Shakespeare, W. (2003) *MacBeth*, Mowat, B.A. & Werstine, P. (eds.), New York: Simon & Schuster.
Shepherd, J. (2015) Consciousness, free will, and moral responsibility: Taking the folk seriously, *Philosophical Psychology*, **28** (7), pp. 929–946.
Siewert, C. (1998) *The Significance of Consciousness*, Princeton, NJ: Princeton University Press.
Smith, A.M. (2005) Responsibility for attitudes: Activity and passivity in mental life, *Ethics*, **115** (2), pp. 236–271.
Smithies, D. (2012) The mental lives of zombies, *Philosophical Perspectives*, **26** (1), pp. 343–372.
Smithies, D. (2013) The significance of cognitive phenomenology, *Philosophy Compass*, **8** (8), pp. 731–743.
Sosa, D. (2009) What is it like to be a group?, *Social Philosophy and Policy*, **26** (01), pp. 212–226.
Sparrow, R. (2007) Killer robots, *Journal of Applied Philosophy*, **24** (1), pp. 62–77.
Sripada, C. (2015) Acting from the gut: Responsibility without awareness, *Journal of Consciousness Studies*, **22** (7–8), pp. 37–48.

Stout, N. (2017) Autism, metacognition, and the deep self, *Journal of the American Philosophical Association*, **3** (4), pp. 446–464.
Strawson, P.F. (1962) Freedom and resentment, *Proceedings of the British Academy*, **48**, pp. 1–25.
Strawson, P.F. (1991) *Individuals: An Essay in Descriptive Metaphysics*, London: Routledge.
Sytsma, J. & Machery, E. (2012) The two sources of moral standing, *Review of Philosophy and Psychology*, **3** (3), pp. 303–324.
Tanney, J. (2004) On the conceptual, psychological, and moral status of zombies, swamp-beings, and other 'behaviourally indistinguishable' creatures, *Philosophy and Phenomenological Research*, **69** (1), pp. 173–186.
Tollefsen, D.P. (2003) Participant reactive attitudes and collective responsibility, *Philosophical Explorations*, **6** (3), pp. 218–234.
Vedantam, S. (2017) Can robots teach us what it means to be human?, *Hidden Brain*, 10 July, [Online], https://www.npr.org/2017/07/10/536424647/can-robots-teach-us-what-it-means-to-be-human.
Wallace, R.J. (1994) *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.

Paper received December 2017; revised July 2018.