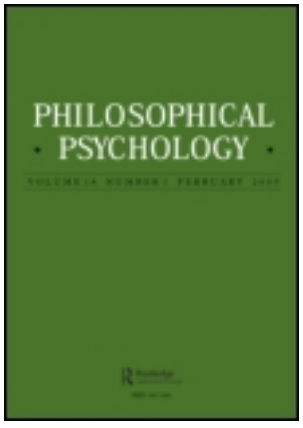

Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cphp20

Automaticity, Consciousness and Moral Responsibility

Simon Wigley

Available online: 23 Apr 2007

To cite this article: Simon Wigley (2007): Automaticity, Consciousness and Moral Responsibility, Philosophical Psychology, 20:2, 209-225

To link to this article: http://dx.doi.org/10.1080/09515080701197122

Philosophical Psychology
Vol. 20, No. 2, April 2007, pp. 209–225

Automaticity, Consciousness and Moral Responsibility

Simon Wigley

Cognitive scientists have long noted that automated behavior is the rule, while conscious acts of self-regulation are the exception to the rule. On the face of it, automated actions appear to be immune to moral appraisal because they are not subject to conscious control. Conventional wisdom suggests that sleepwalking exculpates, while the mere fact that a person is performing a well-versed task unthinkingly does not. However, our apparent lack of conscious control while we are undergoing automaticity challenges the idea that there is a relevant moral difference between these two forms of unconscious behavior. In both cases the agent lacks access to information that might help them guide their actions so as to avoid harms. In response, it is argued that the crucial distinction between the automatic agent and the agent undergoing an automatism, such as somnambulism or petit mal epilepsy, lies in the fact that the former can preprogram the activation and interruption of automatic behavior. Given that, it is argued that there is elbowroom for attributing responsibility to automated agents based on the quality of their will.

Keywords: Automaticity; Automatism; Consciousness; Illusion of Free Will; Responsibility

1. Introduction

An all too familiar experience occurs when we suddenly realize that we have been driving for some time and yet, despite that lack of conscious awareness, we have been able to change the car's gears, adjust its speed and steer without hitting another vehicle or going off the road. Typically we can only, at best, vaguely recollect what occurred whilst we were driving in that automated state. Cognitive scientists have long noted that the automatic performance of well-practiced tasks is pervasive. Automaticity is the rule, while conscious acts of self-regulation are the exception (Baars, 1988, pp. 74–86; Bargh & Chartrand, 1999; Palmeri, 2003; Wegner, 2002, pp. 56–59). The efficiency gains from this form of unconscious behavior are

Correspondence to: Simon Wigley, Faculty of Humanities and Letters, Department of Philosophy, Bilkent University, Bilkent, Ankara, 06800, Turkey. Email: wigley@bilkent.edu.tr

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/07/020209-17 © 2007 Taylor & Francis
DOI: 10.1080/09515080701197122

considerable. Typically virtuoso tasks can be performed significantly quicker than novel tasks that require conscious control. Moreover, automaticity enables mental processes to operate in parallel, whereas it is not possible to carry out more than one consciously controlled process at the same time (e.g., having an absorbing discussion with a passenger while driving in busy traffic). Thus, automaticity frees up cognitive resources that are needed in order to perform novel tasks that demand conscious regulation.

The question that this paper is concerned with is what that means for the attribution of moral responsibility. On the face of it automated actions and their consequences appear to be immune to praise and blame because they are not the product of conscious control. How, therefore, are we to distinguish the moral responsibility of the person who is undergoing automaticity from that of the person who is, say, sleepwalking? Common sense suggests that somnambulism absolves a person of responsibility, while the mere fact that a person is performing a well-versed task "without thinking" does not. But our apparent lack of conscious control while we are undergoing automaticity challenges the idea that there is a relevant moral difference between these two forms of unconscious behavior.

By way of illustration consider the following four vignettes:

- Zelda is driving automatically and hits a child who suddenly steps out onto a pedestrian crossing.
- Zara is driving automatically and successfully avoids hitting a child who suddenly steps out onto a pedestrian crossing.
- Bea, who is a somnambulist, wakes up with the vague recollection that she has struck a child on a pedestrian crossing while driving her car.[1] When she calls the police she realizes that she was involved in a hit-and-run accident during the time she was asleep.
- Bella, who suffers from petit mal epilepsy,[2] has a seizure while driving home from work and strikes a child on pedestrian crossing.

I assume, for the sake of explanation, that each driver would have avoided hitting the child if they were driving attentively. Moreover, each of the drivers is experienced in the sense that, as a result of persistent task repetition, they are able to perform the complex activity of driving unconsciously (for a detailed study of automatic behavior during driving, see Groeger, 2000, ch. 4). However, by definition their routinized behavior is inflexible in that, in itself, it is unable to respond to unexpected circumstances such as a child suddenly stepping off the curb. Thus, when it comes to attributing moral responsibility, we may ask whether there is a relevant difference between the person who may harm others whilst undergoing *automaticity* (Zelda and Zara) and the person who may harm others whilst undergoing forms of *automatism* such as somnambulism and petit mal epilepsy (Bea and Bella). For, during both these forms of unconsciousness the agent is unable to adapt their habitual behavior so as to take into account the occurrence of novel or unpredictable events.

It should be made clear from the outset that in this paper I am primarily concerned with the harms that occur as a result of unconscious behavior that is

over-practiced (e.g., driving to work whilst undergoing somnambulism), rather than unconscious behavior that is unpracticed (e.g., killing a family member whilst undergoing somnambulism). There is an increasing body of literature that is engaged with novel behavior that occurs during somnambulism (e.g., Denno, 2003; Levy & Bayne, 2004; McSherry, 1998; Schopp, 1991; Yeo, 2002). Here, however, I shall be focusing on the harms that arise because of the non-adaptive nature of routinized unconscious behavior. In order to illustrate the discussion I will tend to refer to the activity of driving because of its familiarity and because of the risk of injury that is posed by it. However, the discussion is intended to be applicable to all instances of the unconscious triggering of habituated behavior (including, e.g., aggression, dishonesty, the evaluative stereotyping of a particular group, etc.).

In the following section I clarify the sense in which agents undergoing automaticity and automatism are unconscious (§2). I argue that the nature of consciousness does not enable us to pinpoint the sense in which the automatic agent (such as Zelda and Zara) can be held responsible while the automatistic agent (such as Bea and Bella) cannot be. Both types of agent are characterized by a lack of rational control over their actions, and so, neither appears to be open to moral appraisal. I then turn to consider whether the distinction between these two forms of unconscious behavior derives from the fact that the automatic agent's lack of rational control is a product of her volition. More precisely, I examine whether the automatic agent can choose to activate and interrupt her automatic behavior (§3). I argue that only a modification of that position will enable us to distinguish the automatic agent from the automatistic agent (§4). Elbowroom for moral appraisal emerges if we focus on the person's control over the acquisition and revision of habituated behavior. In the penultimate section I consider and respond to the challenge of the idea of intentional preprogramming and revision that is posed by the claim that our actions are determined preconsciously (§5). I conclude by noting how control over the acquisition of automaticity means that we need not, as some have suggested, abandon the idea that quality of the will is a necessary condition for responsibility (§6).

2. Consciousness and Ignorance

I take it that a necessary condition of moral responsibility is that the decision-making of the agent must be adequately informed (for the classic statement of this line of argument see Aristotle, 1998, bk. 3, ch. 1, 1110b20–1111a20). That is, she must be conscious in the sense that she has access to the information that is necessary to guide her actions so as to avoid harm, or so as to do good. With that prerequisite for blameworthiness and praiseworthiness in mind what I will suggest is that we cannot differentiate the responsibility of the automatistic agent and automatic agent based on their "state of mind" at the time when the harmful or good event occurred.

In the case of automaticity, the periphery (e.g., road and traffic conditions) is not the focus of the driver's attention—nevertheless, she is sufficiently aware of the

periphery to ensure that she does not crash the car. I take this to mean that the person undergoing automaticity is nonconscious in the sense that they are not attending to the periphery's perceptual content. That is to say, they are not using that information very much in their reasoning or in order to guide their actions or, as Ned Block (1997, pp. 399–400) puts it, they lack access consciousness.

The potential excuse available to Zelda, therefore, is that she was not processing the information necessary to guide her actions. Similarly moral praise may not be attributable to Zara if she was not accessing the information necessary to avoid hitting the child. She acted ''without knowing what she was doing.'' If that is the case, then the fact that she avoids hitting the child, while Zelda does not, is nothing more than an instance of moral luck in the ways things turn out (Nagel, 1979, pp. 28–32).

Notice how this account of automatic behavior does not rely on whether or not we are conscious of ourselves as exerting rational control over our actions. That is in spite of the fact that it is introspective consciousness that is typically being referred to by the common sense idea of conscious control. According to that common sense view, to say that a person ''knowingly'' acted in a certain way, or omitted to act in certain way, is to say that they were reflexively aware of what they were doing. However, while introspective consciousness may co-occur with access consciousness, it is by no means clear that introspection is necessary in order to guide speech and action (Block, 2001, p. 206). Below (§5) I discuss whether introspective consciousness can play a causal role or whether it is merely epiphenomenal. There I argue that while introspection may provide us with the phenomenology of control it is not necessary for the rational control of our actions. Thus, for an action to qualify as consciously controlled it is only necessary that the agent possessed sufficient access consciousness. In other words, novel or flexible behavior only requires rational control of speech and action based on input from the periphery. The presence of introspection during moments of conscious control, thus described, only serves to provide us with the feeling of control. Similarly, the co-occurrence of diminished introspective consciousness alongside diminished access consciousness at most helps to explain what it is like to experience automaticity and automatism, namely, the sense of dissociation from one's own actions. Thus, attempts to explain automaticity in terms of the absence of inner perception (Armstrong, 1997, pp. 723–724) or a higher-order thought (Rosenthal, 1997, p. 743) are impoverished if they ignore the agent's accompanying lack of access consciousness.

What bearing does this account of the consciousness have on the attribution of moral responsibility to agents undergoing an automatism and automaticity? Bea and Bella, our somnambulist and petit mal drivers, appear to be unconscious in exactly the same way as Zelda and Zara. For in both cases the driver has sufficient access consciousness to stay on the road, but is otherwise driving based on over-learned motor routines. If that is the case there seems little to differentiate the automatistic driver from the automatic driver. Both are not attending to the periphery, and so, both are equally ignorant of the information needed to modify their behavior so as to avoid hitting the child.

In response to that claim it may be argued that there are at least three different ways in which there is a relevant moral difference between automaticity and automatism. In the first place, it may be pointed out that a driver undergoing automaticity can pay attention to other things (e.g., reading a road map, talking on a mobile telephone, etc.). By contrast the driver who is undergoing somnambulism or a petit mal seizure appears to be unable to attend to anything, let alone driving. But the fact that they appear to be globally rather than locally unconscious has no bearing on their moral responsibility. The point is that both types of unconscious agent lack access to the particular information that is required to respond to the unexpected.

In the second place, it may be argued that while both the automatic driver and the automatistic driver are cognitively deficient with regard to the periphery, only the former is phenomenally conscious of the periphery—what it feels like to sense the black-and-whiteness of the pedestrian crossing and the redness of the child's jacket and so on. There are at least two problems with that suggestion. Firstly, there appears to be no reason to suggest that the automatistic agent is completely devoid of phenomenal consciousness. Indeed it seems that empirical examples of access consciousness—albeit diminished in the case of the sleep walker or petit mal epileptic—in the absence of phenomenal consciousness do not exist (Block, 1997, pp. 397–398, 402). Insofar as that is correct, a more plausible explanation is that both kinds of unconscious agent possess a modicum of access *and* phenomenal consciousness. Secondly, I take it that experiential properties play no direct role in reasoning and the control of speech and action (Block, 1997, p. 401). That is to say, taking preventative measures when unexpected situations arise requires reasoning based on the perceptual content that characterizes the periphery. Thus, in terms of responsibility, the relevant similarity between the two cases holds whether or not phenomenal consciousness based on the periphery is present.

Finally, it may be argued that automaticity is better explained by attention to the periphery coupled with rapid memory loss. That is, unlike the automatistic driver, the automatic driver does in fact have access to the information required for guiding action at each moment in time, but does not put it into memory. If that is the case, then Zelda can not plead ignorance if at a previous point in time she was aware of a sign that said "SCHOOL AHEAD" that she then omitted to store in memory. Undoubtedly automaticity is characterized by memory loss. However, that is consistent with the possibility that the information that may or may not be put into memory by the automatic driver is impoverished due to the fact that she is not attending to the periphery. That is, automaticity may be characterized by inattention followed by memory loss. It is difficult to see how we can determine which of these interpretations—attention coupled with rapid memory loss or inattentiveness coupled with memory loss—provides us with a more accurate explanation of automaticity.[3] More significantly for our present purposes, neither of those accounts permits us to distinguish the responsibility of the automatic agent from that of the automatistic agent. For it may be the case that, as with their automatistic counterparts, automatic agents are unable to inhibit inattentiveness or memory

loss (e.g., the driver undergoing automaticity may not be able to prevent herself from forgetting the sign "SCHOOL AHEAD" before her vehicle reaches the school). If that turns out to be correct then it remains the case that both types of unconscious agent lack access to the information necessary to adapt their actions so as to avoid harm.

The upshot of this is that we cannot differentiate the culpability of the automatic and automatistic agent simply based on their prevailing "state of mind" when the accident occurred.[4] Rather we must turn to examine the extent to which the agent has control over their cognitive deficiency—i.e., whether or not they can choose to attend to the periphery or store the contents of the periphery in memory. If they did have control over the extent to which they are guiding action based on the periphery, then there is room for evaluating the quality of their will. Analogously, a person may be responsible for the harms that occurred whilst she was driving under the influence of alcohol because she could have avoided her temporary cognitive impairment by refraining from drinking too much (Aristotle, 1998, bk. 3, ch. 5, 1113b30–1114a).

3. Quality of the Will

It may be argued that responsibility is attributable to the automatic agent because her behavior is consistent with her action plans. Thus, for example, an agent may decide to drive to work, but that intention is then executed automatically by a menu of over-learned routines such as gear changing, speed control, steering and so on. For Searle (1983, pp. 84–85) this constitutes an example of "intention in action." The first problem with this suggestion is that it does not permit us to differentiate the petit mal epileptic from the person who is behaving automatically. For, in both cases the agent has a prior intention to achieve an end, which is subsequently pursued unconsciously. The second problem is that the harm that we are concerned with here arises because of the inflexible nature of automatic behavior, rather than the agent's decision to initiate an action plan. We should, therefore, be focusing on the agent's control over whether the action plan was executed automatically and not the action plan itself. That is to say, we should examine whether the ignorance that arises because of automaticity reflects the careless or reckless willing of the agent (Fischer & Ravizza, 1998, pp. 87–89; Wallace, 1994, pp. 138–139). Degree of control over the activation and interruption of unconscious behavior, therefore, may be what distinguishes the automatic agent from the somnambulistic or petit mal agent. With that line of argument in mind, I now turn to consider whether we do in fact have control over the instigation and disruption of automaticity.

3.1. Control Over Activation

In the case of Zara, praise may be due if, given her assessment of her own learnt ability and the pre-existing driving conditions, she made the correct decision to drive

automatically. Similarly blame may be attributable to Zelda if her prior decision to drive automatically was, given her learnt ability and the driving conditions, ill-judged. The primary problem with this line of argument is that we do not intentionally choose when automaticity is activated (Bargh & Chartrand, 1999, pp. 471–474; Palmeri, 2003). Automated behavior is triggered by the presence of particular stimuli, rather than intentional choice. A good example of this in experimental psychology is the classic Stroop-effect (Stroop, 1935; for a review see MacLeod, 1991). Naming the ink color of a word takes much longer when the word means a different color (e.g., the word 'green' written in the red colored ink). The standard explanation for this is that the semantic content of words is automatically processed in virtue of the fact that reading is an over-practiced activity. More significantly for our present purposes, an expert reader is unable to stop herself from reading a word even when she has been explicitly instructed not to do so. By extension, the fact that Zelda initiates automatic driving when it is inappropriate to do so is beyond her control. Similarly, the fact that Zara switched to automatic pilot when it was appropriate to do so is beyond her control. In this regard there is little or no difference between the automatic driver and the automatistic driver. Both do not choose to go into automatic pilot.

Nevertheless, we may still refer to the driver's learned ability insofar as it minimizes the chances of an emergency situation arising in the first place. Thus, if a person knows that the activation of automaticity is itself automatic, then they are negligent if they do not take sufficient care over how they preset their behavior. The driver's control over preprogramming is, I contend, the right basis for apportioning responsibility. However, a lot more work needs to be done before we can arrive at a satisfactory account of it. The problem with basing responsibility on the quality of each person's prior training is that it only seems to apply when there are no unexpected or novel events. When the driving conditions are normal then an automatic driver who has trained themselves well should be able to avoid having an accident. However, because our capacity to process new information during automaticity is limited, automatic behavior in itself is unable to adapt to the occurrence of an unexpected or novel event. Both the well-trained and poorly-trained driver require attention to the periphery in order to adapt to such circumstances. At most, automaticity will place the well-trained driver in a situation where they are less likely to be affected by the unexpected (e.g., keeping sufficient distance from surrounding cars, driving at a safe speed, etc.). Thus, even if we suppose that activation is intentional, it seems that the non-adaptability of automatic behavior precludes the attribution of praise and blame when the unexpected occurs. Notice also that quality of training does not appear, at first glance at least, to provide us with a way to differentiate automatic and automatistic agents. The inherent inflexibility of oft-practiced routines means that, whilst they remain in their unconscious state, neither kind of agent can modify their behavior in response to novel or unexpected circumstances.

3.2. Control Over Interruption

We still have not found an adequate way of establishing the common sense distinction between automatic behavior and automatistic behavior. Given that the activation of both forms of unconsciousness is uncontrolled, perhaps that distinction can be established based on the fact that automatic behavior is susceptible to interruption. Thus, when the road conditions unexpectedly change for the worse, it may be the case that the automatic driver can interrupt her unconscious behavior and attend to the periphery. If correct, then although the automatic agent may be as cognitively deficient as the automatistic agent, she can rectify that deficiency. By way of illustration, if a piano wire snaps during a recital the virtuoso pianist who is undergoing automaticity would stop once she hit the dead key. If they were undergoing a petit mal seizure, however, they would continue playing in spite of the missing note. Their performance is ballistic in that it cannot be disrupted and, therefore, runs to completion. If the disruption of automaticity can occur, then there appears to be enough elbowroom for the attribution of moral responsibility (Levy & Bayne, 2004, pp. 213–214). Accordingly, while automatic behavior in itself may be insufficiently flexible, the agent can adapt according to novel circumstances if and when interruption takes place.

Interruption appears plausible once we remind ourselves that inattention to the periphery does not entail the complete absence of access consciousness. Indeed there is a body of evidence that suggests that attention can be gained automatically *even* when we are consciously preoccupied with another task (e.g., while we are engaged in a conversation the sound of one's own name or a baby crying from outside of the conversation may breakthrough) (Baars, 1988, pp. 34–35, 306–307; Moray, 1959). It seems, therefore, that we can, to some extent at least, unconsciously process the meanings of words in the unattended channel (e.g., a road sign whilst we are engrossed in conversation with a passenger). In this regard the automatic semantic processing entailed by the Stroop effect may prove to be particularly useful when it comes to automatic tasks such as driving. For road signs may lead to the interruption of automaticity both because skilled readers automatically process the semantic content of words and because they may contain words (e.g., 'Danger!', 'Stop!', etc.) that trigger attention even if we are consciously engaged with another task. In effect this entails the parallel processing of three layers of automaticity: namely, driving, reading and word recognition.

Aside from the peripheral informational content (including road sign words) it remains possible that the phenomenal content of the periphery (e.g., sensory quality of a red traffic light or stop sign) may also help to trigger attention. That is to say, phenomenal consciousness may play an indirect causal role in guiding speech and action in the sense that it may initiate access consciousness. Nevertheless, empirical evidence suggests that word signs constitute a more effective warning signal than color or shape signs. Thus, for example, when a person is presented with a triangular road sign with the word 'Stop' written in red, the word will be processed before the color, which in turn is processed before the shape. Although the latter priority can be

reversed with training, it does not appear that the processing priority of word signs over color or shape signs can be (Palmeri, 2003, pp. 11–12, 14–15).

The problem here is that, as with activation, interruption is the product of environmental stimuli rather than executive control. I cannot, for example, prevent myself from attending to the sound of my name coming from outside of the conversation that I am engaged in. Hence, control and, therefore, responsibility is only relevant at the point where the agent uses the peripheral information that interruption happens to grant them access to in order to guide action. The fact that the agent has access to that information in the first place, however, appears to be beyond her control.

4. Relocating the Site of Responsibility: Preprogramming and Revisability

The upshot of the discussion so far is that unless access consciousness is triggered by stimuli there is no room for morally appraising the person in an automatic state. Praise or blame can only affix to the way the agent uses the information they are fortunate to have access to. If an agent's attention is not triggered when an unexpected or novel event occurs, then, as with the case of automatism, there is no room for praise or blame.

Is there a way to extend the scope within which we can praise and blame the automatic agent? Is there a way to sharpen the distinction between the automatic agent and the automatistic agent? There is, I believe, if we accept that we have control over the way in which we inculcate our automatic behavior. Thus, during the learning of an activity we can willingly attune ourselves to those stimuli that trigger the interruption of automaticity (e.g., attuning ourselves to latch onto the content of road signs). Equally, we can attune ourselves to those stimuli that activate automaticity when the conditions are right. Thus, praise and blame can be based on the care that the agent took over the way they preprogrammed their automatic behavior. The assumption here is that the agent has sufficient access consciousness during the initial stages of the acquisition of automatic behavior. This claim is consistent with the process of acquisition itself gradually becoming automated (Bargh & Chartrand, 1999, 470–471; Bargh & Ferguson, 2000, pp, 933–934).

We can broaden the scope for moral appraisal even further if it is the case that we can intentionally revise our automated behavior once it has been inculcated. Revisability would also allow us to include as eligible for moral appraisal behavior that was not intentionally acquired in the first place, such as the evaluative stereotyping of a particular group (e.g., elderly, professors, African Americans, etc.) and the automatic moral evaluation of the actions of a person (Ferguson & Bargh, 2004; Haidt, 2001). Moreover, the ability to modify habitual behavior would also counteract the diminution of moral responsibility due to the circumstances in which the behavior is acquired (e.g., having learnt to drive in a country where they drive on the other side of the road, acquisition that begins during childhood, being raised by violent parents, etc.). Susceptibility to reprogramming, therefore, would mitigate

against the excuse that the agent lacked control during the acquisition of her habitual behavior.

The basic claim I am making here is that we do have rational control over our automatic behavior insofar as we can preset or revise the way it will function. It follows that the inappropriate activation of automaticity, or the failure to interrupt it when it is necessary to do so, constitutes blameworthy events. The fact that we do not, at the time when they occur, willingly trigger activation or interruption does not absolve us of responsibility for those harms that occur while we are undergoing automaticity. Hence, returning to our initial examples, Zelda and Zara may be the proper subject of blame and praise, respectively, if they could have preprogrammed or revised the triggering of activation and interruption so as to avoid hitting the child. Similarly, the police officer who acts according to automatic racial stereotypes is open to blame if she could have reformed her unconscious behavior. The interesting upshot of this is that an automatic agent is praiseworthy or blameworthy not because of the immediate actions that led to a good or bad outcome, but rather because of what they did, or omitted to do, in the past.

Note, however, that this approach will not allow us to ground responsibility in those cases where the automatic behavior is both unintentionally acquired and unrevisable. The attribution of praise and blame in such cases would appear to hinge on the ability of the agent to inhibit her speech and action (e.g., the ability of the employer to stop him or herself from hiring based on racial stereotypes). But as we have seen, the activation of automaticity is triggered by the environmental situation, rather than via conscious control. Thus, moral appraisal in those cases where the agent could not have done otherwise may depend on whether she endorses her automatic behavior upon reflection (i.e., based on sufficient access consciousness); that is, whether she would have behaved that way even in the absence of automaticity (Frankfurt, 1988).

Observe, however, that an appeal to higher-order volitions will also render the automatistic agent eligible for moral appraisal (e.g., a person who drives to his parents-in-law's house and kills them while sleepwalking may in fact have explicitly planned to do so while wakeful). In terms of moral responsibility, therefore, there seems to be very little that differentiates the agent who is subject to an uncontrollable automaticity from that of an agent who is subject to an automatism. Indeed the only relevant difference between them is that the automatic agent can regain control if and when interruption is triggered.

5. Actual Control and the Phenomenology of Control

In the discussion so far I have presupposed that moral responsibility requires both that the agent has access to the information necessary for guiding action and has control over whether or not they have that access. I have argued that the automatic agent, but not the automatistic agent, does have control over access in the sense that they can intentionally preprogram or revise when activation and interruption is to

occur as well as a menu of situation-action responses they have available to them. I am assuming, therefore, that the agent has control over the way in which her automatic behavior is preset or revised and the way in which she responds when her automatic behavior is interrupted.[5] But she does not have control over activation or interruption itself because those events are stimulus-driven. It follows that we can assess the quality of the automatic agent's will by tracking back to the acquisition of automatic behavior and forward to her response once her automatic behavior is interrupted.

There will be some, however, who will argue that this still does not permit us to differentiate the automatistic agent from the automatic agent. For them what we have in the case of preprogramming, revisability and evasive action after interruption is the phenomenology of control, rather than actual control. By way of illustration consider Spinoza's (1674/1955, letter 62, p. 390) example of a stone that is set in motion by an external cause. If the stone was suddenly able to think about the fact that it is in motion it might very well conclude that its continued motion is the product of its own wish. Similarly, the driver who takes evasive action when her automatic behavior is interrupted may believe that that action resulted from her conscious intention. Following Spinoza, however, it may be argued that the cause of the driver's action preceded her thought about the upcoming evasive action. In other words, both the triggering of interruption by environmental stimuli and the guiding of action based on the attention that is then paid to the periphery, occurred before the emergence of an intention in the driver's mind. Benjamin Libet (1985), for example, has produced experimental evidence that lends support to the claim that the brain activity that leads to action predates our conscious intention to act. I take this to mean that cognitive activity based on access consciousness precedes the agent's conscious thought about the upcoming action. As Daniel Wegner (2002, ch. 3) puts it, we interpret that thought as an intention that causes the action, even though it is nothing more than a preview of what is about to happen. Hence, we are left with the feeling of willing even though the action (and the preview itself) was preconsciously determined by underlying causal mechanisms. Nevertheless, from the constant conjunction of thought followed by action we are inclined to mistakenly infer that the former caused the latter.

It should be made clear that my limited concern here is to reply to the specific claim, eloquently put by Wegner, that the presence of preconscious brain activity should lead us to conclude that control is illusory. Insofar as my argument is persuasive it will simply show that the presence of mental activity that we are not introspectively aware of does not, by itself, rule out the possibility of voluntariness. It will follow that Wegner's argument does not provide us with sufficient reason to reject the claim that the agent has control over the presetting and revision of her automatic behavior.

According to Wegner's account, both the person who is undergoing automatism and the person who is undergoing automaticity lack the phenomenology of control. In other words, they are not aware of the preview of the impending action. In the

case of automaticity, where the activity is by definition efficient, there is insufficient time for the agent to become aware of the preview. However, when the activity is novel (e.g., learning to drive or responding to the unexpected), we experience the feeling of control because there is a sufficient temporal delay for us to become aware of the preview (Wegner, 2002, pp. 97–98, 2005, pp. 23–24, 27–30). Accordingly, consciousness does not actually play a causal role in the case of novel tasks (i.e., tasks that require serial rather than parallel processing)—it is just that we are conscious of the preview in such cases.

It should be noted, however, that automaticity provides minimal support for the view that, in the case of novel tasks (e.g., learning to drive) consciousness arises after the fact. The unintended upshot of Wegner's analysis is that the only relevant difference between automatic and novel behavior is the speed at which they are executed (Wegner, 2002, pp. 56–59). But clearly the crucial difference between these two modes of behavior is the degree to which they require rational control. Novel behavior is not simply a slower version of automaticity. That is to say, the fact that novel behavior demands more cognitive resources means it is not analogous to an agent who is undergoing automaticity suddenly having a preview of what is about to happen. Thus, because automaticity is not sufficiently similar to novel behavior it only provides minimal support for the claim that conscious willing is illusory. Notice also that the Libet experiments are afflicted by the same shortcoming because they only require the participant to exhibit a modicum of rational control (i.e., move a finger). Providing evidence that cognitively undemanding actions are initiated preconsciously does not allow us to conclude that cognitively demanding actions follow the same causal sequence. I take it, therefore, that it is only unpracticed behavior that occurs whilst a person is undergoing an automatism (e.g., unroutinized actions during a somnambulistic episode) that might provide telling evidence in support of the claim that cognitively demanding behavior is preconsciously determined.

If Wegner's line of argument is correct, then we cannot attribute responsibility based on the quality of the agent's will because she does not choose how her automatic behavior is preset or what evasive course of action should be taken when the unexpected occurs. However, as Daniel Dennett (2003, pp. 236–239, 252–253; see also Rosenthal, 2002, pp. 216–217; Zhu, 2004, p. 317) argues, the problem with that account of thought and action is that it presupposes that willing can only occur at the moment when the agent is introspectively aware of the, supposedly, already initiated action. Hence, the agent can at most veto a course of action that was initiated pre-introspectively. Mental activity prior to that preview is deemed to be an alien or external cause, in the same way as the motion of the stone in Spinoza's example was not caused by the stone itself. Thus, the presumption that choice-making by the agent depends on inner awareness rules out the possibility of agency prior to introspection. In other words, it does not countenance the possibility that decision-making is spread out over time, rather than squeezed into those moments when we happen to be introspectively aware that it is taking place. There may be

prior preparation for conscious decision-making in at least two different ways: (a) pre-introspective reasoning, or, if you like, access consciousness without introspective consciousness; and (b) the preprogramming of a menu of well-rehearsed responses that can be quickly selected from and initiated once automatic behavior is interrupted (e.g., the driver's preprogrammed set of evasive actions). The presence of pre-introspective reasoning and pre-commitment means that we are not simply left in a position where we can only begin to control an upcoming action once we become introspectively conscious of it. The upshot of this is that responsibility need not be based solely on a temporally isolated moment—e.g., the moment in which the driver reflexively considers what evasive action to take.

Nevertheless, responsibility is still attributable even if we insist that it should affix to introspective consciousness, rather than simply access consciousness. For we can still refer to those occasions in which the agent is aware that she is using reason to guide speech and action. Namely, during the (a) preprogramming of activation and interruption, (b) preprogramming of a preset menu of action responses, and (c) choice of action once automaticity is interrupted. In other words, responsibility can track those instances in which we do experience a sense of control or authorship. Indeed, we might wish to adopt such an approach given the epistemic problems associated with trying to attach responsibility to decision-making that is non-introspective and, therefore, cannot be reported. While introspection may be merely epiphenomenal, it seems to be the only means by which we can infer the presence or absence of agency. Hence, with the exception of those rare instances when it misleads us, the phenomenology of control provides us with a reliable indicator of the authorship that occurs both prior to and during moments of introspection (on the fallibility of introspection, see Wegner, 2002, pp. 74–78, ch. 4, pp. 195–201). According to this account the belief that introspection is playing a causal role is an illusion, but it is not a detrimental illusion because introspection does typically provide us with a reliable guide to agency.

Authorship, according to the account presented here, is attributable to those episodes of mental activity where there is sufficient access consciousness. During automaticity and automatisms there is a lack of authorship, as indicated by the lack or complete absence of introspective consciousness. In other words, diminished introspective awareness both indicates depleted rational control and helps (as Wegner suggests) to explain what it is like to experience automatistic and automatic behavior. However, as we have seen, automatic behavior is a product of intentional preprogramming and revisability, and so we can trace authorship back to those moments when introspection indicates that the automatic agent possessed sufficient rational control—e.g., when the novice decided which environmental cues should trigger activation and interruption. By contrast, authorship, and therefore responsibility, cannot be attached to those undergoing automatisms such as somnambulism and petit mal epilepsy because preprogramming or revision could not have prevented their lack of rational control.

6. Conclusion

The philosopher of law Peter Cane (2002, pp. 100–102) has argued that the pervasiveness of automaticity should lead us to conclude that responsibility cannot be attributed based solely on the quality of the agent's will. Indeed, we have seen how the decision-making of a person who is automatically performing an over-practiced task is insufficiently informed when unexpected or novel circumstances arise. That is, their choices are not sufficiently conscious because they are not using the peripheral information very much in order to guide action. We have also seen that, because the activation and interruption of automaticity is triggered by environmental stimuli, the agent does not have immediate conscious control over whether or not they attend to information from the periphery. Hence, the fact that a person's behavior does or does not lead to a harmful or good outcome whilst they are undergoing automaticity appears to be beyond their control. Given that, as well as the ubiquity of automaticity, there appears to be little scope for attributing praise and blame according to the conscious control of the agent. Put differently, quality of the will does not appear to provide us with a way to make the common sense distinction between the responsibility of the person undergoing automaticity and that of a person who is undergoing a sleeping or epileptic fugue. Neither type of agent has sufficient (access) conscious control over their automatic behavior.

The apparent implication of this is that we should not insist on making conscious control a necessary condition for responsibility. For, the agent's unconscious state means that we do not have a basis for evaluating the quality of her will. However, or so it might be argued, that does not entail that she is absolved of responsibility for the burden (e.g., compensating the innocent victim or her family) that results because of her actions. Her lack of rational control may mean that she is not an appropriate target for blame (or praise if things happen to turn out well), but it does not necessarily excuse her of responsibility for the consequences of her actions. All that matters, according to that outcome-based account of responsibility, is whether the agent's behavior was consistent with an objective standard of care (for an extended discussion of that line of argument, see Ripstein, 1999, especially pp. 84–87, 106–107, 268–269, 288–291). Thus, the fact that an agent was undergoing automaticity at the time that she failed to meet the standard of care does not excuse her of responsibility for the harmful outcome.

In response, I have argued that quality of the will remains relevant because we do have control over the presetting and revision of automatic behavior, as well as our response once automaticity is interrupted. In this sense we do have control over accessing and processing the information necessary for guiding action so as to avoid harmful consequences. While we lack immediate control whilst we are behaving automatically, we do have control over the measures that we should preemptively build into our automatic behavior. Hence, the fact that an automatic driver hits a child on a pedestrian crossing whom they would not have hit if they were driving attentively, indicates a failure to take due care over the way in which they acquired the skill of driving. Equally, the police officer who arrests the wrong person because

of unconscious stereotyping is blameworthy if she could have preemptively revised her automatic behavior. And so, the fact that both these agents were acting ''without thinking'' when the harm occurred does not render them beyond moral appraisal.

Acknowledgments

I would like to thank Josh Cowley, Ilhan Inan, Neil Levy, Emre Ozgen, Stephen Voss and Bill Wringe, John Doris and one anonymous referee. An earlier version of this paper was presented to the Joint Session of the Aristotelian and Mind Association, Manchester (UK), July 7–9, 2005. I would also like to thank the audience on that occasion for their insights.

Notes

[1] A person undergoing this form of somnambulism enters into a fugue state whereby they are able to proficiently carry out complex tasks such as driving, mowing the lawn, etc. See Mahowald and Schenck (2000).

[2] During petit mal or minor epilepsy the person suffers a loss of consciousness and amnesia, but does not undergo convulsions. In spite of the seizure they are able to continue driving home, walking or playing the piano, albeit in a routinized way. See Penfield (1975, pp. 38–40) and Searle (1994, pp. 107–108).

[3] According to Dennett's (1991, pp. 141–142) ''Multiple Drafts'' account of consciousness, both explanations—not attending and forgetting—are equally viable.

[4] In addition, for example the physiological difference between these two forms of unconsciousness may not be as great as might be assumed. There is, for example, evidence that suggests that wakefulness, non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep are not always mutually exclusive states. Neurophysiological features of one state may intrude into another state causing parasomias such as somnambulism (Mahowald & Schenck, 2000, pp. 322–323, 325–326). It may be the case, therefore, that automaticity occurs because neurological features that typically characterize NREM or REM sleep intrude into the state of wakefulness; i.e., it may turn out that automaticity can be characterized as a mildly sleeplike phenomenon. The crucial difference, as we shall see, lies in the fact that automatisms are typically impervious to interruption, while automaticity is typically susceptible to interruption.

[5] Note that this is not sufficient for the attribution of responsibility as there may be a lack of eligible alternatives (e.g., swerving on to the crowded footpath in order to avoid hitting the child on the pedestrian crossing is not an eligible alternative), or there may be insufficient time to react (i.e., due to time to impact and the comparative slowness of conscious control). On the need for eligible alternatives see T. M. Scanlon (1998, pp. 279–280, 291–292). In addition, I take it that when the agent is in a position to preset or revise her automatic behavior the attribution of responsibility requires that she is not subject to some form of cognitive impairment (e.g., immaturity may mean that the novice is insufficiently sensitive to reasons). That is to say, they must be sufficiently capable of understanding and assessing reasons when they are attending to the periphery. On the requirement that the agent be sufficiently responsive to reasons, see Fischer and Ravizza (1998, pp. 41–49).

References

Aristotle. (1998). *Nicomachean ethics* (W. D. Ross, trans.). Oxford, England: Oxford University Press.

Armstrong, D. M. (1997). What is consciousness? In N. Block, O. Flanagan & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 722–728). Cambridge, MA: MIT Press.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, England: Cambridge University Press.

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

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925–945.

Block, N. (1997). On a confusion about a function of consciousness. In N. Block, O. Flanagan & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 375–415). Cambridge, MA: MIT Press.

Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition*, *79*, 197–219.

Cane, P. (2002). *Responsibility in law and morality*. Oxford, England: Hart.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Co.

Dennett, D. C. (2003). *Freedom evolves*. London: Penguin.

Denno, D. W. (2003). A mind to blame: New views on involuntary acts. *Behavioral Sciences and the Law*, *21*, 601–618.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge, England: Cambridge University Press.

Ferguson, M. J., & Bargh, J. A. (2004). How social perception can automatically influence behavior. *Trends in Cognitive Sciences*, *8*, 33–39.

Frankfurt, H. (1988). Alternative possibilities and moral responsibility. *The importance of what we care about: Philosophical essays* (pp. 1–10). Cambridge, England: Cambridge University Press.

Groeger, J. A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task*. Hove: Psychology Press.

Haidt, J. (2001). The emotionalist dog wags its rationalist tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Levy, N., & Bayne, T. (2004). Doing without deliberation: Automatism, automaticity, and moral accountability. *International Review of Psychiatry*, *16*, 209–215.

Libet, B. (1985). Unconscious cerebral initiative and the role of the conscious will in voluntary action. *Behavioral and Brain Science*, *8*, 529–566.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 521–524.

Mahowald, M. W., & Schenck, C. H. (2000). Parasomnias: Sleepwalking and the law. *Sleep Medicine Reviews*, *4*, 312–329.

McSherry, B. (1988). Getting away with murder. *International Journal of Law and Psychiatry*, *21*, 163–176.

Moray, N. (1959). Attention and dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.

Nagel, T. (1979). Moral luck. *Mortal questions* (pp. 24–38). Cambridge, England: Cambridge University Press.

Palmeri, T. J. (2003). Automaticity. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 290–301). London: Nature Publishing Group.

Penfield, W. (1975). *The mystery of the mind: A critical study of consciousness and the human brain*. Princeton, NJ: Princeton University Press.

Ripstein, A. (1999). *Equality, responsibility and the law*. Cambridge, England: Cambridge University Press.

Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 729–753). Cambridge, MA: MIT Press.

Rosenthal, D. M. (2002). The timing of conscious states. *Consciousness and Cognition*, *11*, 215–220.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Schopp, R. F. (1991). *Automatism, insanity, and the psychology of criminal responsibility: A philosophical inquiry*. Cambridge, England: Cambridge University Press.

Searle, J. R. (1994). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Spinoza, B. (1955). *Correspondence*. In R. H. M. Elwes (Trans.), *On the improvement of the understanding; The ethics; Correspondence* (pp. 275–420). New York: Dover. (Original work published 1674).

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.

Wegner, D. M. (2002). *The illusion of the conscious will*. Cambridge, MA: MIT Press.

Wegner, D. M. (2005). Who is the controller of controlled processes? In R. R. Hassin, J. S. Uleman & J. A. Bargh (Eds.), *The new unconscious* (pp. 19–36). Oxford, England: Oxford University Press.

Yeo, S. (2002). Clarifying automatism. *International Journal of Law and Psychiatry*, *25*(2), 445–458.

Zhu, J. (2004). Locating volition. *Consciousness and Cognition*, *13*(2), 302–322.