

Restoring Control: Comments on George Sher

Neil Levy

Received: 1 March 2007 / Revised: 21 June 2007 / Accepted: 26 June 2007 /

Published online: 23 August 2007

© Springer Science + Business Media B.V. 2007

Abstract In a recent article, George Sher argues that a realistic conception of human agency, which recognizes the limited extent to which we are conscious of what we do, makes the task of specifying a conception of the kind of control that underwrites ascriptions of moral responsibility much more difficult than is commonly appreciated. Sher suggests that an adequate account of control will not require that agents be conscious of their actions; we are responsible for what we do, in the absence of consciousness, so long as our obliviousness is explained by some subset of the mental states constitutive of the agent. In this response, I argue that Sher is wrong on every count. First, the account of moral responsibility in the absence of consciousness he advocates does not preserve control at all; rather, it ought to be seen as a variety of attributionism (a kind of account of moral responsibility which holds that control is unnecessary for responsibility, so long as the action is reflective of the agent's real self). Second, I argue that a realistic conception of agency, that recognizes the limited role that consciousness plays in human life, narrows the scope of moral responsibility. We exercise control over our actions only when consciousness has played a direct or indirect role in their production. Moreover, we cannot escape this conclusion by swapping a volitionist account of moral responsibility for an attributionist account: our actions are deeply reflective of our real selves only when consciousness has played a causal role in their production.

Keywords Control · Moral responsibility · Attributionism · Sher

Consciousness is puzzling. In the philosophy of mind, the problem of explaining how phenomenal consciousness might arise from neural processes is known as “the hard problem.”¹ In moral philosophy, too, there is a hard problem of consciousness:

¹Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

N. Levy (✉)
Centre for Applied Philosophy and Public Ethics,
University of Melbourne,
Melbourne 3010, Australia
e-mail: neil.levy@philosophy.ox.ac.uk

why and how might consciousness matter to moral responsibility? Recently, George Sher has argued, on the basis of a picture of human agency, surely correct, in which consciousness plays a relatively small direct role in the causal chain leading to morally significant action, that we ought to drop the requirement that agents must be conscious of what they are doing in order to properly count as morally responsible for their actions.² Sher's account is interesting, because as well as advancing a substantive thesis regarding the relationship between responsibility and consciousness, it might also usefully be understood as developing a view midway between two opposed accounts of moral responsibility, the accounts sometimes called attributionism, which stresses the extent to which an action is reflective of an agent's real self, and volitionism, which stresses choice and control.³ Sher is a volitionist inasmuch as he wants to preserve the intuition that responsibility requires control, but his account of what control consists in borrows heavily from attributionism.

In this brief response to Sher, I shall argue that his account of the relationship between moral responsibility and consciousness fails on all counts: it does not succeed in preserving the intuition which guides Sher, that we are responsible only for what we control, and it fails because it concedes too much to attributionism at the expense of volitionism. Moreover, Sher's central claim, that we ought to drop the requirement that agents must be conscious of what they do in order to be responsible for it, is false: morally responsible agency is (directly or indirectly) conscious agency. Sher is right to argue that we must develop a more realistic picture of agency, which accommodates the very significant extent to which unconscious processes play a part in action production. But this picture does not, as he thinks, expand the scope of moral responsibility; rather, it contracts it.

I

Sher builds his essay around a number of cases, involving agents who perform wrongful actions while apparently lacking control over what they do. The agents in Sher's cases suffer lapses in memory (like Alessandra, who is distracted by her children's tale of misbehavior and punishment and forgets that she has left her dog in the car on a hot day), and lapses in concentration (like Julian, the ferry pilot whose mind wanders from his routine task and whose boat ends up on the rocks). They are overcome by strong emotions (like Joliet, who shoots her son when her fear causes her to panic at the sound of an intruder), and make bad judgments (like Father Poteet, who misjudges the speed of an oncoming truck, causing a serious accident, and like Ryland, whose self-absorption prevents her from seeing that her anecdote is hurtful to members of her audience). In all of these cases, Sher claims, the agent

² Sher, G. (2006). Out of control. *Ethics*, 116, 285–301.

³ The terms "attributionism" and "volitionism" are mine, but the distinction itself is made by many moral philosophers. For a clear definition and discussion of the distinction, see Angela M. Smith (2005), "Responsibility for attitudes: Activity and passivity in mental life," *Ethics*, 115, 236–271. Of course, there is a sense of attributionism in which all sides accept that the debate is about attribution: the attribution of acts to agents and the attribution of responsibility to agents. Attributionists differ from volitionists inasmuch as they disagree about the conditions which must be satisfied for this kind of attribution to be appropriate.

would rightly be held responsible for the action and indeed blamed for it; in some, the agent would be liable to punishment.⁴ But the intuition that these agents are responsible for their actions is difficult to square with “the widely held principle that a lack of control excuses an agent from responsibility.”⁵ We do not, it seems, exercise control over our lapses of memory or concentration, over our mistakes in judgments, over what we fail to see.

Of course, there is a sense in which the agents in Sher’s cases control their actions. The problem arises because the sense in which they exercise control does not seem to be the relevant one. If they are responsible because of what they failed to do – their lapses of memory, concentration, and judgment – then it seems that they must have exercised control *over their failure*; over whether or not they brought about the action they failed to perform. Merely guiding the action they did perform does not constitute control over the counterfactual action they neglected, and therefore does not provide for control, in the responsibility-relevant sense. Hence Sher’s problem, what I shall call the relevant control problem: how to reconcile a voluntarist account of moral responsibility with this apparent lack of relevant control.

What to do? We might abandon the principle that responsibility requires control. This is the course of action recommended by attributionism. Attributionists hold that an agent is responsible for an act (or omission) just in case it is expressive of the agent’s real self.⁶ Attributionists typically advance their thesis by putting forward examples like Sher’s; cases in which we are disposed to blame agents for actions over which they apparently lacked control. These cases, they argue, support a “quality of will” thesis: an agent is responsible insofar as her action expresses her good or ill will toward those who benefit from or are harmed by it.⁷ Thus, for instance, they may say that agents who experience lapses in concentration or who fail to notice that their remarks are hurtful are responsible because their actions express their indifference to others: had they cared more for the victims of their actions, they would not have acted as they did. Sher is unwilling to go this route, and with good reason: the principle that responsibility requires control is strongly supported, as he suggests, by the role it plays in explaining our intuitions in a wide range of cases in which agents are excused responsibility.⁸

Rather than abandoning the principle, Sher recommends finding a way to accommodate it, somehow reconciling control with the apparent lack of consciousness or intention in the cases he sketches. He begins by noting that human beings ought not to be identified with “a center of consciousness and will.”⁹ We consist in far more than just the set of our conscious states, and our unconscious beliefs,

⁴ Sher, “Out of control,” pp. 288–290.

⁵ *Ibid.* p. 295.

⁶ Prominent defenders of attributionism include Robert Adams (1985), “Involuntary sins,” *Philosophical Review*, 94, 1–31; T. M. Scanlon (1998), *What we owe to each other* (Cambridge, MA: Harvard University Press) and Angela Smith, “Responsibility for attitudes.”

⁷ The clearest expression of the quality of will thesis, and the phrase itself, is found in T. M. Scanlon, “The significance of choice,” in *Free Will* (pp. 352–371), Ed. Gary Watson, Oxford: Oxford University Press.

⁸ Sher, “Out of control,” p. 295. I have advanced other considerations against attributionism in Neil Levy (2005), “The Good, the bad and the blameworthy,” *Journal of Ethics and Social Philosophy*, 1, 1–16.

⁹ Sher, “Out of control,” p. 296.

desires and attitudes play a causal role in producing our actions. Given that this is the case, Sher argues, we need a conception of control that accommodates unconscious, as well as conscious, mental states. If, he suggests, an agent's full set of beliefs and desires, conscious and unconscious, accounts for or explains their failure to recognize that an action is wrong, they exercise relevant control over it. This, Sher suggests, is true of the agents in his examples, and explains why we rightly hold them responsible for their actions. Despite appearances, they exercise control over their actions, and are therefore responsible for them.

Ryland, for instance, fails to recognize that her remarks are hurtful *because she is self-absorbed*. Whatever combination of mental states, conscious and unconscious, constitutes her self-absorption explains her wrongdoing. Similarly, if Alessandra had cared more for her duties, or for her dog, she would not have acted as she did. The set of conscious and unconscious mental states that constitutes her degree of caring was causally relevant to her act, and in that sense she exercised control over it. Joliet's tendency to panic makes her the person she is, and Father Poteet's misjudgment is rooted in his complex psychology. Since their mental states explain their actions, they can rightly be said to exercise control over them. They are therefore responsible for what they do.

There is something compelling, as well as something deeply puzzling, about Sher's suggestion. It is compelling insofar as it is surely true that our unconscious mental states play a causal role in *all* our actions, and that many of them are produced almost entirely without conscious thought. The majority, probably the overwhelming majority, of our actions are automatic, in the technical sense of the term current in psychology.¹⁰ Classically, automatic processes have been held to be effortless, ballistic (uninterruptible once initiated) and unconscious, but more recent researchers have shown that these features can dissociate: some automatic processes are consciously initiated.¹¹ Nevertheless, a great deal of ordinary human action is initiated and guided by unconscious processes. Given the prevalence of automaticity in human life, contracting the scope of moral responsibility to the sphere of consciousness would apparently contract it to a narrow band.

But Sher's suggestion is deeply puzzling as well. Suppose an agent satisfies Sher's condition for responsibility in the absence of consciousness; suppose, that is, that his failure to believe that he is acting wrongly is explained "by some subset of the other beliefs (desires, attitudes, etc.) that make him the person he is."¹² In what sense, however, does the fact that he satisfies this condition make it the case that he exercises *control* over his action? The relevant control problem arises, recall, because we do not exercise control over anything of which we are unaware. Sher's condition explains *why* the agents in his cases are not aware of the wrongmaking features of their actions, but explaining why they are not aware of these features explains why they do not exercise relevant control: it does not restore it. Sher's solution leaves the relevant control problem untouched.

¹⁰ Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462–479.

¹¹ Bargh, J. A. (1992). The ecology of automaticity. Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, 105, 181–199.

¹² Sher, "Out of control," p. 298.

There may be, as some psychologists have suggested, such a thing as unconscious control.¹³ But if there is, it is a matter of the guidance of behavior by representations, albeit unconscious representations, not (the wholly mysterious) guidance by the *lack* of such representations. In order for an agent to satisfy some kind of control condition, then, she must guide her behavior in the light of her mental states, and that requires far more than Sher's condition allows for.

In fact, Sher's condition is best read not as a volitional account of moral responsibility at all, but as an attributionist account. Agents who satisfy his condition, whose wrongful actions are explained by their mental states, conscious and unconscious, do not (relevantly) control their actions. These actions may, however, be (partially) expressive of their real selves as practical agents. It is because our actions which express our mental states are reflective of who we are that attributionists hold us responsible for them, even though (as they recognize) we do not control them. If attributionism is not viable, then Sher's account fails.

II

Sher's attempt to accommodate volitionism to the fact that our actions are often initiated, and even more often carried out once initiated, without the assistance of consciousness therefore fails. This leaves open the substantive question: should volitionists attempt such an accommodation? Or should they narrow the scope of moral responsibility to the sphere of consciousness? I shall argue that, counter-intuitive though it seems, volitionists ought to hold that responsibility requires a substantial role – direct or indirect – for consciousness, and that therefore responsibility is rarer, perhaps far rarer, than we think. Indeed, I shall go on to argue, attributionists too ought to accept this conclusion, at least so long as they believe that responsibility requires deep attributability. Actions are deeply attributable to agents only when consciousness plays a substantial role in their production.

Each of us has many mental states, attitudes and stereotypes which we have not endorsed, and some of which, were we aware of them, we would not endorse. Sometimes we become aware of the presence of such states, and as a result work, with more or less effect, toward eliminating them. We might achieve considerable success at this task, but we rarely entirely purge ourselves of these states. Consider, for instance, the large literature on racial and sexual stereotyping. A range of tests, such as the many variants of the Implicit Association Test, reveal that most white subjects have a preference for white faces over black faces, as measured by their relative reaction times. It takes longer for whites to associate positive terms with black faces than with white; conversely it takes them longer to associate negative terms with white faces than with black. These implicit associations are probably a more or less inevitable result of growing up in a culture in which negative stereotypes of blacks are common. Indeed, black Americans do not escape the

¹³ See, for instance, Moskowitz, G. B. (2001). Preconscious control and compensatory cognition. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 333–358). Hillsdale, NJ: Erlbaum.

effects of these stereotypes, though – unsurprisingly – they are more resistant to them than are whites. Black Americans show no ingroup favoritism on average, whereas white Americans do; instead they show much more variation in responses.¹⁴ Similarly, most subjects show a bias toward heterosexuals as against homosexuals, and males as against females; once again, gay men and women are sometimes subject to the same biases as heterosexuals, and females may show a bias toward males.

These stereotypes can be shown to alter behavior, leading whites to show a preference, for instance, to white job candidates over black. But they do not condemn us to acting in a racially biased manner. There are a number of things we can do to moderate the influence of these implicit associations. For instance, we can fight automatic responses by priming contrary responses: thinking of admired black individuals causes a significant reduction (though not the elimination) of the preference for white faces.¹⁵ Even being strongly motivated to avoid stereotyping has been shown to be effective in attenuating the effects of such preferences.¹⁶ However, there is overwhelming evidence that combating the effects of these unconscious beliefs requires attentional resources. Subjects under cognitive load (that is, who are required to perform another, simultaneous, task) cannot attenuate the effects of stereotypes on their judgments.¹⁷ In other words, the link between unconscious attitudes and behavior can be broken – but it requires consciousness to break it.

In the light of this, consider the difference in behavior of two psychologically identical agents, Eileen and Eleanor. Both of them are well-intentioned and highly motivated people, with a strong sense of duty. Neither of them could justly be described as racist: indeed, each has campaigned, contributing both time and money, against what they regard as ongoing discrimination against people of non-white descent. Each, however, exhibits a weak preference for white faces over black on the implicit association test (the result of having grown up in a society which, as they recognize, is racist). Aware of this fact, each makes an effort to combat their prejudice. By dint of reading black history, associating with black people and making an effort to judge people only on the basis of their relevant characteristics, they achieve considerable success: they reduce the extent of their preferences, as measured by the Implicit Association Test, and they prevent what small preference remains from impacting in any significant way on their behavior. But now consider what happens when their histories diverge. Suppose that, through no fault of her own, Eileen finds herself without the attentional resources to interrupt the link between stereotype and behavior. Whereas Eleanor is able to prevent her implicit

¹⁴ See Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17, 143–168 for a review of the relevant literature.

¹⁵ Dasgupta, N. & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.

¹⁶ Monteith, M. J., Sherman, J. W., & Devine, P. G. (1998). Suppression as a stereotype control strategy. *Personality and Social Psychology Review*, 2, 63–82.

¹⁷ Fine, C. (2006) Is the emotional dog wagging its rational tail, or chasing it? Reason in moral judgment. *Philosophical Explorations*, 9, 83–98.

attitude from corrupting her actions, because Eileen is under cognitive load – because her attention is divided, because the need to combat the stereotype is crowded out of consciousness by distractions, because she is tired or whatever the case may be – her behavior reflects her prejudice.¹⁸

In that case, Eileen has behaved in a morally wrongful manner, whereas Eleanor has not. Is she to blame for her action? By hypothesis, the only moral difference between the two agents is the product of external factors over which *neither* had any control. Eileen has acted worse than Eleanor, but through no fault of her own. Given that it is Eileen's bad luck, and only her bad luck, that she finds herself with a stain on her moral record, she is not responsible for her action.¹⁹ Since she is not responsible for the conditions that cause her alone to act badly, she is not responsible for the action that is the product of these conditions.

In the absence of consciousness we are at the mercy of automatic responses we are not responsible for acquiring, that we may consciously reject and which we may even have worked hard to eradicate. It is therefore unfair to hold us responsible for actions which reflect such responses when we cannot control them; thus, consciousness is necessary for moral responsibility.²⁰ This implies, as Sher recognizes, that we are responsible less often than we tend to think. But the counterintuitiveness of this conclusion can be softened somewhat by the recognition that though consciousness is necessary for responsible action, it need not be present at the precise moment of action. Many of our automatic action routines exist only because we have, consciously, inculcated them: they are learned behaviors that have become routinized, and are now reliably triggered by environmental stimuli. Moreover, many automatic responses are, in the jargon of psychologists, *goal-dependent*; they are triggered only when we are (consciously) engaged in relevant action.²¹ Since these responses reflect and are the product of our conscious choices, we are responsible for them, at least in ordinary circumstances. When our actions are the direct product of our conscious choices, or their indirect product via our choices to mould our character or to inculcate habits and responses in ourselves, we are responsible for them. When they are produced automatically and are neither the

¹⁸ The hypothesis that Eileen finds herself without the attentional resources to interrupt the link between automatic stereotyping and behavior is, it must be stressed, realistic. Conscious control is a slow, inefficient and relatively rare phenomenon. When we lack time, are under pressure or distracted, our behavior is guided by automatic mechanisms without the benefit of the flexibility that consciousness provides.

¹⁹ Indeed, Eileen might actually be worse off for trying to control her response. There is evidence that though under optimal conditions conscious control is effective, when attentional resources are limited attempts at control produce the very kinds of responses they aim to avoid. See Wegner, D. M. & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 446–496). New York: McGraw-Hill, at p. 474.

²⁰ That is not to say, of course, that agents may not be appropriate targets for moral criticism on the basis of racist attitudes. Such attitudes are morally abhorrent, and we are justified in calling for people to reject them. But moral criticism need not be moral blame; the fundamental attributionist error consists in confusing them. Elsewhere, I have argued that part of the reason for this confusion may consist in this: when we justifiably criticize someone for their moral attitudes, we often create conditions under which it is appropriate to blame them if they do not take steps to correct those attitudes. But we cannot infer from that fact that it appropriate to blame them before or at the same time as we criticize them.

²¹ *Ibid.* p. 469.

direct nor the indirect product of conscious endorsement, we are not responsible for them. We have not chosen to respond in the way we do, and therefore do not exercise control over our action, direct or indirect.²²

The inability of volitionism to accommodate the kind of cases that Sher highlights will be taken by some people as evidence in favor of attributionism: given that volitionist accounts of moral responsibility yield counterintuitive results, we should reject such accounts. In fact, however, attributionism is no help in vindicating our intuitive responses to these cases. Attributionists, too, ought to accept that consciousness is required for moral responsibility, at least if one prominent view of consciousness is correct. According to this theory, widely endorsed by philosophers, psychologists and neuroscientists, consciousness arises from or facilitates (depending upon the account) a *global workspace*, as Bernard Baars calls it.²³ A global workspace is a space in which information, accessed from many different sources – from modular brain systems and from the environment – become accessible to many different parts of the brain simultaneously. Consciousness serves the function of allowing parts of the brain that are otherwise relatively isolated from each other to communicate. Consciousness is or at least facilitates what Dennett calls “fame in the brain,” or “cerebral celebrity;”²⁴ in the terms made famous by Ned Block, phenomenal-consciousness facilitates access-consciousness.²⁵ The global workspace allows all the mechanisms constitutive of the agent, personal and subpersonal, conscious and unconscious, to contribute to the process of decision-making. Hence conscious deliberation is properly reflective of the entire person, including her consciously endorsed values. Whereas, in the absence of consciousness, my decisions reflect only some subset of the subpersonal mechanisms that constitute me (just as Eileen’s judgment reflects only a small portion of her psychology), when I deliberate consciously the resulting decision really reflects my real self. Hence, it is deeply attributable to me. Only when I choose are the conditions upon deep attributability satisfied: attributionism has volitionist conditions.

Sher is perfectly correct in asserting that if we limit the scope of moral responsibility to the clearing of conscious choice, we shall be responsible less often,

²² It is intuitive that we are not responsible for our automatic responses when they conflict with our (non self-deceptively) endorsed values and we are unable to prevent them. What, however, if our automatic response happens to be in line with our values, though there is no causal link between our values and our response? In that case, the conformity of the response to our values is merely accidental, and we deserve neither credit nor blame for its expression (unless we are responsible for the fact that we respond automatically).

²³ Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. New York: Oxford University Press. Other prominent global workspace theorists include Stanislas Dehaene and Lionel Naccache (2001), “Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework,” *Cognition*, 79, 1–37; Anthony I. Jack and Tim Shallice (2001), “Introspective physicalism as an approach to the science of consciousness,” *Cognition*, 79, 161–196; Daniel Dennett (1991), *Consciousness explained* (London: Penguin); David Chalmers, “Availability: The cognitive basis of experience?,” in *The Nature of Consciousness: Philosophical Debates* (pp. 421–424), Eds. Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, MA: The MIT Press.

²⁴ Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, 79, 221–237.

²⁵ Block, N. On a confusion about a function of consciousness. In Block, N., Flanagan, O., & Güzeldere, G. (Eds.) *The nature of consciousness: Philosophical debates* (pp. 375–415). Cambridge, MA: MIT.

perhaps far less often, than we tend to think. Even if we expand the scope of conscious choice in the ways I have suggested, by allowing that consciousness need not play a *direct* role in responsible agency, we shall probably discover that moral responsibility is a relatively scarce good. We ought, as Sher argues, to build our picture of moral responsibility around a realistic picture of human agency. When we do so, however, we discover that some of the people we are disposed to hold responsible are not. Alessandra, Ryland and the rest *may* be responsible; it's hard to tell on the basis of the information Sher provides. If they are, however, it will be because their responses are the product, directly or indirectly, of conscious choices, and not for the kinds of reasons Sher suggests. If they have not chosen to respond as they do, then they are not responsible. This may conflict with our knee-jerk responses to the cases that Sher presents, but if we follow Sher's advice, and base our conception of responsibility upon a realistic picture of human agency, we ought to be prepared for such surprises.

Acknowledgement I thank George Sher and two anonymous reviewers for *Philosophia* for very helpful comments that have helped me to improve this paper enormously.