Gregg D. Caruso

Précis of Neil Levy's Consciousness and Moral Responsibility

Neil Levy's *Consciousness and Moral Responsibility* (2014) provides the most comprehensive and clear-headed examination of the relationship between consciousness and moral responsibility in the literature to date. The goal of the book is to argue for and defend the *consciousness thesis*, which maintains that 'consciousness of some of the facts that give our actions their moral significance is a necessary condition for moral responsibility' (Levy, 2014, p. 1). Levy defends this thesis by bringing to bear a probing discussion of consciousness along with a rigorous survey of relevant work in the behavioural, cognitive, and neurosciences. He also provides a nuanced treatment of the implications of the consciousness thesis — rejecting *both* the arguments of those who deny the consciousness thesis, as well as those who accept it but conclude that we are therefore *never* morally responsible. According to Levy, the consciousness thesis entails that people are responsible less often than we might have thought, but it does not entail that we are never morally responsible (at least not for reasons having to do with consciousness — more on this in a moment).

In this précis I will provide a clear and concise summary of Levy's arguments in *Consciousness and Moral Responsibility*, without any additional commentary or critical comments. This will allow readers to evaluate the arguments for themselves. It will also allow the articles to follow in this book symposium — those by Chandra Sripada, Philip Robichaud, myself, and Neil Levy — to provide their comments and replies without having to summarize the arguments in full. Readers

Correspondence:
Email: gcaruso@corning-cc.edu

Journal of Consciousness Studies, **22**, No. 7–8, 2015, pp. ?–?

can then refer back to this précis for additional clarification or details on any particular argument.

1. Précis of Neil Levy's *Consciousness and Moral Responsibility*

It's important to begin by placing *Consciousness and Moral Responsibility* (hereafter CMR) into the context of Levy's previous work on free will and moral responsibility. In *Hard Luck* (2011), Levy argues that *luck* undermines free will and moral responsibility, and that no one is ever morally responsible for anything (for reasons independent of consciousness). He has not changed his mind about that thesis. Some may wonder, then, why he would bother arguing in CMR that consciousness is a necessary condition for moral responsibility. Levy, however, notes that 'exploring the commitments of our concepts is worth doing in its own right' (Levy, 2014, p. x). That is, even if one is a sceptic about free will and moral responsibility, it remains an interesting question whether consciousness is a necessary condition for moral responsibility.

Secondly, Levy believes that arguing for the consciousness thesis may be a more effective way to reduce the unjust practice of blaming people and holding them morally responsible in the *basic desert* sense — the sense that would make us *truly deserving* of praise and blame (see Levy, 2011; Pereboom, 2001; 2014). Convincing people that they are never morally responsible (in the basic desert sense) for reasons having to do with luck is a difficult task, but 'establishing that consciousness is a necessary condition for moral responsibility is an easier task and one that might meet with greater approval' (Levy, 2014, p. x). Hence, Levy believes the task is worthwhile since it 'will lead to fewer people being unjustly held morally responsible' (*ibid.*).

Lastly, Levy claims that 'the consciousness thesis matters for our moral lives even in the absence of a sufficient condition for moral responsibility' (*ibid.*, pp. x–xi). A big part of Levy's arguments (as we will see) is that only when agents satisfy the consciousness thesis do their actions and omissions express their personal-level attitudes, and only when they satisfy the consciousness thesis are they capable of exercising reasons-responsiveness. The consciousness thesis is therefore of practical importance according to Levy since: (a) it will aid our ability to morally assess people and their behaviour (since 'only when our actions are expressions of our selves can we be appropriately identified with them, such that we can be assessed on their basis'), and

(b) it will help us in deciding when agents ought to be sanctioned for pragmatic purposes (since 'we need to know whether their behavior was reasons-responsive' and 'we need to know whether their behavior expressed their attitudes, and which attitudes it expressed, because this knowledge will allow us to predict their future behavior') (*ibid.*, p. xi).[1]

Keeping in mind, then, that Levy is only seeking a necessary condition for moral responsibility, not a sufficient one, we can now examine his argument for the consciousness thesis. Chapter 1 begins by sketching the motivation of those who deny the consciousness thesis as well as reviewing the scientific evidence for the ubiquity and power of non-conscious processes.

Recent work in the behavioural, cognitive, and neurosciences has shown that the causes that move us are often less transparent to ourselves than we might assume — diverging in many cases from the conscious reasons we provide to explain and/or justify our actions. These findings reveal just how wide open our internal psychological processes are to the influence of external stimuli and events in our immediate environment, without knowledge or awareness of such influence. They also reveal the extent to which our decisions and behaviours are driven by implicit biases. No longer is it believed that only 'lower level' or 'dumb' processes can be carried out non-consciously. We now know that the higher mental processes that have traditionally served as quintessential examples of 'free will' — such as evaluation and judgment, reasoning and problem solving, and interpersonal behaviour — can and often do occur in the absence of conscious choice and guidance (Bargh and Ferguson, 2000, p. 926). Levy calls this the 'automaticity revolution' and it consists in 'recognizing the major role that automatic processes play in psychology, and therefore behavior' (Levy, 2014, p. 4).

While some conclude that these findings represent a serious threat to desert-based moral responsibility (since they indicate that we often lack consciousness of the facts that give our actions their moral significance), others go in the opposite direction, arguing that we need *not* be conscious of these facts to be responsible. In fact, there are a growing number of contemporary philosophers who reject the consciousness thesis. Prominent examples include Nomy Arpaly (2002),

[1] The sanctioning under consideration here would be for reasons of deterrence and incapacitation, not *just deserts*.

Angela Smith (2005), and George Sher (2009). These opponents of the consciousness thesis typically rely on everyday examples of agents who *appear* responsible but who act for reasons of which they are apparently unconscious, yet they also gesture at the fact that their view has the virtue of consistency with contemporary cognitive science since it gives the unconscious 'its due' (Levy, 2014, p. 3). As Levy describes it, 'Spurred by advances in cognitive science… and buttressed by philosophical argument, it is becoming increasingly fashionable to downplay the significance of consciousness' (*ibid.*, p. 2). The opponents of the consciousness thesis claim that 'their work is psychologically realistic and in tune with recent trends in cognitive science', and that it is the defenders of the consciousness thesis, like Levy, who are 'quaint and out of touch' (*ibid.*, p. 10). Levy's goal in CMR is therefore to defend the consciousness thesis against such opponents and to show that agents do, indeed, need to be conscious of the morally significant facts to which they respond in order to be morally responsible.

Levy begins his defence of the consciousness thesis in Chapter 2 by explaining what kind of consciousness, and with what content, is required for moral responsibility. He maintains that the work of Benjamin Libet and Daniel Wegner (Libet, 1999; Libet *et al.*, 1983; Wegner, 2002), which is often cited in this area, is simply irrelevant to moral responsibility since 'it makes no different whether or not consciousness has the powers they contend it lacks' (Levy, 2014, p. vii). Libet's pioneering investigations into the timing of conscious intentions and Wegner's work on the illusion of conscious will are often interpreted as showing that consciousness is *epiphenomenal* — lacking any causal role in action production. Levy, however, provides two reasons for considering these findings irrelevant. 'The first is simply that the twin challenges from neuroscience and social psychology adverted to are not very substantial' (*ibid.*, p. 14). That is, the work of Libet and Wegner does not 'amount to a substantive challenge to *any* interesting consciousness thesis' (*ibid.*, p. 15, italics added). According to Levy, important empirical and interpretive criticisms exist which challenge the purported conclusions of both (see, for example, Nahmias, 2002; Metzinger, 2004; Mele, 2009; Dennett, 1991; Rosenthal, 2002; Schurger, Sitt and Dehaene, 2012). The second reason is that the focus of the challenge (such as it is) is different than the focus of Levy's book: 'the consciousness thesis they have been taken to challenge is a different thesis to the one I have in mind' (Levy, 2014, p. 15).

Libet's challenge is specifically to the role of consciousness in initiating action. His experimental results are interpreted by many, including Libet himself, as showing that consciousness comes on the scene too late to play a role in initiating action. Wegner's work has 'a similar (apparent) moral' — it 'aims to demonstrate that our consciousness of action initiation is illusory' (*ibid.*, p. 16). For those that interpret these (apparent) results as undermining free will and moral responsibility (e.g. Spence, 1996; Wegner, 2002; Pockett, 2004), it is assumed consciousness must be involved in action control and initiation for an agent to be held morally responsible. Levy, however, disagrees. According to Levy, these theorists are 'wrong in claiming that it is a conceptual truth that free will (understood as the power to act such that we are morally responsible for our actions) requires the ability consciously to initiate action' (Levy, 2014, p. 16). For Levy, what is of true importance is the causal efficacy of deliberation — i.e. 'we want it to be the case that our conscious deliberation — our conscious consideration of reasons for and against a particular action — is causally efficacious' (*ibid.*, p. 24). *This* consciousness thesis is not threatened by the findings of Libet and Wegner (see, *ibid.*, pp. 18–25). Hence, for Levy, not only does the evidence adduced by Wegner, Libet, and their followers fail to establish the conclusions for which they argue, they also have the wrong consciousness thesis as their target.

Levy's consciousness thesis maintains that consciousness of the facts that give our actions their moral significance is a necessary condition for moral responsibility. The kind of consciousness Levy has in mind is *not* phenomenal consciousness but rather states with *informational* content. He limits himself to philosophically arguing for the claim that 'contents that might plausibly ground moral responsibility are *personally* available for report (under report-conducive conditions) and for driving further behavior, but also occurrent [in the sense of] shaping behavior or cognition' (*ibid.*, p. 31). What Levy means here by *personally available* is the following: 'Information is personally available… when the agent is able to effortlessly and easily retrieve it for use in reasoning *and* it is online' (*ibid.*, p. 33). In turn, information is available for easy and effortless recall if 'it would be recalled given a large range of ordinary cues; no special promoting (like asking a leading question) is required' (*ibid.*, p. 34). This notion of personal availability is closely akin to what Ned Block (1995) has called *access consciousness* — though the two are not exactly equivalent. Levy prefers not to build into his definition of *personal availability* the fact

that the information involved must also be available to a broad variety of consuming systems. Whereas Block builds such availability into his definition of access consciousness, Levy prefers to leave it as an open empirical question. But since Levy thinks the empirical question is answered in the affirmative — i.e. personally available information *is* information broadcast to a broad variety of consuming systems in the mind — it turns out that 'as a matter of empirical fact the two [notions] are coextensive' (Levy, 2014, p. 35).

We can now say that, on Levy's formulation of the consciousness thesis, information of the right kind must be personally available to ground moral responsibility. But what kind of information is the right kind? Here Levy writes, 'if the thesis is that agents must be conscious of *all* the mental states that shape their behavior, no one would ever be responsible for anything' (*ibid.*, p. 36). Rather than demanding consciousness of all relevant mental states, Levy argues that when agents are morally blameworthy or praiseworthy for acting in a certain manner they must be conscious of certain facts which play an especially important role in explaining the *valence* of responsibility. Valence, in turn, is defined in terms of moral significance: 'facts that make the action bad play this privileged role in explaining why the responsibility is valenced negatively, whereas facts that make the action good play this role in explaining why the responsibility is valenced positively' (*ibid.*, p. 36). Additionally, the morally significant facts that determine the valence need not track the actual state of affairs that pertain, but 'the facts that the agent *takes* to pertain' (*ibid.*, p. 36). The consciousness thesis can now be stated as follows:

> The consciousness thesis is the claim that an agent must be conscious of (what she takes to be) the facts concerning her action that play this important role in explaining its moral valence; these are facts that constitute its moral character. (*Ibid.*, p. 37)

According to the consciousness thesis, then, if an action is morally bad the agent must be conscious of (some of) the aspects that make it bad, and conscious of those aspects under an appropriate description, in order to be blameworthy for the action (*ibid.*, p. 37).

Chapters 3 and 4 are dedicated to spelling out the functional role of awareness and defending the *Global Workspace Theory* of consciousness (Baars, 1988; 1997). It is here that Levy advances an account of the role that consciousness plays in behaviour and why it is necessary for desert-based moral responsibility. According to Levy, since consciousness plays the role of integrating representations, behaviour

driven by non-conscious representations are inflexible and stereotyped, and only when a representation is conscious 'can it interact with the full range of the agent's personal-level propositional attitudes' (Levy, 2014, p. vii). This fact, Levy argues, entails that consciousness of key features of our actions is a necessary (though not sufficient) condition for moral responsibility since consciousness of the morally significant facts to which we respond is required for these facts to be assessed by and expressive of the agent him/herself.

Levy reiterates that the consciousness he is interested in is *not* phenomenal consciousness but (essentially) access consciousness. The so-called 'hard problem' (Chalmers, 1995) is not one he therefore needs to address. Philosophical zombies, for instance, may be phenomenally unconscious, 'but they lack nothing in the way of awareness; therefore their (alleged) conceivability raises no problem for us. Our question is about the functional role of awareness, and that is a much easier question' (Levy, 2014, pp. 38–9). In spelling out the functional difference that awareness makes, Levy turns to the Global Workspace Theory — which he claims is 'accepted by almost all neuroscientists working on the topic and a majority of philosophers as well' (*ibid.*, p. 39).

According to the Global Workspace Theory (hereafter GWST), awareness allows information to be made available to a broad variety of systems for further consumption, 'this allowing or even ensuring (depending on the details of the theory under consideration) that these systems play complementary roles in behavior' (*ibid.*, p. 39). On such an account, awareness plays an important *integrative* role (Morsella, 2005), and this role is key to understanding its function.

> It is unclear whether the information is actually bound in a single representation at any point, but the integration metaphor gets at something important: even if the representations are distributed across a broad range of brain systems, the distributed parts function together in a manner that produces integration of processing and behaviour. (Levy, 2014, p. 39)

The GWST was first proposed by Bernard Baar (1988; 1997) and subsequently developed into a neurobiological theory by Dehaene, Naccache, and colleagues (e.g. Dehaene and Naccache, 2001; Dehaene, Changeux and Naccache, 2011). The theory maintains that conscious information is broadcast to a broad range of consuming systems, which are distinct and dissociable components of the mind. Information that is not conscious, on the other hand, is less widely available (Baars, 1988; 1997). On Levy's account, then, the functional

role of awareness is to broadcast information to widely distributed brain regions which allows integration to take place and, in turn, rational domain-general information processing.[2]

After introducing the GWST and defending it against critics, Levy turns in Chapter 4 to explaining how the functional integration of information plays an important role in action. Here Levy argues that '[t]he integration of information that consciousness provides allows for the flexible, reasons-responsive, online adjustment of behavior'. Without such integration, 'behaviors are stimulus driven rather than intelligent responses to situations, and their repertoire of responsiveness to further information is extremely limited' (Levy, 2014, p. 39).

Consider, for example, cases of *global automatism* (Levy and Bayne, 2004). Such cases provide dramatic illustrations both of the stakes of the debate — 'by illustrating the seriousness of some of the cases in which justified attributions of moral responsibility turns (in part) on the truth of the consciousness thesis' — and also of the power of non-conscious processing to drive complex behaviour (Levy, 2014, p. 70). The most familiar example is somnambulism (but global automatisms may also arise as a consequence of frontal and temporal lobe seizures and epileptic fugues). Take, for instance, the case of Kenneth Parks, the Canadian citizen who on 24 May 1987 rose from the couch where he was watching TV, put on his shoes and jacket, walked to his car, and drove 23 kilometres to the home of his parents-in-law where he proceeded to strangle his father-in-law into unconsciousness and stab his mother-in-law to death. He was charged with first-degree murder but he pleaded not guilty, claiming he was sleepwalking and suffering from 'non-insane automatism'. He had a history of sleepwalking, as did many other members of his family, and the duration of the episode and Parks' fragmented memory were consistent with somnambulism. Additionally, two separate polysomnograms indicated abnormal sleep. At his trial, Parks was found not guilty and the Canadian Supreme Court upheld his acquittal.

While cases like this are rare, they are common enough for the defence of non-insane automatism to have become well established (Fenwick, 1990; Schopp, 1991; McSherry, 1998). Less dramatic,

---

[2] The remainder of Chapter 3 is dedicated to defending the GWST against criticisms, presenting the empirical evidence for the theory (see, for example, Dehaene and Naccache, 2001; Dehaene, Changeux and Naccache, 2011), and arguing for a 'broadcasting' version of the view over an 'access' version.

though no less intriguing, are cases involving agents performing other complex actions while, apparently, asleep (Levy, 2014, p. 72). Siddiqui, Osuna and Chokroverty (2009), for example, recently described a case of sleep emailing. As Levy describes it: 'After the ingestion of zolpidem for treatment of insomnia, the patient arose from her bed, walked to the next room and logged onto her email. She then sent three emails in the space of six minutes, inviting friends for dinner and drinks the next day. She had no recall of the episode afterwards' (Levy, 2014, p. 72).

According to Levy, '[t]hese cases illustrate the complexity of the behaviors in which agents may engage in the apparent absence of awareness' (*ibid.*). The capacities required for sleep emailing are rather complex: 'typing a relatively coherent message, entering a subject line, pressing "send" — all seem to require a high level of cognitive capacity' (*ibid.*, p. 73). This all raises the following question: if somnambulism (and other global automatisms) is a disorder of consciousness characterized by a dramatically reduced level of awareness of behaviour and surroundings, how can we explain the complex behaviours exhibited by sleep emailers or by Parks? It is here that Levy introduces the notion of *action scripts*:

> Skills the acquisition of which requires the engagement of brainscale distributed networks may be carried out efficiently and in the absence of consciousness, by networks of brain areas that are more localized. The skills that sleep emailing or sleep walking agents exercise are, in the jargon of psychology, overlearned… As a consequence they may be carried out efficiently in the absence of consciousness. Agents who experience disorders of consciousness follow what we might call *action scripts*, which guide their actions, I suggest, where a script is a set of motor representations, typically a chain of such representations, that can be triggered by an appropriate stimulus, and which once triggered runs ballistically to completion. (*Ibid.*, pp. 74–5)

An example would be learning to change gears in a manual car: 'we learn an extended series of movements, each of which serves as the trigger for the next' (*ibid.*, p. 75). In acquiring these scripts, we acquire capacities that may allow us to act efficiently in the absence of consciousness.

Levy, however, also points out that behaviours driven by action scripts tend to be inflexible and insensitive to vital environmental information. The behaviours of somnambulists, for instance, exhibit some degree of responsiveness to the external environment — e.g. navigating the layout of streets (*ibid.*, p. 75) — but they also lack

'genuine flexibility' of response. To have genuine flexibility of response, or 'sensitivity to the content of a broad range of cues at most or all times' (*ibid.*, p. 76), consciousness is required.

> …I have suggested that though nonconscious processes may drive complex behaviors, they do so in an inflexible manner. Behaviors for which there exist automatized action scripts may be triggered by environmental inputs, the behavior may respond to further cues, but the degree of sensitivity to such cues, and the range of cues utilized, will be limited to what has become automated. We might think of these action scripts as encapsulated: they may utilize perceptual information, but only in certain ways and only information which falls within a certain range (the ways and ranges are themselves automatized, reflecting the learning history of the agent). Perceptual information outside this range cannot be utilized even in cases in which the agent, were she aware of it, would recognize it as relevant to whether or how the script should unfold. As a consequence, the agent's behavior is inflexible and stereotyped. (*Ibid.*, pp. 78–9)

Conscious processing is therefore needed for flexible modulation of action scripts according to Levy.

> Only when the agent is conscious do a very broad range of internally and externally generated cues modulate behavior. It is only under these conditions that behavior is sensitive to such a broad range of cues, because only under these conditions are these cues integrated into the representations in the GWS, and subsequently broadcast to the consuming systems that drive behavior. (*Ibid.*, p. 79)

On Levy's account, then, integration buys us flexibility — 'without it, behavior is driven only by the cues which trigger action scripts, and by those cues to which overlearned scripts incorporate sensitivity' (*ibid.*, p. 79).

How does this all relate to moral responsibility? In the final two chapters Levy argues that the truth of his account of the functional role of awareness 'entails that agents satisfy conditions that are widely and plausibly thought to be candidates for necessary conditions of moral responsibly only when they are conscious of facts that give to their actions their moral character' (*ibid.*, p. 87). More specifically, Levy argues that the two leading accounts for necessary conditions for moral responsibility — *real self* (or *evaluative accounts*) and *control-based* accounts — are committed to the truth of the consciousness thesis despite what proponents of these accounts maintain. Levy argues that (a) only actions performed consciously express our evaluative agency, and that expression of moral attitudes requires consciousness of that attitude; and (b) we possess responsibility-level

control only over actions that we perform consciously, and that control over their moral significance requires consciousness of that moral significance.

Chapter 5 discusses *real self* accounts of moral responsibility (or *evaluative accounts*, as Levy prefers to call them) and aims at establishing conclusion (a) above. Contemporary proponents of such accounts advocate for an updated version of what Susan Wolf (1990) influentially called the *real self* view, inasmuch as they ground an agent's moral responsibility for her actions 'in the fact… that they express who she is as an agent' (Smith, 2008, p. 368). Many contemporary real self theorists deny that expression of who we are as agents requires that we be conscious either of the attitudes we express in our actions or the moral significance of our actions (Levy, 2014, pp. 87–8), but Levy claims these theorists are wrong. In trying to demonstrate the importance of the consciousness thesis to real self accounts of moral responsibility, Levy examines cases of global automatism where there is an absence of creature consciousness, as well as cases where there is creature consciousness but there is an absence of *state consciousness* of a fact that gives an action its moral significance. In both cases, Levy maintains, agents fail to be morally responsible in the desert-based sense since they are not conscious of the facts that give their actions their moral significance.

Consider again the Kenneth Parks case. Assuming that Parks was in a state of global automatism on the night of 24 May 1987, he acted without consciousness of a range of facts, each of which gives to his actions moral significance — '[h]e is not conscious *that he is stabbing an innocent person*; he is not conscious *that she is begging him to stop*, and so on' (*ibid.*, p. 89). These facts, argues Levy, 'entail that his actions do not express his evaluative agency or indeed any morally condemnable attitude' (*ibid.*, p. 89). Because Parks is not conscious of the facts that give to his actions their moral significance, these facts are not globally broadcast; and because these facts are not globally broadcast, 'they do not interact with the broad range of the attitudes constitutive of his evaluative agency' (*ibid.*, p. 89). This means that they 'do not interact with his *personal-level* concerns, beliefs, commitments, or goals' (*ibid.*, p. 89). Because of this, Parks' behaviour is 'not plausibly regarded as an expression of his evaluative agency' — agency caused or constituted by his personal-level attitudes (*ibid.*, p. 90).

Now, it's perhaps easy to see why agents who lack creature consciousness, or are in a very degraded global state of consciousness, are

typically excused moral responsibility for their behaviours, but what about more common everyday examples where agents *are* creature conscious but are not conscious of a fact that gives an action its moral significance? Consider, for instance, an example drawn from the experimental literature on implicit bias. Uhlmann and Cohen (2005) asked subjects to rate the suitability of two candidates for police chief, one male and one female.[3] One candidate was presented as 'streetwise' but lacking in formal education, while the other one had the opposite profile. Ulhmann and Cohen varied the sex of the candidates across conditions, so that some subjects got a male streetwise candidate and a female well-educated candidate, while other subjects got the reverse. What they found was that in both conditions subjects considered the male candidate significantly better qualified than the female, with subjects shifting their justification for their choice. That is, they rated being 'streetwise' or being highly educated as a significantly more important qualification for the job when the male applicant possessed these qualifications than when the female possessed them. Obviously, 'a preference for a male police chief was driving subjects' views about which characteristics are needed for the job, and not the other way around' (Levy, 2014, p. 94).

Is this kind of implicit sexism reflective of an agent's *real self* such that he should be held morally responsible for behaviours stemming from it? Levy contends that, 'though we might want to say that the decision was a sexist one, its sexism was neither an expression of evaluative agency nor does the attitude that causes it have the right kind of content to serve as grounds on the basis of which the agent can be held (directly) morally responsible' (*ibid.*, p. 94). Let us suppose, for the moment, that the agent does not consciously endorse sexism in hiring decisions — i.e. had the agent been conscious that the choice had a sexist content he would have revised or abandoned it. Under this scenario, the agent was not conscious of the facts that give his choice its moral significance. Rather, Levy argues, 'they were conscious of a confabulated criterion, which was itself plausible (it is easy to think of plausible reasons why being streetwise is essential for being police chief; equally, it is easy to think of plausible reasons why being highly educated might be a more relevant qualification)' (*ibid.*, p. 95). Since it was this confabulated criterion that was 'globally broadcast, and which was therefore assessed in the light of the subjects' beliefs,

[3] The description below is drawn from Levy's summary on pages 93–4.

values, and other attitudes', the agent was unable to evaluate and assess the implicit sexism against his personal-level attitudes. It is for this reason that Levy concludes that the implicit bias is 'not plausibly taken to be an expression of [the agent's] evaluative agency, their deliberative and evaluative perspective on the world' (*ibid.*, p. 95).

The conclusion of Chapter 5, then, is that if moral responsibility requires expression of evaluative agency, then agents like those discussed above are excused moral responsibility.

Chapter 6 makes a similar argument regarding control-based accounts of moral responsibility. According to control theorists: 'an agent is (directly) morally responsible for those actions over which he or she exercises the capacity for (sufficient) control' (*ibid.*, p. 109). Of course, what constitutes 'control' is a matter of contention, but one influential account that Levy discusses is that of Fischer and Ravizza (1998). Fischer and Ravizza argue that responsibility requires not *regulative* control — actual access to alternative possibilities — but only *guidance* control. 'Roughly speaking, we exercise guidance control over our actions if we would recognize reasons, including moral reasons, as reasons to do otherwise, and we would actually do otherwise in response to some such reason in a counterfactual scenario' (Levy, 2014, p. 109). Levy maintains that guidance control (like other control-based accounts) requires satisfaction of the consciousness thesis.

Levy's argument in Chapter 6 follows the same structure as the previous chapter — he examines cases of global automatism and cases of implicit bias, and he argues that: (a) responsibility-level control requires creature consciousness, and (b) in addition to being a conscious agent, the agent must be conscious of the moral significance of their actions in order to exercise responsibility-level control over it (*ibid.*, p. 111). With regard to (a), Levy begins by examining whether creature consciousness is required for guidance control.[4] An agent exercises guidance control over an action if it is caused by a moderately reasons-responsive mechanism; and a mechanism is moderately reasons-responsive if it is appropriately receptive and reactive to reasons, including moral reasons (Fischer and Ravizza, 1998, pp. 85–

[4] Levy proceeds by way of a consideration of guidance control since it is 'an undemanding kind of control' — i.e. if agents who lack creature consciousness, or agents who retain creature consciousness but lack awareness of a fact that gives an action its moral significance, do not have the capacity for guidance control, we can generalize the findings to any plausible account of responsibility-level control (Levy, 2014, p. 111).

8). Can an agent in a state of global automatism, such as Kenneth Parks, exercise guidance control? Levy argues that they cannot.

Levy acknowledges that Parks' behaviour may be *weakly* responsive to reasons: 'there is some scenario in which the mechanisms that cause behavior would be receptive and reactive to a reason to do otherwise' (Levy, 2014, p. 112). But weak reasons-responsiveness is not sufficient for guidance control — Fischer and Ravizza (1998) require *moderate* reasons-responsiveness. 'A mechanism is moderately reasons-responsive when it is regularly receptive to reasons.' That is, the mechanism must be 'responsive to reasons, including moral reasons, in an understandable pattern' (Levy, 2014, p. 113). For Levy, this condition entails that 'agents like Parks do not exercise guidance control over their behavior, because the mechanism upon which they act (the action script) is not regularly receptive to reasons' (*ibid.*, p. 113). The kind of broad sensitivity Parks lacks requires that 'contents be filtered through the GWS, where they can be assessed for consistency and for conflict with the agent's personal level attitudes' (*ibid.*, p. 113). Hence, according to Levy, in the absence of consciousness we get, at most, weak reasons-responsiveness, and that is not sufficient for responsibility-level control.

What about cases of implicit bias where agents are creature conscious but they are not aware of a fact that gives an action its moral significance? Here too, Levy argues, agents fail to be directly morally responsible. Consider again the subjects in Uhlmann and Cohen's (2005) experiment.

> These subjects were, of course, conscious agents, but they were… not conscious of the implicit attitudes that biased their information processing, thereby producing their confabulated criteria for job suitability. This implicit attitude imparted to their decision its morally significant content: its sexism. But because the subjects were conscious neither of the attitude nor its effect on their decision, they could not detect conflicts between either their attitudes or their decision, on the one hand, and their personal-level attitudes, on the other. What was globally broadcast, and therefore assessed for consistency and conflict, was the confabulated criterion; the attitude that caused the confabulation was neither broadcast nor assessed. (Levy, 2014, p. 115)

Because these agents were conscious neither of the implicit attitude that caused the confabulation, nor of the moral significance of the

decision they made, Levy maintains that they could not exercise guidance control over either (*ibid.*, p. 115).[5]

The conclusion of Chapter 6, then, is that if moral responsibility requires guidance control, then agents like those discussed above are excused moral responsibility.

In a very short concluding section, Levy addresses the concerns of theorists like myself (Caruso, 2012) who worry that the ubiquity and power of non-conscious processes either rules out moral responsibility or severely limits the instances where agents are justifiably blameworthy and praiseworthy for their actions. Levy attempts to avoid the following dilemma: either abandon the consciousness thesis, or surrender to scepticism since the effects of non-conscious processes are *so* pervasive that nothing can satisfy this demanding condition. Levy takes my view as one of his main targets:[6]

> This objection has been most forcefully pressed by Gregg Caruso (2012). Caruso opts for the second horn of the dilemma: he believes that the pervasiveness of nonconscious processes rules out moral responsibility. So much happens outside the sphere of our awareness that we simply can't possess the kind of control we need in order to be morally responsible for our action, he argues. (Levy, 2014, p. 131)

In replying to this type of concern, Levy argues that 'even in the kinds of cases Caruso cites (and which he rightly takes to be unexceptional), agents often possess sufficient consciousness of the central features of their behavior to count as responsible' (*ibid.*).

Levy's reply draws on an important distinction between cases of global automatism and implicit bias, on the one hand, and cases drawn from the *situationist* literature on the other. In the former, agents either lack creature consciousness, or they are creature conscious but fail to be conscious *of* some fact or reason which nevertheless plays an

---

5 Recall that guidance control requires moderate reasons-responsiveness, and moderate reasons-responsiveness requires regular receptivity to reasons, including moral reasons. But Levy argues that '[i]nsofar as our behavior is shaped by facts of which we are unaware, we cannot respond to these facts, nor to the conflict or consistency between these facts and other reasons' (Levy, 2014, p. 115). On Levy's analysis, '[w]e exercise guidance control over those facts of which we are conscious, assessing them as reasons for us', but in this case 'the contents that came up for assessment were confabulated… and the contents that caused the confabulation could not be recognized as reasons' (*ibid.*, p. 116). Hence, Levy concludes that these agents 'failed to exercise control over the central fact concerning their decision: *that it had a sexist content*' (*ibid.*).

6 I provide a reply and some corrections to the way Levy represents my view later in this issue.

important role in shaping their behaviour. In situational cases, however, Levy argues that the actions of agents are driven by non-conscious situational factors (factors the influence of which agents are not conscious) *but* they still remain aware of the nature of the actions they perform.

Consider, for example, the famous Good Samaritan experiment (Darley and Baston, 1973) where Princeton seminarians were asked to give a talk — some on the parable of the Good Samaritan, others on various topics. Once their talk was prepared, students were told to go to a building across campus to present their sermon. Some of the seminarians were told to hurry because they were running late while the others were told to take their time. On their way to the talk they passed a man who appeared to be in need of assistance but was actually a confederate of the investigators. The actor sat slumped over in a doorway, moaning and coughing. The researchers found that only 10% of the students who were told to hurry stopped to help, compared to 63% of the students without a time constraint placed on them.

While Levy acknowledges that experiments like these 'demonstrate that whether an agent does the right thing or not — helping someone apparently in distress, or helping to pick up dropped objects (Macrae and Johnston, 1998), say — may be strongly influenced by non-conscious factors', he contends that 'this concession does not seem all that threatening to moral responsibility' (Levy, 2014, p. 132). According to Levy, since agents (apparently) remain aware of the nature of the actions they perform in such experiments, the moral significance of the act is consciously available to them, and globally broadcast. It is therefore assessed for consistency with the agent's overarching attitudes and values (*ibid.*). Levy goes on to conclude:

> This assessment is biased by manipulation of the experiment, but not so much that agents are induced to act in a manner that genuinely *conflict* with their values. These manipulations, I suggest, modulate behavior such that agents are more likely to act in ways consistent with an interpretation of their values biased by the manipulation; they perform acts that accord with their values given a selfless, or selfish, spin. But given that they remain aware of the nature of these acts, the manipulation does not cause them to act in ways that genuinely conflict with their values. (*Ibid.*, pp. 132–3)

It is by drawing a distinction, then, between cases where agents understand the nature of what they are doing (e.g. the Good Samaritan case) and cases where agents fail to be conscious of facts that give our actions their moral significance (e.g. cases of global automatism and

implicit bias) that allows Levy to defend the consciousness thesis without embracing full-scale scepticism (at least for reasoning having to do with consciousness).

II. Conclusion

Levy's *Consciousness and Moral Responsibility* makes a strong case for the consciousness thesis and is essential reading for anyone interested in free will, consciousness, and moral responsibility. Whether one agrees with him or not, the arguments presented in CMR are important ones and should be taken up in future discussions of the topic. As Levy cautions, '[w]e should not turn philosophy over to the scientists', but likewise, 'when we are concerned with the nature of the objects of science, we ignore that science at our peril' (Levy, 2014, p. 135). Levy's approach is a refreshing one and it offers us a promising roadmap forward, one that requires a reciprocal engagement of science and philosophy. As Levy says in the final line of his book: 'It is in exploring this interchange between science and philosophy… that I believe progress on these issues will be made' (*ibid.*, p. 135).

References

Arpaly, N. (2002) *Unprincipled Virtue: An Inquiry Into Moral Agency*, New York: Oxford University Press.

Baars, B. (1988) *A Cognitive Theory of Consciousness*, Cambridge: Cambridge University Press.

Baars, B. (1997) *In the Theater of Consciousness*, New York: Oxford University Press.

Bargh, J.A. & Ferguson, M.J. (2000) Beyond behaviorism: On the automaticity of higher mental processes, *Psychological Bulletin*, **126** (6), pp. 925–945.

Block, N. (1995) On a confusion about a function of consciousness, *Behavioral and Brain Sciences*, **18**, pp. 227–287.

Caruso, G.D. (2012) *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*, Lanham, MD: Lexington Books.

Darley, J. & Batson, C.D. (1973) From Jerusalem to Jericho: A study of situational and dispositional variables in the helping behavior, *Journal of Personality and Social Psychology*, **27**, pp. 100–108.

Dehaene, S. & Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, **79**, pp. 1–37.

Dehaene, S., Changeux, J.P. & Naccache, L. (2011) The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications, in Dehaene, S. & Christen, Y. (eds.) *Characterizing Consciousness: From Cognition to the Clinic?*, Berlin: Springer-Verlag.

Dennett, D.C. (1991) *Consciousness Explained*, London: Penguin Books.

Fenwick, P. (1990) Automatism, medicine and the law, *Psychological Medicine Monograph*, **17**, pp. 1–27.
Fischer, J.M. & Ravizza, M. (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
Levy, N. (2011) *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, New York: Oxford University Press.
Levy, N. (2014) *Consciousness and Moral Responsibility*, New York: Oxford University Press.
Levy, N. & Bayne, T. (2004) A will of one's own: Consciousness, control and character, *International Journal of Law and Psychiatry*, **27**, pp. 459–470.
Libet, B. (1999) Do we have free will?, *Journal of Consciousness Studies*, **6** (8–9), pp. 47–57. Reprinted in Kane, R. (ed.) *The Oxford Handbook of Free Will*, New York: Oxford University Press, 2002.
Libet, B., Gleason, C.A., Wright, E.W. & Pearl, D.K. (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**, pp. 623–642.
Macrae, C.N. & Johnston, L. (1998) Help, I need somebody: Automatic action and inaction, *Social Cognition*, **16**, pp. 400–417.
McSherry, B. (1998) Getting away with murder: Dissociative states and criminal responsibility, *International Journal of Law and Psychiatry*, **21**, pp. 163–176.
Mele, A. (2009) *Effective Intentions: The Power of Conscious Will*, New York: Oxford University Press.
Metzinger, T. (2004) Inferences are just folk psychology, *Behavioral and Brain Sciences* **27**, p. 670.
Morsella, E. (2005) The function of phenomenal states: Supramodular interaction theory, *Psychological Review*, **112**, pp. 1000–1021.
Nahmias, E. (2002) When consciousness matters: A critical review of Daniel Wegner's *The Illusion of Conscious Will*, *Philosophical Psychology*, **15**, pp. 527–541.
Pereboom, D. (2001) *Living Without Free Will*, New York: Cambridge University Press.
Pereboom, D. (2014) *Free Will, Agency, and Meaning in Life*, Oxford: Oxford University Press.
Pockett, S. (2004) Does consciousness cause behavior?, *Journal of Consciousness Studies*, **11** (2), pp. 23–40.
Rosenthal, D. (2002) The timing of conscious states, *Consciousness and Cognition*, **11** (2), pp. 215–220.
Schopp, R.F. (1991) *Automatism, Insanity, and the Psychology of Criminal Responsibility: A Philosophical Inquiry*, Cambridge: Cambridge University Press.
Schurger, A., Sitt, J.D. & Dehaene, S. (2012) An accumulator model for spontaneous neural activity prior to self-initiated movement, *Proceedings of the National Academy of Sciences*, **109**, pp. E2904–2913.
Sher, G. (2009) *Who Knew? Responsibility Without Awareness*, New York: Oxford University Press.
Siddiqui, F., Osuna, E. & Chokroverty, S. (2009) Writing emails as part of sleepwalking after increase in zolpidem, *Sleep Medicine*, **10**, pp. 262–264.
Smith, A. (2005) Responsibility for attitudes: Activity and passivity in mental life, *Ethics*, **115**, pp. 236–271.

Smith, A. (2008) Control, responsibility, and moral assessment, *Philosophical* Studies, **138**, pp. 367–392.
Spence, S.A. (1996) Free will in the light of neuropsychiatry, *Philosophy, Psychiatry and Psychology*, **3** (2), pp. 75–90.
Uhlmann, E.L. & Cohen, G.L. (2005) Constructed criteria: Redefining merit to justify discrimination, *Psychological Science*, **16**, pp. 474–480.
Wegner, D.M. (2002) *The Illusion of Conscious Will*, Cambridge, MA: Bradford Books/MIT Press.
Wolf, S. (1990) *Freedom and Reason*, Oxford: Oxford University Press.