Neuroethics (2010) 3:121–133
DOI 10.1007/s12152-009-9054-8

The BCN Challenge to Compatibilist Free Will and Personal Responsibility

Maureen Sie · Arno Wouters

Received: 2 February 2009 / Accepted: 23 November 2009 / Published online: 15 December 2009

Abstract Many philosophers ignore developments in the behavioral, cognitive, and neurosciences that purport to challenge our ideas of free will and responsibility. The reason for this is that the challenge is often framed as a denial of the idea that we are able to act differently than we do. However, most philosophers think that the ability to do otherwise is irrelevant to responsibility and free will. Rather it is our ability to act for reasons that is crucial. We argue that the scientific findings indicate that it is not so obvious that our views of free will and responsibility can be grounded in the ability to act for reasons without introducing metaphysical obscurities. This poses a challenge to philosophers. We draw the conclusion that philosophers are wrong not to address the recent scientific developments and that scientists are mistaken in formulating their challenge in terms of the freedom to do otherwise.

Keywords Compatibilism · Acting for reasons · Reasons-responsiveness · Personal responsibility · Free will · Determinism

M. Sie · A. Wouters (✉)
Department of Philosophy,
Erasmus University of Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: wouters@fwb.eur.nl
URL: http://www.xs4all.nl/~morepork/

M. Sie
e-mail: sie@fwb.eur.nl
URL: http://web.mac.com/mmsksie/

Introduction

The behavioral, cognitive, and neurosciences (hereafter: the BCN-sciences) are gradually beginning to reveal the mechanisms that make us who we are. The success of this enterprise has led some to worry that these sciences will undermine the notion of free will and the idea that people are responsible for what they do [e.g. 1–3].

The feared challenge of the BCN-sciences is often seen as a denial of the idea that persons are able to act differently than they in fact do. However, many philosophers have abandoned the idea that the ability to do otherwise is relevant to free will and responsibility long ago and they tend to dismiss the challenge as directed at an outdated view [e.g. 4, 5].[1]

According to a strong and influential current in philosophy, it is rather the ability to act for reasons that is crucial to our everyday practices of personal responsibility [e.g. 9–11]. We shall call this view 'new compatibilism'. One important reason to favor

[1] We suggest that the fears raised by the results of Libet's [6] experiments on the timing of consciousness in relation to brain activity and bodily movement and bold titles such as Wegner's *The Illusion of Conscious Will* [7] concern the thesis that consciousness does not influence our behavior rather than their alleged support for determinism. Such a lack of influence (if true) would threaten compatibilist and incompatibilist positions alike. Several compatibilists have recognized this threat and argued in response that the impotence of consciousness does not follow from the experimental results (e.g. [8]). We agree with that conclusion.

the new compatibilist account of responsibility over accounts in terms of the ability to do otherwise is that it seems so obvious that we act for reasons, whereas it is unclear and highly controversial what kind of ability the ability to do otherwise would be (especially if our behavior turns out to be determined by genes and environment). Because such determinism seems not to preclude us from acting for reasons, research that merely seems to strengthen determinism is perceived as irrelevant by these compatibilists.

Ironically, as we shall show, the ability to act for reasons is central to a great deal of interesting research in the BCN-sciences. This research indicates, as we shall argue, that it is not as obvious as it seems that the ability to act for reasons can serve as an unproblematic basis to justify our daily practices of responsibility. This does not imply, of course, that this research shows that we do not act for reasons, but it does pose a challenge to new compatibilist philosophers, as we shall explain.[2] This challenge deserves full philosophical attention.

This paper has two aims. We would like to invite those who think that the BCN-findings challenge our views of free will and personal responsibility to explicitly address the new compatibilist view, and we aim to convince the new compatibilist that the results from the BCN-sciences provide an interesting challenge to what we believe to be the core strength of their position.

In "The Classical Problem", "Two Crucial Turns", "New Compatibilism" we present a short en sketchy introduction to the view we call 'new compatibilism'. These sections are not meant as a review of the state of the art of the contemporary free will debate, but as an introduction to one influential and attractive family of positions in this debate, namely the positions we collect under the heading 'new compatibilism'. In "The Classical Problem" we introduce the idea of free will as the ability to do otherwise and the debate about the compatibility of this notion with the thesis of determinism. In "Two Crucial Turns" we describe two important contributions to the discussion that shifted attention away from the ability to do otherwise and determinism. Peter Strawson moved the focus of the debate to our everyday practices of holding ourselves and each other responsible for what we do. Harry Frankfurt argued that in those practices talk of 'free will' does not refer to alternative possibilities, but to something we did willingly. In "New Compatibilism" we describe the new compatibilist view and explain why it is so attractive. In "BCN Findings" we summarize some relevant BCN-findings. In "New Problems" we explain how these findings pose a challenge to new compatibilist philosophers. In "Conclusion" we draw some conclusions.

The Classical Problem

The problem of free will is one of the oldest and most frequently discussed in philosophy. It is traditionally framed as the problem of how to reconcile freedom and determinism. This difficulty arises out of the tension between our view of ourselves as persons who can be held responsible for what they do, and the scientific view that depicts our actions as the combined result of genes and environment. The traditional view is that humans are responsible for what they do to the extent that they have the freedom to do otherwise. Determinism poses a threat to personal responsibility because, if true, it is not clear how it could ever be possible that someone could ever do otherwise.

Classical compatibilists (beginning with Thomas Hobbes in the 17th century) argued that determinism does not exclude human freedom by interpreting the principle of alternative possibilities in a conditional way. According to them being able to do otherwise means that one would have done otherwise if one had willed or chosen to do otherwise.

Incompatibilists such as Thomas Reid (18th century) have objected that, while this interpretation might perhaps salvage human freedom, it does not restore personal responsibility. Suppose, for example that, unbeknown to me, some gum that I chewed contained nicotine, and that the taste of nicotine triggered my former addiction. Most people would agree that, in this situation, I am less responsible for taking the cigarette on the table in front of me than I would have been had I not chewed the gum. However, in either scenario I willed or chose to take the cigarette and would not have taken the cigarette if I had willed or chosen otherwise. The argument, of

[2] As the ability to act for reasons also figures in other accounts of personal responsibility, some or all of the challenges for new compatibilism identified by us, may apply to other views as well. We leave it to others to point that out.

course, is that my will has been seriously tampered with. So, according to traditional incompatibilists one cannot be blamed for doing what one chooses to do if the choice is not free. Hence, personal responsibility can be salvaged only if one could have chosen otherwise and this possibility does not exists in a deterministic world. Incompatibilists come into two kinds: libertarians who hold that incompatibilist free will exists and that determinism is untrue and hard determinist who hold the opposite view.

In some senses, the BCN-sciences aggravate the problem of determinism. As long as we did not know the workings of the brain, we could believe that incompatibilist free will enters somewhere in the brain. However, as the brain sciences advance and we learn more and more about our brains, the lack of any scientific basis for incompatibilist free will strengthens the idea that this kind of free will does not exist.[3]

Given the emphasis on the determinism issue in the discussion about the BCN-sciences, it is easy for compatibilist philosophers (which has been the dominant view in philosophy for many centuries) to maintain that those sciences do not pose a threat to free will. If free will and determinism are compatible, the alleged support of determinism by the results of the BCN-sciences cannot pose a threat to free will. Moreover, in the eyes of many compatibilists, libertarianism had already been dismissed on philosophical grounds, so why bother about a threat to this outdated view?[4]

As we argue below, this response ignores the specific character of the new compatibilist answer to the incompatibilist challenge. The new compatibilist approach is based on the idea that personal responsibility is grounded not in our assumed ability to choose otherwise, but in our ability to decide and act on the basis of reasons. In the example of the gum chewing smoker the new compatibilist would point out that the unknown presence of nicotine in the gum prevented the smoker from deciding on the basis of the relevant reasons. In this view it is the impossibility to take an influence (the presence of nicotine) into account that lessens responsibility, not the diminished ability not to take the cigarette (whatever that would be). Ironically, it is our ability to act for reasons in relation to all the factors that influence us (including those we are not aware of) rather than our ability to do otherwise on which the BCN-sciences focus. In order to explain new compatibilism and the challenges posed by the BCN-sciences we first need to discuss two landmark insights that explain how compatibilist came to see the ability to do otherwise as irrelevant to free will and personal responsibility.

Two Crucial Turns

Nowadays many compatibilists propose to sidestep the whole issue of the compatibility of the freedom to do otherwise with the thesis of determinism. They focus on a concept of personal responsibility that does not require such a freedom. A first landmark in the development of this idea is Peter Strawson's seminal "Freedom and Resentment" [14]. Strawson pointed out that our interpersonal relations are inextricably intertwined with spontaneous reactive attitudes such as gratitude and resentment, attitudes that constitute the practice of holding ourselves and each other responsible for what we do. In Strawson's view the abolition of this practice is impossible and undesirable because it constitutes the natural human way of relating to each other.

Strawson observes that our natural tendency to react to wrongdoing with sentiments such as guilt, resentment, blame, and moral indignation is lessened only when we discover that the wrongdoer did not mean harm (excuses) or that she was not a full-blown adult human being at the moment of her action (exemptions). In the latter case, we switch to what Strawson calls the 'objective attitude' towards this person, no longer regarding her as a human equal but as someone to be treated or manipulated or otherwise prevented to act wrongly. On the basis of this observation, Strawson argued that the thesis of determinism would threaten our everyday practices of moral responsibility only when it could be understood as implying 'no one ever meant any harm' or that 'everyone should be regarded from the objective attitude'. This, according to him, is an absurd suggestion. Hence, the thesis of determinism is irrelevant to our views of moral responsibility.[5]

[3] See Kane [12] for an honest attempt to locate free will in the brain.

[4] For an excellent discussion of the literature on this subject, see Roskies [5]. We wrote a short reaction in the spirit of the argument of this paper [13].

[5] We discuss this argument elaborately elsewhere [15].

Strawson's arguments have influenced many contemporary positions[6] and also inspired many objections.[7] Relevant to our purposes is that his essay changed the way in which philosophers discuss personal responsibility. In the wake of Strawson's essay our every day practices of responsibility, the distinctions we make in those practices, the conditions in which we excuse and exempt wrongdoers, and the reasons we give for our moves in those practices became the central focus of the debate. Contemporary philosophical accounts of responsibility aim to capture, understand, and evaluate the conditions for responsibility and the excusing and exempting conditions that we accept in everyday life.

Nowadays many philosophers are convinced that alternative possibilities are not among those everyday conditions. This idea was brought home by Harry Frankfurt [20]. Frankfurt argued that when we talk about 'free will' in our everyday practices, about 'doing something of our own free will', we do not refer to indeterminist intuitions of any kind. Rather, we are simply referring to something that we did willingly. For example, if I witness a crime and do not warn the police out of disinterest, the fact that I could not have warned them because—unknown to me—my phone has been disconnected, seems irrelevant to my responsibility for not warning the police. How exactly to define what it means to 'do something willingly' is, even today, subject to controversy. For our purposes, it is sufficient to understand it as in some relevant sense 'authorized' by the self/person; as something we 'agree with,' 'accept,' or 'positively endorse' ('decisive identification' and 'wholeheartedness' are the phrases Frankfurt [21, 22] uses).

Central to Frankfurt's hierarchical view is the idea that, as human beings we are not only directed at the world by desiring certain things (e.g. chocolate, a career, to stay dry when it rains), but are also directed at the desires that move us into action. Hence, we might prefer not to have the desire for chocolate, but we might be glad for and embrace our desire to work. It is within this framework that we should understand 'free will'. I act of my own free will when the desires that move me into action are the desires I authorize; I typically lack free will when the desires that move me into action are desires I resist, feel alienated from, and do not identify with.

Frankfurt's work opened up a new way of understanding free will without the need to meet the challenge posed by determinism and/or the conceptual problems attached to indeterminism. Frankfurt also offered an attractive explanation for the widespread and, in his eyes, mistaken idea that libertarian free will matters to our everyday practices of responsibility. He pointed out that we sometimes seem to excuse people for wrongdoing because they 'could not have done otherwise,' for example, when someone hurts us because she is pushed into us by another or by the abrupt movement of the train we are both on. However, Frankfurt points out that the reason that we accept this as an excuse is that we believe that the person would have done otherwise *if she could have done otherwise*. In other words, we believe that if she were not pushed, she would have avoided hurting us. According to Frankfurt, it is misleading to say that we excuse her because she lacks alternatives; we excuse her because she hurts us 'only because' she could not have done otherwise. It is not the lack of alternatives but the fact that we assume that, if she were not pushed, she would not have hurt us [20].

Frankfurt's criticism of the idea that responsibility is grounded in the ability to do otherwise is one of the defining landmarks in contemporary discussions on free will and responsibility. It needs to be addressed by everyone who defends or argues against a libertarian account of free will. However, it is not easy to formulate adequate conditions for personal responsibility based on Frankfurt's hierarchical view. Whereas the fact that someone wholeheartedly identifies with a certain action might be sufficient to hold her personally responsible for it even if alternate possibilities were absent (we will come back to this below), there is little reason to excuse someone for moral wrongdoings only because she did not wholeheartedly identify with them e.g. if someone steals my wallet, her reluctance to do so (because she knows that I need the money) does not in itself seem to lessen her personal responsibility.

[6] For example the work of new compatibilists such as Wallace [10] and Wolf [16]. Also see [17].

[7] See the above mentioned authors, but also Nagel's famous objection against the idea that we could prevent a slide from 'internal' to 'external' criticism of our everyday practices of moral responsibility [18] and Paul Russell excellent essay about the tension between several of the arguments Strawson endorses to prevent this so-called slide from internal to external criticism [19].

New Compatibilism

Many contemporary compatibilists looking for a justification of our common practices of responsibility have turned to a concept that we, for the purposes of this paper, will label 'reasons-responsiveness' [e.g. 9–11].[8] This is the approach that we call 'new compatibilism.' It is currently one of the most important and influential views in philosophy.

According to new compatibilism the ultimate justification of our practices of responsibility lies in our ability to act for reasons. Roughly, the idea is that we are responsible for what we do because we are the kind of beings that can figure out what to do and respond correspondingly. We are 'sane' human beings who are able to distinguish good from bad and are able to adjust our behavior in the light of it [16]. What capacities exactly make up our ability to act for reasons is a controversial matter. However, it is not controversial that in the case of human beings we can make a distinction between, on the one hand, actions done for reasons (stopping your car because others have right of way), and, on the other hand, bodily movements that are mere reactions to internal and external causes (tripping on the carpet). New compatibilism claims that we can understand and justify our everyday practices of responsibility on the basis of this distinction without further appeal to an assumed ability to do otherwise.

Unlike Frankfurt, new compatibilists do not believe that our personal responsibility can be grounded purely on a person's attitude towards the desires she acts upon. New compatibilists argue that in our practices of responsibility, in order to be held responsible, it does not suffice that an agent did something because she freely willed it; she should also be a sane and morally competent being (at the time of her action and with regard to it). That is, she should not only act on the basis of considerations that she accepts as reasons, but these considerations should also be real reasons. For example, if someone murders her neighbor because she understands this neighbor to, literally, be the devil; this person might be free in the sense that the murder was one she wholeheartedly embraced as the necessary and required thing to do at that moment, yet she cannot be held personally responsible for it because she does not qualify as a sane and morally competent being.[9]

Much more can be said about the notion of reasons-responsiveness and the conditions for responsibility based on this idea. However, for our purposes, it suffices to point out that, what is central and common to these views is the idea that some people clearly are reasons-responsive in that they act for reasons they can explain and justify, whereas others are not, either temporarily (e.g. when acting under conditions of extreme personal distress) or more permanently (e.g., when suffering from a mental illness that affects their moral competence).

This, we believe, is the strength and attraction of the new compatibilist view. Contrary to views that connect personal responsibility with the ability to do otherwise, new compatibilism does not need to assume the existence of a metaphysical obscure *counterfactual* freedom, i.e. 'that we could have done otherwise than we actually did.' They just point out that we are the kind of beings who regularly act for reasons and that it is our ability to do so that determines our status as responsible agents. It seems obvious that this is the case. We often manage to observe traffic regulations (e.g. to stop for the reason that another has right of way), play difficult games such as chess, and know exactly how and when to be polite, to give a few examples. Within that picture, we are excused if, and only if, our ability to act for reasons was imposed upon (e.g. because we were pushed or constrained) or undermined (e.g. because we were under hypnosis or suffering from a mental disease). In this way, new compatibilism can have its cake and eat it too. It enables us to do full justice to

[8] We use the term 'reasons-responsiveness' in line with Wolf [16] and Wallace [10] as an indication of the tenor of the new compatibilist approach, not as a specific condition for responsibility based on this approach. This use of the term 'reasons-responsiveness' differs from Fischer & Ravizza's [9] use of that term. The latter develop the general idea of reasons-responsiveness (in Wolf's and Wallace's sense) by specifying two conditions for responsibility: 'ownership' and 'reasons-responsiveness' (in a more specific sense).

[9] To be sure, these claims are vulnerable to all kinds of complicating objections. For example, even though many of us fail to grasp fundamentalist motivations and worldviews, we do hold those who act in accordance with them personally responsible. Therefore, who determines who is to count as sane and what actions are to count as 'morally deviant' instead of the result of mental illness? We have elaborated on these problems elsewhere [23].

the everyday moral distinctions that we make without getting into 'metaphysical trouble', or so it seems.

BCN Findings

Before we explain how the BCN-sciences challenge new compatibilist accounts of responsibility, let us first turn to the BCN literature that deals with our ability to act for reasons. According to this literature the reasons we give for our behaviors do not tell us about the inner deliberations that motivated our actions, but rather explain and justify these behaviors in a post-hoc manner, confabulating reasons if needed and rationalizing our behavior if that suits. In the next section we will explain how these findings challenge new compatibilism. Our point will not be that the BCN-sciences indicate that we are not reasons-responsive, but rather that in the light of the BCN-findings the concept of reasons-responsiveness turns out to be not as clear and free of metaphysics as it seems.

Confabulation was discovered by Michael S. Gazzaniga in the early seventies when he experimented with patients whose *corpus callosum* (the only direct nerve connection between the two hemispheres of the brain) has been severed [24]. Split-brain patients (as these patients are known) cannot describe pictures presented to only the left part of the visual field. This is because the information from the left visual field is routed to the right hemisphere and hence, is not available to the speech center in the left hemisphere if the brain is split. However, these patients are able to pick up with their left hand, a card related to a picture in the left visual field. In one experiment, Gazzaniga showed a boy two pictures at the same time, one in the left and one in the right visual field, each visible to a different hemisphere. The right hemisphere saw a snow scene, the left hemisphere a chicken claw. When asked to pick up cards related to what he saw, the boy picked up a picture of a shovel with his left hand and a picture of a chicken with his right. He explained this by saying that he picked the chicken because he saw a claw; and the shovel because you have to clean the chicken shed with a shovel. He had no idea that he was making up plausible reasons to explain his behavior, caused by factors of which he was not aware. Gazzaniga has argued that this tendency to confabulate reasons is not an aberration of split-brain patients, but a process that occurs often when people are not aware of the causes of their actions.

In their discussion of our ability to introspect about our higher cognitive processes, Nisbett and Wilson [25] go a step further than Gazzaniga. Not only do they suggest we easily confabulate reasons when we do not know the causes of our actions, but that we also lack introspective insight into those causes. Instead, causes are inferred based on 'a priori causal theories' originating from experience and the social environment. These theories tell us the possible causes of certain actions. When asked to explain an action, we determine which of the possible causes were present at the time of the action and cite that as the reason for the action. If we cannot find a plausible reason, we confabulate one. Causes that escaped our attention, causes that are not easily remembered, and causes that are not within our known range of possible causes will never be cited.

Wegner [7] takes this one step further, arguing that even the feeling of having acted is the result of the application of certain general principles of causal inference, rather than direct experience. We infer that we initiated a certain action in the same way as we infer that the movement of one billiard ball caused the movement of another. More specifically, people experience themselves as the originator of a certain event when they have a conscious thought corresponding to that event just prior to it and when they do not observe an alternative cause for that event. To support this theory, Wegner and his colleagues devised some very ingenious experiments. In the 'I Spy' study [26] a participant and a confederate together moved a computer mouse to guide a cursor on a screen with many small pictures. They were asked to stop cursor movement after a signal. Unknown to the participant, the confederate forced the cursor to stop at a certain picture. It turned out that, when the experimenter induced thoughts about the relevant picture either 1 or 5 s before the movement was halted, the participants reported feeling that they intentionally stopped the cursor at that picture. When the thoughts were induced 30 s before the movement was halted or one second after it, they said that they merely allowed the other to stop the cursor. In another experiment [27], the experimenters asked participants, allegedly as part of a study of psychological influences on health, to

perform a voodoo curse upon another person, in the presence of that person. The other person was a confederate who, at the end of the experiment, pretended to have a headache. Participants that were led to have negative thoughts about the confederate (for example, because he behaved badly) attributed the headache much more frequently to the curse than participants in whom no such thoughts were induced.

What this research shows is that we do not have access to the inner processes that connect the causes of our actions with the actions. When asked for reasons, we interpret our behavior with the help of a priori theories. Of course, it does not follow from this that we do not act for reasons, nor that conscious thoughts do not influence our actions. Nisbett and Wilson emphasized that a priori theories usually give quite reliable estimates about the real causes of our actions. Wegner admits that the application of principles of causal inference usually lead to correct identification of the originator of an action.[10]

Why should we bother about the mistakes people tend to make in ingeniously constructed, highly unusual, experimental conditions? It is not the fact that people make mistakes that challenges new compatibilism, but rather the new picture of the process of giving reasons that is supported by these mistakes. The mistakes indicate that the process of providing reasons is quite different from what it seems and only loosely connected to the processes that generate the actions. Initially one might think that when we give reasons we recollect the motives that drove our actions. The fact that people can make real errors in reporting reasons (errors that are neither the result from conscious or unconscious distortion of what they perceived, nor of an unwillingness to perceive their motives) shows that we have no direct access to our motives and, hence, that we do not recollect our motives when asked for reasons, but infer them on the basis of the information we do have. As in any experiment, the assumption is that the processes operating in the experimental conditions are the same as in the normal conditions. Compare this to the study of optical illusions. Just like the mistakes we discussed above, many optical illusions occur only in unusual and artificial conditions (when looking at drawings, movies or rooms that are especially designed to elicit the illusion), but this does not prevent them from providing insight in the manner in which we normally process visual input. In the same way the reasons brought up in experimental conditions and the mistakes people make in such situations provide insight in the way in which the process of giving reasons normally operate.

Another line of research however, does seem to lead to the conclusion that reasoning does not influence our moral judgments. This line of research originates in work emphasizing the role of intuitions and emotions in human decision-making [28] and in work that suggests that most of our everyday life is determined by automatic processes triggered by our social environment and operating without conscious awareness or guidance [29]. In an exciting and controversial article, the moral psychologist Jonathan Haidt [30] combined these insights and applied them to moral thinking to argue for what he called the 'Social Intuitionist Model of Moral Judgment'. According to this model, our moral judgments are not caused by moral reasoning, but rather are the result of intuitions: more or less instantaneous feelings of approval or disapproval that pop up in our minds and are generated by rapid unconscious processes. Moral reasoning is, in this view, just a post hoc attempt to justify these feelings and influence other people. Later research indicates that this model needs important qualifications,[11] but the important point for our purposes is the suggestion that, overall, moral deliberation might be post-hoc.

Finally, there is a line of research known as 'situationist psychology' that indicates that our moral judgments are influenced by many irrelevant factors (e.g. being in a hurry),[12] some of which might even escape our attention (e.g. in cases where we are 'primed'). For example, in the famous 'Good Samar-

[10] Wegner repeatedly suggests that thoughts and actions are the parallel results of a common unconscious process and that thoughts do not influence actions. This conclusion is consistent with his research, but it is not the only explanation for the results. One might also see the principle as giving reliable insight in what actually happens.

[11] In a subsequent paper Haidt himself took a more nuanced view on this issue [31].

[12] Whether a factor is irrelevant or not is of course, open to discussion. What we refer to in this case are the considerations that the agents themselves would not mention in accounting for their decision, action, or behavior or would not consider relevant.

itan' experiment by Darley and Batson [32], people's willingness to help a stranger in distress varied in accordance with the degree to which they believed themselves to be in a hurry. This, in itself, might not sound so shocking in an age where time is precious and helping strangers is not always without danger, however; the experiment was performed more than three decades ago with students in training for a helping profession (students of a theological seminary). The degree of time pressure was manipulated by telling them they were 'early,' 'right on time,' or 'late' for a video recording of a sermon about the biblical story of the Good Samaritan as part of their training [32, 33]. Surely, one would expect people training for a helping profession to be sensitive to the needs of someone in distress, regardless of being in a hurry. More recently, Wheatley and Haidt [34] hypnotically primed people to react with disgust to an arbitrary word and asked them to assess the severity of certain moral transgressions. They found that the participants were harsher in their judgment when the description contained the primed words. Schnall et al. [35] discovered that such things as the cleanliness of the table or the presence of certain odors could influence the severity of moral judgments.

To sum up these findings, BCN-research suggests that 'providing reasons' for our behavior is an interesting and complicated process that is better described as 'interpreting our behavior' than 'recalling what moved us'. This process is guided by a priori theories that inform us of plausible causes for our actions. Many factors that influence our behavior escape our attention and we are inclined to fill in the gaps in our knowledge with fabrications that are experienced as real. This process is tailored to justify our behavior and convince others, rather than to providing the truth about the motives of our actions. In the next section, we explain how this idea challenges the new compatibilist view of personal responsibility.

New Problems

As we discussed in "New Compatibilism" among the main attractions of new compatibilism are its foundation in the seemingly obvious fact that we act for reasons, as is evidenced by the efficacy of our daily practice of giving and asking for reasons, and the idea that our daily practices of responsibility (including the assumption that at least some people sometimes do things for which they are not responsible and our habit to hold people by default responsible for what they do) can be legitimized on the basis of this fact, without further appeal to obscure notions such as the ability to do otherwise. In this section we will argue that the BCN-findings discussed in "BCN Findings" cast doubts on this latter claim.

The BCN-evidence indicates that many actions for which the actor can give reasons are automatic responses to external stimuli, many of which are not recognized by the actor. These findings have led many scientists to the conclusion that, normally, people do not act for reasons. We do not think that this conclusion follows. Saying that a certain act was performed for a particular reason does not necessarily mean that that act resulted in one way or another from a conscious process of reasoning. There might be ways to accommodate the BCN-finding that normally people act automatically while maintaining that people normally act for reasons.

Yet, we do think that these findings spell trouble for the new compatibilist, for, if these findings are true, new compatibilists must find a way to view automatic actions as actions for a reason if they are to avoid the conclusion that acting for reasons is exceptional. Because new compatibilists are also committed to the thesis that the ability to act for reasons distinguishes actions for which the actor is responsible from action for which she is not, it follows that new compatibilists must come up with an account of what distinguishes automatic actions for a reason from automatic actions that were not for a reason. It is far from obvious that such a distinction can be made without introducing obscurities such as the ability to do otherwise.

This challenge is aggravated by the BCN-thesis that giving reasons is more a matter of interpretation than of recollection. As we explained in "New Compatibilism" the project of new compatibilism is to legitimate our common practices of responsibility independent of the issue whether someone ever could have acted otherwise. In practice, the issue of personal responsibility typically arises when someone behaves immorally without an adequate excuse or exemption. In such cases of wrongdoing we assume that there is a distinction between cases in which

wrongdoers are responsible for doing wrong, and those in which they are not. We shall call the first type of cases 'blameful wrongdoing' and the latter 'blameless wrongdoing'.

According to new compatibilism, the distinction between blameful and blameless wrongdoing in our daily practice is made and should be made based on the ability to act for reasons. However, someone who acts immorally without an adequate excuse or exemption in fact ignores the reasons or moral justifications for not acting in such a way. Clearly, someone who steals without an adequate excuse fails to respond to the reasons not to steal. That is, cases of wrongdoing by their very nature disclose a failure to respond adequately to reasons (see [23], Chapter 3). This means that the distinction between blameful and blameless wrongdoing cannot be viewed as a distinction between wrongdoings in which one did respond to reasons and wrongdoings in which one did not; it must be seen as a distinction between being able to respond to reasons and not being able to do so. It follows that the new compatibilist is committed to the view that in cases of blameful wrongdoing a person did not respond to the reasons there were, although she could have done so. This sounds astonishing like saying that she would have been able to do otherwise!

At this point, new compatibilists might want to point out that our every day practices provide evidence for the view that in some cases of wrongdoing wrong doers are capable of responding to reasons although in fact they did not do so. The kleptomaniac is deemed unaccountable because she communicates that she feels awful about her behavior and freely commits herself to treatment (showing she is not able to respond to reasons). The thieving student is deemed accountable either because her remorse shows her to be aware of having made a wrong choice (showing that she is able to respond to reasons); or, alternatively, because her anger at us, shows her unwillingness to accept our norms and values, which we firmly believe to be valid (showing that she responds to reasons, albeit to reasons we do not accept). So although we have a problem to explain how it is possible that people are sometimes able to respond to reasons to which they did not respond, we might be sure that this sometimes is the case.

However, if the BCN-sciences are right that reasons are inferred after the fact, it is far from clear that this practice provides the evidence needed by the new compatibilist. For the BCN-view suggests that differences between the self-reports of the thieving student and the kleptomaniac might result only from differences in their ability or willingness to account for their failure to respond to the reasons there are, rather than from differences in their capability to act for reasons.

The problem is not that people can make mistakes in their self-reports or lie about their motivations. The problem is that if giving reasons is a matter of post hoc interpretation, the alleged fact that we can make a distinction between blameful and blameless wrongdoers on the basis of the reasons they provide, does not provide evidence for the view that there ever are wrongdoers who are able to respond to reasons although they do not do so. It might only indicate that some people are better than others in rationalizing their failure to respond to the reasons there are. The thieving student might be just as incapable to act for reasons as the kleptomaniac although she thinks otherwise (and due to her rationalizing talents is able to convince others of it too).

Let us stress that we do not think that the BCN-sciences have shown that it is impossible for someone not to respond to reasons while being able to do so, or that the BCN-sciences show that our everyday ways of judging whether some wrongdoer can be held responsible for what she did are illegitimate. Our point is rather that in the light of the BCN-sciences it is unclear what it would mean to fail to respond to reasons while being capable of doing so and hence that it is unclear how on the new compatibilist view someone could ever be responsible for doing wrong. The new compatibilist needs an account of what it means to be able to respond to reasons while in fact failing to do so, an argument based on this account that shows that it is a real possibility that someone who did not respond to reasons was nevertheless able to respond to those reasons, and an argument that this account legitimates our common practices of deciding about the blamefulness of wrongdoings.

Fischer & Ravizza ([9] Chapter 3] provide an account that seems tailor made to solve this problem. Put very simply, the idea is that a person who in a certain case did not respond to certain reasons is nevertheless able to respond to those reasons if there is a scenario available in which she would have

responded differently.[13] Suppose someone stole a wallet. If she would not have stolen it when a police officer was standing next to her, this would show that she was nevertheless able to respond to the reasons not to steal. As Fischer and Ravizza put it: "we believe that, if an agent's mechanism reacts to some incentive ... this shows that the mechanism can react to any incentive" [9: 73].

At first sight this is an interesting and appealing idea. However, the argument is flawed. It does not help to distinguish between the thieving student and the kleptomaniac. For, a kleptomaniac, too, would be able not to steal under conditions such as having a police officer stand beside her. However, most of us do not need a police officer to prevent us from stealing. Therefore, the very fact that someone would steal means that she does not respond to the reasons to which most of us would respond i.e. internal reasons. Why would the fact that a person reacts to other reasons (the presence of a police officer) indicate she was nevertheless able to react to the reasons to which she did not respond? Fischer and Ravizza [9: 74] suggest that this inference is justified to the extent that the decision mechanism is the same in both situations. Yet, this response only displaces the problem, for it gives rise to the question of how to decide whether the same mechanism is in play. How can we be sure that this is the case and on what basis do we infer it?

The assumption that the ability to react to one reason provides evidence that one is able to react to all the reasons one recognizes is seriously challenged by situationist psychologists who have investigated the influence of features that we do not notice or believe to be relevant in explaining behavior [e.g. 33]. The general view of this paradigm is that our behaviors and actions are better explained in terms of the particularities of the situation that often go unnoticed or are deemed irrelevant than in terms of internal character traits, virtues, and morality. If it really is the case that the particularities of the situation strongly influence our ability to respond to reasons, our ability to respond to reasons in situations in which we respond to reasons, does not indicate that we were able to respond to reasons in situations in which we do not do so. So in the light of the BCN-sciences it is not clear how, on a new compatibilist account, it would ever be possible to do something wrong while being responsible for it.

The new compatibilist is not only committed to explain how it is ever possible that a wrongdoer was able to respond to reasons, although in fact she did not do so. The new compatibilist should also explain why in cases of wrongdoing it is reasonable to assume *by default* that the wrongdoer was able to respond to reasons. After all, in our daily practices we hold wrongdoers by default responsible for what they did. However, if situationist psychologists are right about the influence of the situation on what we do and decide, this practice must be reconsidered. If the particularities of the situation rather than differences in agential decision procedures determine whether or not someone acts morally it seems not reasonable to hold the agent responsible for acting wrong. As Doris observes, the challenge of situationism is that the sensible default attitude would be: "[...] a general agnosticism about responsibility attribution, since we could never confidently rule out the presence of competence defeaters" [33: 138].

This is a problem for new compatibilism, which argues that the assumption that we are reasons-responsive beings is metaphysically modest. Therefore, new compatibilism needs our reasons-responsiveness to be an unproblematic, easy to observe aspect of our

[13] Fischer & Ravizza's account is of course much more sophisticated. They distinguish two components of reason-responsiveness: receptivity to reasons and reactivity to reasons. 'Receptivity to reasons' refers to the ability of a person to recognize the relevant reasons, 'reactivity to reasons' refers to the ability to respond adequately to the reasons that are recognized. In their view a person is responsible for what she did if the mechanism leading to her action was 'moderately responsive to reasons', that is 'regularly receptive' to a variety of reasons, some of which are moral reasons, and 'weakly reactive to reasons'. 'Regularly responsive' means that the person recognizes an understandable pattern of reasons. Which reasons are moral reasons is to be determined by the community in which the person lives. 'Weakly reactive' means that the person appropriately reacts to at least one reason if recognized. Organisms that do not meet the receptivity requirement (Fischer & Ravizza mention animals, young children and psychopaths) fail to be moral beings at all and should not be treated as such. Both the thieving student and the kleptomaniac in our examples are regular receptive to reasons. The requirement of weak reactivity is meant to explain how it is possible that someone who on a certain occasion does not respond to certain reasons can nevertheless be thought to able to respond to those reasons in that very occasion (and, hence, be held responsible for not responding). We limit our discussion to the requirement of weak reactivity because this is the one that should solve the problem to which we draw attention.

everyday lives.[14] When the reasons we provide for acting immorally can be taken at face value, we can infer that at least sometimes some of us act immorally and are to blame for it e.g. the thieving student, but not the kleptomaniac. The BCN-sciences challenge this assumption.

So, the BCN-sciences pose at least three problems to the new compatibilist. First, the new compatibilist should find a way to accommodate the insight that most of our everyday life is determined by automatic processes triggered by external cues without introducing obscurities. Second, in the light of the BCN-thesis that giving reasons is a matter of post hoc interpretation, the new compatibilist cannot point to our every day practice to determine whether someone was responsible for doing wrong on the basis of the reasons they give to rebut the challenge that the new compatibilists' idea that some wrongdoers are able to respond to reasons to which they do not respond is as obscure as the libertarians' idea that those wrongdoers were able to do otherwise. Third, the new compatibilist should answer the challenge that everyone's ability to act for reasons is heavily compromised by the influence of the situation.

Before concluding, let us emphasize that in our view these challenges are problems that should be addressed rather than findings that undermine new compatibilism. We ourselves are not convinced of the new compatibilist view in all its aspects. However, we do believe that it correctly points to the importance of our abilities to respond to reasons as a key to understanding our moral practices. For that reason, we believe that our understanding of free will and responsibility will be improved if we take BCN-findings concerning these abilities into account. We suspect that, in doing so, some version of incompatibilist free will will reappear, but this does not alter our view that our common point of focus should be the efficacy of reasons, moral deliberation, and our general reasons-responsiveness.

[14] Elsewhere we argued that, if we restrict the new compatibilist position as applicable only to our everyday moral wrongdoings—cycling on the pavement, ignoring traffic regulations, and so on—, its justifying strength can be saved [36].

Conclusion

Our main point has been that the alleged threat of the BCN-sciences to free will and personal responsibility must be discussed in connection with our ability to act for reasons. New compatibilism, which we regard as the most influential compatibilist view of personal responsibility today, seeks the basis for our ascriptions of personal responsibility in our ability to act for reasons. The main attraction of this view is the fact that our ability to act for reasons seems so mundane and undeniable, quite unlike an assumed ability to do otherwise. Our daily practice of asking and giving reasons seems to show that we are acting for reasons all the time. However, as we discussed, developments in the BCN-sciences suggest that it is not as obvious as it seems that our ability to act for reasons can serve as an unproblematic basis for our views of free will and responsibility. The BCN-findings indicate that most of daily life consists of automatic responses to external stimuli. To accommodate this insight, the new compatibilists must find a way to distinguish automatic actions for a reason from automatic actions that were not for a reason. It is not obvious that this distinction can be made without an appeal to something like the freedom to do otherwise. Furthermore, developments in the BCN-sciences suggest that our self-reports and self-understanding are not necessarily evidence of the ability to act for reasons. This underscores a problem that arises independent of the BCN-findings: How to justify our everyday ascriptions of personal responsibility for wrongdoings (including us taking responsibility for our own wrongdoings). Wrongdoings typically disclose a failure to respond adequately to the reasons that exist. So the new compatibilist seems committed to the view that at least in certain cases wrongdoers were capable of responding to the reasons to which they, in fact, did not respond. This sounds as obscure as being able to do otherwise than one, in fact, did, but the new compatibilist might point out that in our everyday practices we routinely infer that some people are responsible for wrongdoings based on the reasons they provide. However, if the BCN-science are right that giving reasons is a matter of *post hoc* interpretation rather than of recalling motivations it might be that the differences between those who are deemed to be responsible for their wrongdoings and those who are not, have more to do with their ability to interpret

what they did than with their ability to act for reasons. This brings to the fore the situationist challenge that the particularities of the situation might adequately explain wrongdoings, hence, excuse the wrongdoing agent. If this is correct, then our everyday beliefs about personal responsibility need serious revision; if it is not, we need to determine how to respond to the overwhelming and fascinating evidence from the BCN-sciences that problematizes our reasons-responsive abilities.

References

1. Brooks, David. 2007. The morality line. *The New York Times*, 19 April 2007.
2. Horgan, John. 2002. More than good intentions: Holding fast to faith in free will. *New York Times*, 31 December 2002.
3. Wolfe, Tom. 1996. Sorry, but your soul just died. *Forbes Magazine* 158(13): 210 e.v.
4. Dennett, Daniel C. 2003. *Freedom evolves*. New York: Viking.
5. Roskies, Adina. 2006. Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences* 10(9): 419–423.
6. Libet, Benjamin, C.A. Gleason, E.W. Wright, and D.K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain* 106 (3): 623–642.
7. Wegner, Daniel M. 2002. *The illusion of conscious will*. Cambridge: MIT Press.
8. Mele, Alfred R. 2009. *Effective intentions: The power of conscious will*. Oxford: Oxford University Press.
9. Fischer, John Martin, and Mark Ravizza. 1999. *Responsibility and control: A theory of moral responsibility.* Cambridge: Cambridge University Press.
10. Wallace, R.Jay. 1994. *Responsibility and the moral sentiments*. Cambridge: Harvard University Press.
11. Wolf, Susan. 1981. The importance of free will. *Mind* 90 (359): 386–405.
12. Kane, Robert. 1996. *The significance of free will*. Oxford: Oxford University Press.
13. Sie, Maureen, and Arno G. Wouters. 2008. The real challenge to free will and responsibility. *Trends in Cognitive Sciences* 12(1): 3–4.
14. Strawson, Peter F. 1962. Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.
15. Sie, Maureen. 1998. Goodwill, determinism and justification. In *Human action, deliberation and causation*, ed. J. Bransen, and S. Cuypers, 113–129. Dordrecht: Kluwer.
16. Wolf, Susan. 1990. *Freedom within reason*. Oxford: Oxford University Press.
17. Watson, Garry. 1987. Responsibility and the limits of evil; variations on a Strawsonian theme. In *Responsibility, character, and the emotions*, ed. F. Schoeman, 256–286. Cambridge: Cambridge University Press.
18. Nagel, Thomas. 1979. Moral luck. In *Mortal questions*. Cambridge University Press.
19. Russell, Paul. 1992. Strawson's way of naturalizing responsibility. *Ethics* 102: 287–302.
20. Frankfurt, Harry G. 1969. Alternative possibilities and moral responsibility. *Journal of Philosophy* 66(23): 829–839.
21. Frankfurt, Harry G. 1976. Identification and externality. In *The identities of persons*, ed. Amélie O. Rorty. Berkeley: University of California Press.
22. Frankfurt, Harry G. 1987. Identification and wholeheartedness. In *Responsibility, character, and the emotions*, ed. F.D. Schoeman. New York: Cambridge University Press.
23. Sie, Maureen. 2005. *Justifying blame. Why free will matters and why it does not*. Amsterdam: Rodopi.
24. Gazzaniga, Michael S., and Joseph E. LeDoux. 1978. *The integrated mind*. New York: Plenum.
25. Nisbett, Richard E., and Timothy D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3): 231–259.
26. Wegner, Daniel M., and Thalia Wheatley. 1999. Apparent mental causation: Sources of experience of the will. *American Psychologist* 54(7): 480–492.
27. Pronin, Emily, Daniel M. Wegner, Kimberly McCarthy, and Sylvia Rodriguez. 2006. Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology* 91(2): 218–231.
28. Damasio, Anthonio R. 1994. *Descartes' error: Emotion, reason and the human brain*. New York: G.P. Putman's Sons.
29. Bargh, John A., and Tanya L. Chartrand. 1999. The unbearable automaticity of being. *American Psychologist* 54(7): 462–479.
30. Haidt, Jonathan. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4): 814–834.
31. Greene, Joshua, and Jonathan Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6(12): 517–523.
32. Darley, John M., and C.Daniel Batson. 1973. From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology* 27(1): 100–108.
33. Doris, John M. 2002. *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press.
34. Wheatley, Thalia, and Jonathan Haidt. 2005. Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16(10): 780–784.
35. Schnall, Simone, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34 (8): 1096–1109.

36. Sie, Maureen. 2006. Ordinary wrongdoing and responsibility worth wanting. *European Journal of Analytic Philosophy* 1(2): 67–82.

Maureen Sie (PhD practical philosophy, Utrecht University, 1999) is associate professor metaethics at the Erasmus University Rotterdam, The Netherlands. She seeks to develop a concept of moral agency that accommodates those recent developments in the cognitive and neuroscience that can be gathered together under the heading of the 'Adaptive Unconscious'. She was recently awarded a VIDI grant by the Netherlands Organisation for Scientific Research, which allowed her to start a research group on this theme.

Arno Wouters (PhD philosophy of science, Utrecht University, 1999) works as postdoc researcher in the VIDI project of Maureen Sie. His research focuses on the implications of recent developments in the neurosciences for anthropological notions like 'free will', 'responsibility', 'self' and 'person' for philosophical action theory.