Book Section:
Holroyd, J.D. and Kelly, D. (2016) Implicit Bias, Character and Control. In: Masala, A. and Webber, J., (eds.) From Personality to Virtue Essays on the Philosophy of Character. Oxford University Press. ISBN: 9780191063787.

5

Implicit Bias, Character, and Control

Jules Holroyd and Daniel Kelly

Implicit biases are automatic associations, often operational without the reflective awareness of the agent, which influence action.[1] This influence can be malign—affecting negatively individuals' evaluations and judgements of, and interactions with, individuals in stereotyped or stigmatized groups (social identities such as race, gender, religious identity, age, and mental illness have all been studied). The effects may be relatively minor: implicit biases can increase the number of times one blinks one's eyes when interacting with a member of another race, which in turn can degrade the quality of those interracial interactions. Or the effects may be of grave consequence, such as increasing the likelihood of shooting a black man reaching for an ambiguous object (could be a gun, could be a wallet or mobile phone). Both kinds of effects systemically contribute to and re-entrench patterns of discrimination and marginalization.[2] In the studies that have proliferated on implicit bias, one outcome is indisputable—almost all of us harbour, and are influenced by, some kinds of implicit bias, to some degree.[3]

[1] We would like to thank the following people for useful feedback on earlier presentations and drafts of this material: Michael Brownstein, Natalia Washington, Alex Madva, the attendees of the Implicit Bias and Philosophy workshops held at the University of Sheffield, audiences at the University of Edinburgh, and the editors of this volume, Jonathan Webber and Alberto Masala, and anonymous reviewers for this volume.

[2] Amodio, Harmon-Jones, and Devine 2003; Payne 2005.

[3] For an excellent overview of the range of biases and their pervasive effects, see Jost et al. 2009.

Our focus here is on whether, when influenced by implicit biases, those behavioural dispositions should be understood as being a part of that person's character: whether they are part of the agent that can be morally evaluated.[4] We frame this issue in terms of control. If a state, process, or behaviour is not something that the agent can, in the relevant sense, control, then it is not something that counts as part of her character. A number of theorists have argued that individuals do not have control, in the relevant sense, over the operation of implicit bias. We will argue that this claim is mistaken. We articulate and develop a notion of control that individuals have with respect to implicit bias, and argue that this kind of control can ground character-based evaluation of such behavioural dispositions.

First we introduce two perspectives on implicit bias and character (section 5.1). In section 5.2 we evaluate the arguments for the conclusion that individuals lack the relevant sense of control with respect to implicit bias. In sections 5.3 and 5.4 we elaborate on one sense of control—Clark's 'ecological control'—which we argue is a sense of control that enables us to consider implicit biases as 'part of who the agent is', and hence something that is a legitimate candidate for normative evaluation. We go on (in section 5.5) to show that this requires some emendation to recent ways of thinking about the development of moral character and self-regulation.

5.1 Character, Evaluation, and Implicit Bias

Some philosophers who have written on implicit bias have homed in on the idea that such implicit associations are 'rogue' processes, which are not properly seen as part of the agent's character—not indicative of 'who she is'. For example, Jennifer Saul has suggested that we should reject the thought that 'acknowledging that one is biased means declaring oneself to be one of those bad racist or sexist people' (~~ms~~, ~~21~~). Whilst this is recommended for pragmatic reasons (namely, getting people to acknowledge that they may harbour implicit biases), we take it that the underlying assumption here is that having negative race or gender biases

[4] For other questions that have been raised in relation to implicit bias, see Gendler 2011, Holroyd 2012, Kelly and Roedder 2008, Saul ~~ms~~. Many other fruitful papers can be found here: <http://www.biasproject.org/recommended-reading>.

does not reflect badly on one's character. Merely being influenced by implicit bias does not mean that one has the character of a racist or sexist person; it takes something other than the operation of implicit racial biases to be properly ascribed the character *trait* racist. Likewise, Joshua Glasgow has emphasized that many of us are *alienated* from such attitudes, and tend to regard implicit bias and the behaviours influenced by them as 'actions and attitudes that do not represent us in any real sense at all' (ms, 7). Glasgow's remarks here resonate with the idea that implicit biases do not constitute 'who we are', or form part of our characters. Finally, Neil Levy (2012) has indicated that the associative structure of implicit biases, and the fact that they are not subject to 'rule-based-processing' in the way that our explicit attitudes are, means that such states cannot play a role in unifying us as agents and can actually undermine our ability to express 'who we are' in our actions over time.

This treatment contrasts with that of Lorraine Besser-Jones, who implies that implicit biases can legitimately be taken into consideration when evaluating a person's character. Insofar as behavioural dispositions are a part of character, and implicit biases produce certain behavioural dispositions, then those implicit biases qualify as part of an individual's moral character, for which they can be evaluated (and perhaps blamed). This emerges in her treatment of the following example (we quote at length):

> Take the case of Jane and Mark. Jane believes that everyone deserves equal treatment, and condemns racism. Yet, whenever she sees a black man walking down the street, she averts her eyes and, if possible, crosses the street. She feels guilty when she does so, but nonetheless cannot help herself from acting in these ways. Mark, on the other hand, holds racist beliefs about the inferiority of African Americans. He, too, averts his eyes and crosses the street when encountering a black man. He, too, feels guilty when he does so, after all, it is not as if he holds anything 'personal' against the man, but is simply acting on the basis of what he thinks is right.
>
> Jane and Mark share very similar dispositions—both behavioral dispositions, and dispositions to feel certain ways in response to their behavioral dispositions—nonetheless, all would agree that they have different moral characters: Jane is what psychologists label an 'aversive racist'—one who is 'consciously non-prejudiced yet unconsciously prejudiced'; her character is significantly different than Mark's, whose character is prejudiced on all levels. (2008, 317, see also discussion in her 2014, 82)

Besser-Jones uses this example to make the point that an individual's beliefs, and not simply her behavioural dispositions, are the basis of

character: the right account incorporates both, she argues. What is striking from our point of view is that Besser-Jones nonetheless sees the behavioural dispositions involved in Jane's 'aversive racism'—a form of racism underpinned by negative implicit biases—as a legitimate (partial) basis for evaluating her moral character. A full evaluation of her moral character should take into account those dispositions, and the behaviour that manifests them, contra the starting assumptions of Saul, Glasgow, and Levy.

Which view should we endorse? Both perspectives have some appeal: insofar as implicit biases are not under an individual's control and are attitudes from which she is alienated, it does seem somewhat unfair to treat them as evaluable in the same way as her considered and endorsed beliefs and attitudes. Moreover, one might think that implicit biases reflect mere epistemic mistakes rather than character flaws, and that being influenced by them can be evaluated only by tracing back to these original errors, which are not flaws of character.

But there is intuitive appeal in Besser-Jones's position also. Her emphasis is on the extent to which individuals' moral commitments and beliefs interact with and shape their behavioural dispositions (and their practical attitudes towards success or failure in doing so) (2008, 322). Moreover, we think it defensible to consider the operation of implicit bias in terms of character, since (as we will see) the operation of implicit biases interact with agents' values in various ways. And, even if an epistemic error is at issue, this does not render questions of character irrelevant: the mistakes individuals make are often reflective of who they are (their values, with respect to what they are scrupulous, say).

The operation of implicit bias *can* be thought of in terms of character, then; but should it be? Besser-Jones's comments direct us to the important question of the extent to which individuals can exercise control over the expression in behaviour and judgement of their own implicit biases. Are these processes and the behavioural dispositions they produce the kind of things that individuals can regulate or control so as to bring their actions into line with their considered evaluative commitments? If not, then the manifestation of implicit bias starts to look more like the spasm of a hand, or perhaps the intrusion of compulsive or phobic thoughts—uncontrollable, and not attributable to the agent in a way that enables us to evaluate her, or her evaluative commitments, on the basis of such behaviours. Crucial in answering this question, then, is assessing whether

individuals have control over the expression of implicit biases. We start to address this question in the following section.

5.2 Control Conditions and Implicit Bias

It has been commonplace to contrast implicit associations with *controlled* processing.[5] The idea that implicit biases are not under the agent's control has contributed to and underpinned the claim that individuals are not responsible for implicit biases, and that they don't 'stand for' who the agent is. This is bound up with the nature of such biases as 'implicit': 'without the consciousness of having an implicit race bias, it seems difficult or impossible to exert control to correct it' (Cameron et al. 2010, 275). But this is too simple. Notions of control proliferate in the philosophical literature,[6] and to ascertain whether individuals are able to exercise control over their implicit biases in a way that can legitimize evaluation of them as part of one's character, we need to take care to be clear about what sense(s) of control might be at issue. In this section, we examine some candidate notions of control, as they are found in the current literature.

5.2.1 Direct control

Saul has suggested that individuals' lack of *direct* control over implicit biases should exempt individuals from moral responsibility for them: individuals 'do not [when made aware of biases] instantly become able to control their biases, and so they should not be blamed for them' (2013, 55). This sounds reasonable, but many traits and dispositions that are typically thought of as part of our characters are not under any kind of direct or immediate control, either. Indeed, the cultivation of stable dispositions to act that are underpinned by evaluative commitments has

[5] Dovidio, Kawakami, Johnson, and Johnson 1997.

[6] Some of the more interesting notions taken from the philosophical literature include: ultimate control (Kane 1989), regulative and guidance control (Fischer and Ravizza 1999), rational control (Smith M. 2001, Smith A. 2008), intervention control (Snow 2006), valuative control (Hieronymi 2006), indirect control (Arpaly 2003), long-range control (Feldman 2008), narrative control (Velleman 2005; he never uses the terminology, but we feel this captures the general idea), ecological or soft control (Clark 2007), fluent control (Railton 2009), habitual control (Romdenh-Romluc 2011), skilled control (Annas 2011), and dialogic control (Doris 2015).

traditionally been considered to require time and practice.[7] We submit that less direct types of control are often actually paradigmatic of the kind of control we have over aspects of our character—witness, for instance, the Aristotelian notion of habituation of virtuous traits. As such, being subject to such indirect types of control permits a mental state or disposition to count as a component of a person's character, and so to be properly the object of character-based moral evaluations.

However, even if individuals have indirect control over implicit biases there remains further unpacking of the kind of indirect control at issue. Moreover, doubts may remain about the significance of other important forms of control that we don't have over the operation of implicit biases, such that it may indeed be inappropriate to consider the agent responsible for their influence. We now consider three further kinds of control that, arguably, individuals lack with respect to implicit biases.

5.2.2 Unified agency and reflective control

Levy argues that 'agents should be excused responsibility for actions caused by implicit attitudes' (2011, 17). One central idea that supports this claim, from Levy, is that implicit attitudes have an associative structure, and as such cannot play a role in unifying the agent. On Levy's account, being an agent—being an individual with evaluative commitments that structure one's plans and projects—requires being able to manifest a kind of 'diachronic unity'. Diachronic unity requires being able to plan and pursue projects, which in turn requires that one's attitudes must be norm-governed such that they can satisfy certain desiderata like avoiding inconsistency and meeting rationality requirements of means-ends reasoning. Explicit attitudes and beliefs are governed by these sorts of norms, and are a kind of 'rule-based processing' (Levy 2011, 13), which enables individuals to bring their evaluative commitments into a coherent structure, and plan according to them. In this way, we can see 'where the agent stands' over time. But Levy argues that the associative, non-inferential structure of implicit biases (and other implicit associations and processes) simply means that they are not so governed, and so not the kind of state that can underpin this

[7] For detailed evaluation of this control condition for moral responsibility in relation to implicit bias, see Holroyd 2012.

sort of unification.[8] Since they are not under the agent's control in a way that enables the unification of agency and the expression, over time, of who the agent is, they are not the proper objects of evaluative assessments, nor relevant to the evaluation of the agent's character.

We disagree for two reasons. First, we see no reason to suppose that implicit, associative processes in general, as a kind of mental structure, cannot contribute to the unification of an agent. Consider for example the findings from Moskowitz and Li (2011): individuals who were actively committed to 'egalitarian' goals (goals of treating individuals fairly) were better able to regulate the expression of negative implicit biases. This 'active commitment' involved being strongly committed, in reports on explicit beliefs, to such egalitarian values, and having the specific goal of treating people fairly activated. As goal activation can be automatic, this need not involve conscious strivings or effort. Because the regulation was not due to conscious effort (subjects were placed under cognitive loads to prevent this), the results indicate that automatic implicit processes, outside of the awareness of reflective agency, are serving to bring behaviour into conformity with explicit values. If this is right, then there are cases in which non-inferential, associative, processes seem to be able to play a role in unifying agency. (See also Devine et al. (2002), who also hypothesize that in individuals strongly committed to fair treatment, preconscious and automatic regulatory systems inhibit the expression of implicit bias.)

Second, even if we accept Levy's claim that such states disrupt or fail to contribute to unified agency, this by itself does not entail that such states are not candidates for moral evaluation in an agent who meets some threshold of unity by other means. For example, it is implausible that behavioural dispositions, in isolation or on their own, would achieve Levy's required kind of unity. Rather, beliefs and commitments that structure these behavioural dispositions are also needed to achieve that unity. But this does not mean that, once an agent meets the threshold of unity, and is able to pursue projects coherently over time, their behavioural dispositions remain outside the scope of moral evaluation. Analogously, we can happily accept the claim that implicit biases, operating in isolation and on their own, do not themselves have or provide the

[8] It is not uncontroversial that implicit processing is associative (see Mandelbaum 2015), but we grant this assumption for the purposes of argument.

structure required for unified agency. This, however, does not entail that the operation of those implicit biases contribute *nothing* to who the (otherwise unified) agent is, or that they cannot be the target of evaluation. Consider an individual's associations concerning gender and leadership: many have been found to have stronger associations between men—rather than women—and leadership qualities (see Valian 2005). This association alone could of course not provide the unity Levy requires for agency. (It is hard to see how any association by itself, implicit or otherwise, could do that.) But once the agent meets the threshold of unity that makes her qualified for normative evaluation in general, this component of her cognitive make-up also becomes a viable target of assessment, and we can consider whether it reflects well or badly on her.

5.2.3 Alienation and evaluative control

One might still find something troubling in the idea that certain actions are influenced by processes (e.g. negative implicit biases) which conflict with explicit attitudes (commitment to fair treatment), and which one is unable to bring into line with one's explicit attitudes. That is, one might worry that implicit biases are not subject to the sort of 'evaluative control' that we have over other parts of our cognitive structures.[10] One might insist that if an agent's action or judgement 'is not responsive . . . to the beliefs, values, and so on of the agent, then the person has lost control over it' (Levy 2011, 5) and therefore that action or judgement cannot be indicative of where she stands as an agent—it is not part of her character.

The cases of implicit bias that have garnered the most attention are those in which the individuals' explicit attitudes come apart from her implicit associations. For example, a person has explicit attitudes endorsing anti-discrimination, whilst being influenced by implicit associations between black men and negative stereotypical associations. In such cases, actions driven by the person's implicit biases appear to be candidates for actions which are out of control, in exactly this sense: they are unresponsive to the agent's explicit evaluative attitudes (at least her considered evaluative attitudes), and in that respect she is alienated from her implicit attitude.

[10] See also Smith 2008.

The notion of *alienation* has been used to capture this psychological state of affairs, where a person is alienated from certain states or processes that are present in her psychological make-up. The notion has also loomed large in 'real-self' views of moral responsibility, where some have argued that 'alienation' can (in at least some cases) exculpate because those alienated states or dispositions are not part of the 'real self'. However, we see no good reason to suppose that alienation entails that the states from which the agent is alienated should not be part of our evaluation of that agent. As Besser-Jones claimed in relation to Mark and Jane, it is not only that Jane has implicit biases, but also *that she is alienated from them* that should be reflected in our evaluation of her character (as less bad in this respect than Mark—but, we add, less good than some third character who neither has nor manifests implicit race bias). Here we also find ourselves sympathetic to aspects of Josh Glasgow's treatment of alienation from implicit bias. Focusing on a case of dissociated implicit and explicit attitudes, Glasgow identifies such implicit attitudes as ones from which the agent is alienated. Nonetheless, asking us to reflect on our own judgements about our relationship to those attitudes from which we are alienated, Glasgow strongly asserts that, whilst confident in his alienation from the implicit biases he harbours, and 'wholeheartedly disavow[ing] such nonsense', he would not doubt their presence, nor that he is culpable for their presence: 'alienation is not sufficient for exculpation' (ms, 8). In short, while *being alienated from* one's implicit biases is relevant to the evaluation of one's character, it does not completely absolve one from *having* or *being influenced by* those implicit biases.

Whilst primarily concerned with moral responsibility, Glasgow here holds two distinctive claims: a) that implicit biases are not part of the agent's moral character (due to alienation, or lack of evaluative control), and b) that nonetheless, the presence of those biases can influence the moral evaluation of her and her actions.[11] We agree with Glasgow that moral evaluation is appropriate. However, we reject the idea that lack of evaluative control (the sort of 'alienation' identified here) entails that implicit biases are not part of 'who the agent is'. In the following section (5.3), we articulate a kind of control that suffices for considering implicit biases ~~are~~ part of moral character.

[11] See also Stump 1996.

5.2.4 *Intervention control*

So far, we have examined claims about the lack of direct control over implicit biases, about their associative structure that does not underwrite unified agency, and about the fact that agents are often 'alienated' from the implicit biases they may be influenced by. We have argued that none of these considerations provides sufficient reason to suppose that implicit biases are not part of an agent's character, and so should be left out of evaluations of 'who she is' and 'what she stands for'.

However, a further concern may remain: if implicit biases influence behaviour and judgement without the agent being able to prevent them, then we should not consider the influence of implicit biases in behaviour—such behavioural dispositions—as part of the agent's character. These behavioural dispositions should be considered more like an agent's disposition for her hand to spasm under certain circumstances—something outside of her control, and not therefore attributable to her. We might identify the worry here in terms of a lack of 'intervention' or 'inhibition' control—the agent's ability to inhibit certain actions that are not expressive of her core values and moral commitments (see Levy 2011, 16). Nancy Snow articulates the notion of inhibition or intervention control in her discussion of virtuous agency and its relation to the automatic processing involved in the pursuit of virtuous goals. Snow notes that whilst automatic processes can be initiated and run without the guidance of reflective deliberation, they are nonetheless attributable to the agent insofar as she has intervention control over them. An agent exercises intervention control when she intervenes in an instance of behaviour that is unfolding smoothing and relatively unthinkingly on its own (such as cycling while on 'autopilot mode')—either inhibiting that behaviour or redirecting it (2006, 549).

A question that this raises is whether individuals have the ability to intervene in, and bring under reflective control, the manifestation of their own implicit biases. Surely if one cannot in any way prevent the expression of implicit biases, one is not properly morally evaluable for those expressions?

Some empirical studies indicate that there is indeed trouble lurking here; attempts to consciously suppress implicit biases are notoriously problematic, and an individual's lack of awareness of the extent to which

she is influenced by implicit biases seems to threaten her ability to inhibit their influence.[12]

Accordingly, we are willing to concede that understood thus, individuals will often lack intervention control over implicit biases. However, we believe—and argue in the next two sections—that this articulation of 'intervention control' is too narrow, focusing only on the ability to inhibit implicit processes by bringing them directly and immediately under reflective control. To make our case, we articulate a model that draws together more extensive resources for intervening in behaviour and bringing one's actions and judgements into line with one's evaluative commitments. We argue that this model of control can underpin the idea that implicit biases should legitimately be included in the evaluation of an agent's character, before articulating (in section 5.5) how this prompts us to revise the notion of self-regulation and character development.

5.3 Clark on Ecological Control

In this section we develop an account of control according to which implicit biases are (or can be brought) under an agent's control. We also argue that in light of this account of control, implicit biases can be thought of as part of a person's character, and are thus relevant to the moral evaluation of that character. We draw on and develop Andy Clark's notion of 'ecological control', using Clark's articulation of this idea as our point of departure to simplify our own exposition, and because he has done so much to unpack and defend it. We situate the model within Clark's larger picture before going on to develop and apply it in the context of empirical research on implicit biases.

5.3.1 Motivating the idea

What is it to exercise ecological control in cognition and action? Clark elucidates it as follows:

[12] See Galinsky and Moskowitz 2000; and Saul ~~ms~~ for pursuit of the question of whether this unawareness should lead us to quite radical sceptical conclusions. On the other hand, empirical findings from Monteith and Voils (1998) indicate that at least sometimes individuals are aware that their actions are not in conformity with their ideals, and able to attribute this to implicit biases.

> Ecological control is the kind of top-level control that does not micro-manage every detail, but rather encourages substantial devolvement of power and responsibility . . . And it allows (I claim) much of our prowess at thought and reason to depend upon the robust and reliable operation, often (but not always) in dense brain-involving loops, of a variety of non-biological problem-solving resources spread throughout our social and technological surround. (Clark 2007, 101)

One well-known element of Clark's model is alluded to here, namely his commitment to the idea that the boundaries of minds, and even selves, are fluid, and need not coincide with the biological boundaries of the organisms they animate (see Clark and Chalmers 1998; cf. Sterelny 2003, 2010). Clark holds that this is especially true of human beings, who exploit and incorporate features of the physical structure of their own bodies and surrounding environment into their strategies of acting and pursuing goals, thus exemplifying the type of ecological control in which he is interested. Various processes employed by 'ecological controllers' incorporate features of their bodies and of the external world, integrating them into 'whole new unified systems of distributed problem-solving' (Clark 2007, 103).

Ecological control is devolved, distributed, diffuse, decentralized, often spread out over time, and typically not accompanied by the kind of rich consciousness-awareness characteristic of higher-level and reflective cognition. This type of control is obviously involved in some lower-level physiological systems. However, Clark claims it is not restricted to 'the "autonomic" functions (breathing, heart-beat, etc.)' but instead holds that 'all *kinds* of human activities turn out to be partly supported by quasi-independent non-conscious sub-systems' (Clark 2007, 110). Indeed, ecological control often involves the effective coordination and calibration of such subsystems: corralling, nudging, tweaking their coordinated operation in the service of a particular goal, rather than micromanaging every detail of each individual component. Consider the simple act of reaching out our hand to pick up a coffee mug. While this may seem simplicity itself, and a behavior under near complete conscious control, Clark argues that even in this case the appearance is misleading, and that 'fine-tuned reaching and grasping involves the delicate use of visually-received information by functionally and neuro-anatomically distinct sub-systems operating, for the most part, outside the window of conscious awareness' (Clark 2007, 109). Even in this mundane case, much of the control of behaviour is distributed throughout the 'non-conscious

circuitry that guides the most delicate shape-and-position sensitive aspects of reach and grasp' (Clark 2007, 109–10).

So pathways of ecological control—'loops' of influence, as Clark and others sometimes call them—are mediated by *sub-personal* structures and subsystems of which the organism is not directly aware or in immediate, direct, control at the *personal* level. Ecological controllers often exploit the features of those sub-personal structures and subsystems to their advantage, to simplify problem-solving and fine-tune performance in the service of achieving their personal goals and complying with their personal values. (Clark uses the personal/sub-personal distinction, but the terminology originates with Dennett 1969, see also Elton 2000.)

As mentioned above, though, Clark holds that these kinds of sub-personal structures and subsystems need not be confined to the organism's skin. Organisms who employ ecological control to manage their behaviour (of whom humans are the example par excellence) often do so by reshaping, even incorporating, elements of their extra-bodily environment into the process of fine-tuning their performance. Loops of ecological control can extend into the environment and back again. Humans do not just exploit the structure they *find* outside their own bodies and heads, but rather take an active hand in *shaping* and *organizing* that environment in such a way that it, too, helps them to fine-tune actions, solve problems, realize their intentions, and express their values. In other words, one central feature of human agency involves supplementing the internal sub-personal mechanisms that guide behaviour by engineering their world, calibrating 'external' sub-personal structures so that they help simplify cognition and bring out the kinds of behaviours and outcomes to which they aspire. A particularly compelling example of this is the effort people put into organizing their offices, to do things like maximize productivity, manage their own moods (or at least ward off despair!) and generally make things epistemically easier on themselves. As Clark points out, in organizing an office, an individual attunes her surroundings to her own particular epistemic needs. In taking such measures, individuals seek to structure and stabilize their environments in ways that 'simplify or enhance the problem-solving that needs to be done' (Clark 2007, 115).

A crucial aspect of the notion, as developed by Clark, is that one goal that people might attempt to achieve using ecological control is to *further*

calibrate the operation of specific sub-personal mechanisms. We might shape our external environments as a means to more effectively manipulating sub-personal operations. Here we see the recursive use of control to enhance and heighten control itself. An agent can do this by fine-tuning the role of subsystems which in turn help produce dispositions and behaviours that can better fulfil her more distal goals, thus allowing her to better behave in ways that more precisely reflect her intentions, and more crisply conform to her considered ideals and values. Ultimately, a person can calibrate subsystems that guide behaviour until eventually they operate, on their own, in precisely the way she wants them to operate, even when she is not consciously and explicitly attending to them.

5.3.2 Developing Clark's model

With the idea of ecological control outlined, an ambiguity in our presentation of it so far can now be teased out. Sometimes, what we call the *exercise* of ecological control involves executing actions or parts of actions without the guidance of reflective, deliberative control. A professional tennis player exercises ecological control, for example, when instinctively responding—without deliberate, reflective, thought—to a baseline shot in tennis. However, individuals can *take* ecological control of something when they reflectively decide to manipulate their mental states or environment, so as to shape their cognitive processes, thus enabling the exercise of ecological control in the future. Given this distinction, what we are calling 'taking' ecological control does require that the agent is at least sometimes able to reflectively control their behaviour. So we acknowledge that the capacity for and occasional exercise of this more deliberate and reflective kind of control is necessary for evaluable action, although its exercise on any one occasion is not.

We mentioned that one way of *taking* ecological control is by calibrating subsystems, and one way to do that is to *practise*. Rehearsal is a common way to deliberately hone a skill, calibrating the sub-personal structures that implement it, making the whole process routine enough that eventually intentions can be expressed smoothly, automatically, and without deliberation, and the performance of the fine-tuned, goal-directed behaviour is natural, easy, and unthinking. This kind of agency can involve and is often accompanied by reflective control—the ability to reflectively decide upon a high-level course of action and carry it

out—but with a view to ultimately obviating the need for reflective agency (or much of it) over lower-level, component parts of the behaviour being practised. In other words, one of the uses of ecological control is that it can be used to attain what Railton (2009) calls *fluid agency*. Once achieved, high-level intentions consciously formed at the personal level can be fluidly enacted via pathways of ecological control mediated by those previously calibrated sub-personal mechanisms, without each individual step being directly controlled or consciously attended to along the way (and perhaps without need for reference to intention; see Romdenh-Romluc 2011). As Clark notes, examples of this use of ecological control can be drawn from sport: 'This is no surprise, I am sure, to any sports player: it doesn't even seem, when playing a fast game of squash, as if your conscious perception of the ball is, moment-by-moment, guiding your hand and racket' (Clark 2007, 17). But as we will see (in section 5.4.3), the particular line of thought can be extended from sports to other domains, and from the sub-personal subsystems that underlie behaviour to those that underlie evaluation and judgement.

Another way of taking ecological control is to co-opt aspects of what is intuitively thought of as our external environment, shaping it so that it can better guide our cognitive processes in a way congenial to our goals and values. In short, beyond the distinction between exercising and taking ecological control, there are several ways in which we can exert ecological control over mental entities. And, as we set out in the next section, there are ways in which we can take and exercise ecological control over implicit biases.

5.4 Ecological Control and Implicit Bias

Implicit biases are troublesome not just because they can contribute to morally problematic outcomes, but also because of their psychological profile. They pose a problem for moral evaluation because they are not introspectively accessible, and so can coexist unknown and alongside explicit attitudes to the contrary; because they can operate outside of conscious awareness; are associative, not under our direct evaluative control, and once activated their expression in judgement and behaviour is difficult to directly manage or completely suppress. But, as we argued in section 5.2, none of these features suffices to exempt it from moral evaluation. Here, we address two questions: first, do we have ecological

~~control over implicit biases~~ (do we exercise ecological control, and can we take ecological control)? Second, is ecological control sufficient (or part of a set of jointly sufficient conditions) for moral evaluation?

Whilst not subject to the kinds of control we considered in section 5.2, empirical work has shown that implicit biases are 'malleable' in a number of ways. This empirical research has investigated the most effective ways of impeding the influence of implicit biases, and revealed that some strategies work while others don't (and still others seem to make matters worse). It has also shown that intuition and common sense tend to be a poor guide to which strategies are likely to be successful. But most importantly it has shown that there is a very real sense in which a person can exert control over her implicit biases. We hold that Clark's notion of ecological control provides a useful way to think about the nature of that control. We identify three different kinds of ecological control that might be exercised in relation to implicit biases. Drawing on the distinction introduced above, we present the first two as concerning instances in which individuals might reflectively *take* ecological control, to better enable their actions to conform to their values. The third example of *exercising* ecological control indicates that individuals' actions can draw on implicit processes so as to be influenced by and calibrated with ~~our~~ values, even when intentional strategies have not been implemented to ensure this.

5.4.1 Environmental props consciously employed for guiding cognitive processes

Clark emphasizes that ecological control can involve the 'offloading' of cognitive structures onto the world beyond the boundaries of the skin (recall the example of structuring one's office to facilitate one's professional activities). One might use similar means to help mitigate one's implicit biases as well. For instance, early studies (Dasgupta and Greenwald 2001) showed that the influence of some implicit racial biases could be weakened simply by exposing participants to pictures of admired black celebrities and other counter-stereotypical images. So, a person might rein in the expression of her own implicit racial biases by putting up pictures of admired black celebrities around her office, thus taking indirect, ecological control over those biases so that her judgements and actions more ~~fluidly~~ express her character and values. A person might engineer her 'external' epistemic environment in other ways to ensure that her intentions and values are more fluidly expressed

in her actions and judgements, and not distorted by the operation of implicit biases. For instance, if one is (justifiably) worried about implicit biases corrupting the assessment of candidates in a job search, one can take measures to remove information from application dossiers that may trigger those implicit biases in the first place.

5.4.2 Cognitive props consciously employed for guiding cognitive processes

Another promising strategy for mitigating the influence of implicit biases has been investigated under the name of 'implementation intentions' (Webb, Sheeran, and Pepper 2012). For example, an individual seeking to exert control over her implicit biases might deliberately repeat to herself, 'If I see a Black face, I will think "safe",' practising this line of thought enough that it becomes routine and automatic, thus defeating her implicit racial bias. One might draw an analogy to Clark's discussion of a sports player who practises, calibrating the operation of sub-personal subsystems to bring them in line with intentions, and thus developing a certain kind of fluid and unthinking control. In the case of implicit biases a person does not practise, say, a backhand tennis stroke or the subtle biomechanics required to throw a curveball, but rather rehearses a psychological process aimed at forming a particular kind of counter-bias association. Once successfully formed, that new association will be able to bring the way she makes snap judgements into line with her more thought out intentions, thus better reflecting her character and values.

5.4.3 Automatic processes as props unconsciously employed for guiding cognitive processes

Following Clark, we have argued that some features of our environment can shape our cognitive processes, and can be manipulated so as to influence the subsystems that will produce judgements and actions. We also hold that automatic features of our mental processes can be manipulated in similar ways, and to similar ends. Recall the finding, mentioned above (Moskowitz and Li 2011), that individuals who were actively committed to egalitarian goals (without consciously reflecting on such goals themselves) manifested less implicit racial bias in experimental testing. Their explanatory hypothesis for this effect was that certain goals automatically block other competing goals (goals such as speed and efficiency, which have been found to encourage the reliance on

implicit associations). The agent's values and goals themselves, then, can play a role qua mechanisms that influence and calibrate the subsystems that run without reflective or direct control. This is a case of one element of a person's psychological economy influencing another. The agent's values 'keep in check' the operation of implicit bias, such that pursuing certain values is one way of exercising ecological control even when one is not actively monitoring one's actions with respect to whether they promote (or depart from) those values. Crucially, this can be so without the agent expressly intending, at any point, to put in place mechanisms for this purpose. One can exercise ecological control, then, without having reflectively taken (in senses 5.4.1 and 5.4.2 above) ecological control.

There is much more that can be said here (indeed, the literature on implicit biases and how to influence them is growing at a tremendous rate), but for present purposes we can simply point out that the idea that an agent's implicit biases are beyond her control in any relevant sense is simply false. Whilst implicit biases themselves are rarely under our direct, evaluative, or intervention control, empirical studies indicate that the kind of ecological control that we exercise in many actions and cognitions can, in many ways, be extended to implicit biases as well. While some of these strategies for controlling implicit biases may look exotic, we hold that another virtue of Clark's discussion is that he articulates a notion of control that seems to capture something at the core of many of these strategies, whilst making clear that use of this kind of control is actually quite mundane. It is a type of control that underlies a vast swathe of human behaviour and problem-solving.

This observation supports our second claim, that susceptibility to ecological control is sufficient for evaluation, and that in virtue of their susceptibility to this kind of control, implicit biases and actions influenced by them are proper targets of character-based evaluation. We believe our discussion renders this claim plausible by showing that there is nothing unusual about the processes implicated in implicit biases, and that denying they are the proper objects of moral evaluation would commit the proponents of this claim to far wider scepticism about the moral evaluation of our cognitive states and actions than they seem to acknowledge.[13] Unless some further way in which such processes

[13] Nomy Arpaly 2003, for instance, has argued that automatic actions of tennis players are apt targets for evaluation; see also Doris 2002, 2009.

differ from our other cognitive processes can be identified, denying that such states are evaluable as part of 'who the agent is' will entail denying that many of our commonplace actions and judgements are apt candidates for character-based evaluation also. Whilst embracing scepticism is always an option, we do not believe that it would be charitable to attribute this to our interlocutors. ~~There are two further challenges to address.~~

5.4.4 Two challenges

The first challenge to this conclusion asks whether all kinds of control that might fall under the rubric of 'ecological control' are sufficient for biases (that are potentially controllable in these ways) to be evaluable. To see the worry here, note that some of the methods of ecological control we identified (conscious use of environmental props) fall into the category of the 'negative programme' of character development, described by Webber in his chapter on this volume. The negative programme involves acknowledging our susceptibility to the influence of implicit biases (and other forms of situational manipulation), and aiming to mitigate or eliminate those influences (changing the environment to prime for counter-stereotypical exemplars, say). Is it really possible, Webber's challenge goes, to try to counteract every possible bias one might be susceptible to (race, age, gender, height, class, educational background, and so on and so on)? If not, then is it really fair to take such biases into account when we evaluate individuals, finding it to their discredit should some such influences persist—especially in the absence of comprehensive and easily accessible information about how to combat all forms (and potential interactions) of such bias?

We argue that it is reasonable to maintain that implicit biases can be part of the target of character-based evaluation, even if such information is not (yet!) readily accessible, and even if it would be extremely demanding to undertake the negative programme in relation to each bias. Firstly, the fact that it would be very demanding does not mean that it would not be to the agent's credit to undertake such a project, and to their discredit to the extent that they have failed to do so. But secondly, and most crucially, it is worth noting that the second two strategies for ecological control that we identify—consciously employing cognitive props, and the role of automatic processes unconsciously employed in regulating implicit biases—fall under what Webber

describes as the 'positive programme' of character development. This programme consists of the entrenchment of attitudes and dispositions to act well. Ecological control strategies of this variety are properly thought of as part of a positive programme of developing cognitive and motivational habits that enable agents to be robustly resistant to implicit biases. It is precisely because the susceptibility to implicit bias is bound up with the strength of agents' habits and commitments that such dispositions are apt candidates for character-based evaluation.

The second challenge asks whether, given the extent that an individual's dispositions are dependent upon these kinds of 'props', those dispositions should be considered part of that individual's character at all. Don't we give up on the notion of character altogether once we recognize how fragile those character trait-like dispositions are, and how dependent they are on the support of such environmental resources? We think not. Rather, we respond by pointing out that while this challenge may seem decisive to those in the grip of a common picture of what character amounts to, it is not the only picture, and perhaps not the best one. Rather, our view fits quite nicely what Maria Merritt (2000) calls the Humean model of character. She distinguishes the Aristotelian model of character, on which an individual's dispositions are 'firm and unchanging' and are motivationally sufficient to produce virtuous action, from the Humean model that she prefers (in no small part because she holds it is better equipped to withstand the situationist critiques of virtue ethical theories). Merritt shows how the Humean model permits stability of character to be supported by a range of mechanisms that prominently include social relations and environmental settings (2000, 377–80). By her lights, the normative task is then to ~~seek~~ the social and environmental factors that best support stable dispositions to act well. In this, we see ourselves as fellow travellers with Merritt, and assert that our discussion of character in this paper is best understood on the Humean model she champions. And so in our view, what falls within the scope of character evaluations is not only the behavioural dispositions implicated in implicit bias themselves, but also the agent's sensitivity (both conscious and otherwise) to the kinds of environmental settings that permit those dispositions to take hold—or those which function as effective regulatory mechanisms in serving the agent's endorsed goals and values.

In sum, then, we hold that since a person can exert intervention and ecological control over implicit biases, there is a real sense in which whether or not they influence an individual's behaviour is very much a reflection of that person's character. Thus, the behavioural dispositions implicated in implicit biases, and the behaviours they influence, are an appropriate subject matter for character-based evaluation.

5.5 Character and Self-Regulation with Ecological Control

We have argued that it is possible for individuals to exercise control over their implicit biases on the model of ecological control—devolving the task of mitigating the influence of implicit bias to parts of their environment, or to mechanistic ~~conscious~~ or automatic cognitive responses. In this section, we make an important qualification of this claim, and tease out two important implications of our conclusion.

5.5.1 Ecological control and epistemic conditions

We have argued that an individual *can* take and exercise ecological control over her implicit biases—it is possible for her to do so—and so she cannot, on grounds of lack of control, maintain that her implicit biases are not evaluable.

However, our conclusion is vulnerable to the following objection: we claim that it is possible for individuals to exercise such control (see sections 5.4.1–3) over implicit biases. But whether, for any individual, she can in fact exercise ecological control depends on whether she is aware of these possibilities (and indeed, aware of the phenomena of implicit bias, and that she may be affected by it). So the mere possibility of having ecological control is not sufficient for implicit biases to be considered as 'part of the agent' and hence morally evaluable. In addition to the control conditions, epistemic conditions must also be met as well.

For present purposes, we are willing to accept that ecological control can permit moral evaluation only if other conditions also obtain. (Indeed, the authors' opinions differ on whether epistemic conditions are necessary. One of us believes that awareness of implicit bias is not required, because awareness of all processes involved in the production of action cannot be a condition on that action being evaluable (Holroyd

2012), and because normative conditions about what individuals should know rather apply (Holroyd ~~forthcoming~~). The other believes that they are, but that these epistemic conditions are sensitive to context in various ways, including the social role occupied by the individual in question, as well as the contents of her external epistemic environment (see Kelly and Washington forthcoming).)

However, this dispute can be set aside: our key claim is that a certain argumentative move cannot be made. That move is to deny that implicit biases are part of who the agent is—or that they can be evaluated for being influenced by them—because the agent lacks control over such mental entities. We have identified an important sense in which agents *can* have control over such mental entities. If implicit biases are rogue states beyond moral evaluation, then, it is not because they are beyond the scope of the agent's control.

If this is endorsed, then we believe there are two further implications for our understanding of control, character, and self-regulation.

5.5.2 Intervention control on an ecological model

Recall Snow's idea that even if one lacks reflective, direct control over one's cognitions and behaviours, they remain morally evaluable if one retains intervention control—the ability to exert influence on autonomously running processes by stopping them or redirecting how they shape action. According to this understanding of intervention control, individuals lack it in relation to implicit biases—it is very difficult to prevent behavioural manifestation of implicit bias via direct reflective control. The job interview panellist cannot effectively intervene on the operation of implicit biases as they influence cognition, simply by thinking: 'Oops, there it goes; better get my cognitive processes back on track and stop that biased evaluation.'

If the ecological control model is endorsed, then we claim that a broader understanding of 'intervention control' is warranted. An agent can intervene in some automatic process not by bringing it under direct reflective control at the moment of its activation, but by diverting its activation by means of some environmental or cognitive prop put in place to derail unwanted cognitive or behavioural patterns. If this more expansive understanding of intervention control is endorsed, then agents can retain control and be morally evaluable for actions even when those actions cannot be brought under direct reflective control as typically

understood. Thus the range of behaviours that are candidates for evaluation as moral or virtuous action could be greater than on Snow's narrow construal of intervention control—although its full extent, and how that class might be delineated in a full account of ecological control, is beyond the scope of this paper.

5.5.3 Character development on an ecological model

We initially framed our discussion by considering whether implicit biases are properly thought of as part of 'who the agent is', and showing how this question turned on the kind of control the agent might have over the mental entities involved in implicit bias. Our main claim has been that taking ecological control is one way in which people can shape the processes and behaviours that constitute character. This claim also has implications for how to understand character development.

Our inquiry was motivated by Besser-Jones's remarks that indicated that implicit biases were a proper candidate in the moral evaluation of character. We have attempted to vindicate this claim by showing that implicit biases are appropriately understood as under the agent's control in a way that renders them objects of character-based moral evaluation. But even given her willingness to include such implicit attitudes in the set of character-constituting attitudes, we find her remarks on character development in need of some fleshing out. In her early work (2008) Besser-Jones focuses on self-regulation by means of prescriptions that remind the agent of the priority of acting well, and strategic rules for action or a decision procedure to be relied on in difficult cases.

Given the significant role that implicit processes have in influencing action, it is clear that both of these strategies will be at best partial, as they speak to the agent's regulation of wayward influences that are within the field of her cognitive attention. In her later (2014) work, Besser-Jones articulates two more strategies that might enable the development of virtuous character even in the face of challenges presented by automatic processes that can hinder the pursuit of virtuous goals. The first strategy involves the articulation of a hierarchical structure of goal-directed activity that will help translate abstract ideals into sequences for concrete actions in fulfilment of those goals. This, in turn, will set up feedback loops that will facilitate *evaluating* whether those actions are in fact serving those goals (2014, 150–2). The second type of strategy for developing virtue also appeals to self-regulation, this time via the

articulation of plans for action which are then used to *anticipate* opportunities for acting well, and helping ensure those opportunities are seized (152–4). For example, Besser-Jones recommends the use of implementation intentions, which enable agents to be cued to situational features that should trigger goal pursuits, rather than irrelevant situational features of the sort shown to hinder acting well.

These two strategies seem much better placed to address the kinds of implicit cognitions that may both serve, or hinder, the development of good character and action upon it. However, given the concerns about implicit bias, we propose two further ways in which these strategies for the development of virtue could be supplemented.

First, in relation to the evaluation of whether action sequences are serving goals, we note that it may not always be obvious whether an agent's actions are hindering one's goals. Consider the goal of acting justly; if agents are (for various reasons) unaware that their actions are inflected with implicit bias, then it may not be apparent to them that certain actions they perform are not serving this goal. Accordingly, agents should supplement their hierarchical goal structures with active investigation into possible ways their actions might be hindering their goals which are not available to introspection.

Secondly, with respect to the plans for action, we note that Besser-Jones gives most attention to the manipulation of cognitive resources (such as implementation intentions) that help agents act in accordance with their commitments. Our discussion of ecological control, and the way that external environments can be harnessed to enable agents to better act in the service of their goals, shows the importance of supplementing these strategies with ones that utilize environmental as well as cognitive props to enable individuals to effectively take and exercise ecological control.

These friendly extensions to Besser-Jones's prescriptions are attuned specifically to how one might ensure the development of virtuous character and action in the face of challenges from research on implicitly biased cognitions.

5.6 Concluding Remarks

We have argued that there is an important sense in which individuals have control over implicit biases, and this kind of control is a sort that is

commonplace in our exercise of agency. Ecological control is the structuring of one's environment and cognitive habits such that autonomous processes and subsystems can effectively fulfil one's person-level goals. We set out at least three ways in which individuals can take or exercise ecological control with respect to their implicit biases.

There are two central implications of this claim: firstly, that agents can have this kind of control means that (subject to other necessary conditions being met) such implicit attitudes can be the appropriate target of character-based evaluation. Secondly, that agents can exercise this control means that (subject to other necessary conditions obtaining) such implicit attitudes can be properly regarded as part of 'who the agent is'—part of her character, which is as a whole subject to moral evaluation. This does not, of course, preclude the agent taking up a stance of disgust or alienation towards that part of her character; that stance is part of 'who she is' too.

If one's character involves not only one's beliefs, behavioural dispositions, and attitudes towards these mental entities, but also the cognitive habits that one uses environmental and mechanistic strategies to shape, then a model of character development and regulation must also make adequate prescriptions for the exercise of ecological control.

Works Cited

Amodio, D., Harmon-Jones, E., and Devine, P. 2003. Individual Differences in the Activation and Control of Affective Race Bias as Assessed by Startle Eyeblink Response and Self-Report. *Journal of Personality and Social Psychology* 84(4): 738–53.

Annas, J. 2011. *Intelligent Virtue.* Oxford: Oxford University Press.

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

Besser-Jones, L. 2008. Social Psychology, Moral Character, and Moral Fallibility. *Philosophy and Phenomenological Research* 76 (2): 310–32.

Besser-Jones, L. 2014. *Eudaimonic Ethics: The Philosophy and Psychology of Living Well*. New York: Routledge Press.

Cameron, D., Payne, K., and Knobe, J. 2010. Do Theories of Implicit Race Bias Change Moral Judgment? *Social Justice Research* 23: 272–89.

Clark, A. 2007. Soft Selves and Ecological Control. In *Distributed Cognition and the Will*, ed. D. Spurrett, D. Ross, H. Kincaid, and L. Stephens. Cambridge, MA: The MIT Press, 101–21.

Clark, A., and Chalmers, D. 1998. The Extended Mind. *Analysis* 58 (1): 7–19.

Dasgupta, N., and Greenwald, A. 2001. On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals. *Journal of Personality and Social Psychology* 81 (5): 800–14.

Dennett D. 1969. *Content and Consciousness*. London: Routledge & Kegan Paul.

Devine, P., Plant, E., Amodio, D., Harmon-Jones, E., and Vance, S. 2002. The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond without Prejudice. *Journal of Personality and Social Psychology* 82 (5): 835–48.

Doris, J. 2002. *Lack of Character: Personality and Moral Behavior*. New York: Cambridge.

Doris, J. 2009. Skepticism about Persons. *Philosophical Issues* 19: 57–91.

Doris, J. 2015. *Talking To Ourselves*. Oxford: Oxford University Press.

Dovidio, J. F., and Gaertner, S. L. 2000. Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science* 11: 319–23.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., and Howard, A. 1997. On the Nature of Prejudice: Automatic and Controlled Processes. *Journal of Experimental Social Psychology* 33: 510–40.

Elton, M. 2000. Consciousness: Only at the Personal Level. *Philosophical Explorations* 3 (1): 25–42.

Feldman, R. 2008. Modest Deontologism in Epistemology. *Synthese* 161: 339–55.

Fischer, J., and Ravizza, M. 1999. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Galinsky, A. D., and Moskowitz, G. B. 2000. Perspective-Taking: Decreasing Stereotype Expression, Stereotype Accessibility, and In-group Favoritism. *Journal of Personality & Social Psychology* 78 (4): 708–24.

Gendler, T. S. 2011. On the Epistemic Costs of Implicit Bias. *Philosophical Studies* 156: 33–63.

Glasgow, J. ~~ms~~ 'Alienation and Responsibility',

Hieronymi, P. 2006. Controlling Attitudes. *Pacific Philosophical Quarterly* 87 (1): 45–74.

Holroyd, J. 2012. Responsibility for Implicit Bias. *Journal of Social Philosophy* 43 (3): 274–306.

Holroyd, J. 2015. Implicit Bias, Awareness and ~~Epistemic Innocence~~, *Consciousness and Cognition* 33: 511–23.

Jost, J. T., Rudman, L. A, Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., and Hardin, C. D. 2009. The Existence of Implicit Bias is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore. *Research in Organizational Behavior* 29: 39–69.

Kane, R. 1989. Two Kinds of Incompatibilism. *Philosophy and Phenomenological Research* 69: 219–54.

Kelly, D., and Roedder, E. 2008. Racial Cognition and the Ethics of Implicit Bias. *Philosophy Compass* 3 (3): 522–40.

Kelly, D., and Washington, N. forthcoming. Who's Responsible for This? Implicit Bias and the Epistemology of Moral Responsibility. In *Philosophy and Implicit Bias*, ed. M. Brownstein and J. Saul. Oxford: Oxford University Press.

Levy, N. 2011. Expressing Who We Are: Moral Responsibility and Awareness of our Reasons for Action. *Analytic Philosophy* 52(4): 243–61.

Levy, N. 2012. Consciousness, Implicit Attitudes, and Moral Responsibility. *Noûs*: 1–22 doi: 10.1111/j.1468-0068.2011.00853.x.

Mandelbaum, E. 2015. Attitude, Inference, Association: On the Propositional Structure of Implicit Bias. *Noûs.* doi: 10.1111/nous.12089.

Merritt, M. 2000. Virtue Ethics and Situationist Personality Psychology. *Ethical Theory and Moral Practice* 3 (4): 365–83.

Monteith, M. J., and Voils, C. I. 1998. Proneness to Prejudiced Responses: Toward Understanding the Authenticity of Self-Reported Discrepancies. *Journal of Personality and Social Psychology* 75 (4): 901–16.

Moskowitz, G. B., and Li, P. 2011. Egalitarian Goals Trigger Stereotype Inhibition: A Proactive Form of Stereotype Control. *Journal of Experimental Social Psychology* 47 (1): 103–16.

Payne, B. K. 2005. Conceptualizing Control in Social Cognition: The Role of Automatic and Controlled Processes in Misperceiving a Weapon. *Journal of Personality and Social Psychology* 81: 181–92.

Railton, P. 2009. Practical Competence and Fluent Agency. In *Reasons for Action*, ed. D. Sobel and S. Wall. Cambridge: Cambridge University Press, 81–115.

Romdenh-Romluc, K. 2011. Agency and Embodied Cognition. *Proceedings of the Aristotelian Society* 111 (1): 79–95.

Saul, J. 2013. Implicit Bias, Stereotype Threat and Women in Philosophy. In *Women in Philosophy: What Needs to Change*?, ed. F. Jenkins and K. Hutchinson. Oxford: Oxford University Press, 39–60.

Saul, J. ms. 'Implicit Bias, Stereotype Threat and Women in Philosophy'.

Smith, A. 2008. Control, Responsibility, and Moral Assessment. *Philosophical Studies* 38: 367–92.

Smith, M. 2001. Responsibility and Self-Control. In *Relating to Responsibility: Essays in Honour of Tony Honore on his 80th Birthday*, ed. P. Cane and J. Gardner. Oxford: Hart Publishing, 1–19.

Snow, N. 2006. Habitual Virtuous Actions and Automaticity. *Ethical Theory and Moral Practice* 9: 545–61.

Sterelny, K. 2003. *Thought in a Hostile World*. New York: Blackwell.

Sterelny, K. 2010. Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences* 9 (4): 465–81.

Stump, E. 1996. Persons: Identification and Freedom. *Philosophical Topics* 24: 183–214.

Valian, V. 2005. Beyond Gender Schemas: Improving the Advancement of Women in Academia. *Hypatia* 20 (3): 198–213.

Velleman, D. 2005. The Self as Narrator. In *Autonomy and the Challenges to Liberalism: New Essays*, ed. J. Christman and J. Anderson. New York: Cambridge University Press, 56–76.

Webb, T., Sheeran, P., and Pepper, J. 2012. Gaining Control over Responses to Implicit Attitude Tests: Implementation Intentions Engender Fast Responses on Attitude-Incongruent Trials. *British Journal of Social Psychology* 51 (1): 13–32.