1 Free will in light of neuroscience

Walter Glannon

ABSTRACT

In this chapter, I provide a historical overview of the main philosophical positions on free will. Neuroscience has motivated a shift in the debate on free will from an external to an internal focus, from features of the external world to features of the brain and how it enables the mental capacities necessary for free and responsible agency. Neither neuroscientific determinism nor mechanism poses a real threat to these capacities and whether or how they are exercised. Neurological and psychiatric disorders may impair the will by disabling the neurobiological structures and processes that mediate the ability to reason and make decisions. But nothing about a normally functioning brain challenges the conviction that we have this ability and can act freely and responsibly. I provide summaries of the other chapters in this volume and the authors' perspectives on how our knowledge of the brain influences how they analyze theoretical and practical aspects of free will and moral and criminal responsibility.

Introduction

For centuries, philosophers have attempted to explain how human agents can choose and act freely and have an impact on events in the physical universe. The most serious challenge to the conviction that we have free will has come from causal determinism. This is the thesis that a complete description of the state of the world at some time T, in conjunction with a complete formulation of the natural laws, entails every truth about the physical state of the world at later times. Causal determinism implies that any action one performs at a given time is the only action one could have performed at that time. Insofar as free will requires the ability to do otherwise and causal determinism rules out this ability, it seems to follow that we lack free will. And insofar as moral responsibility presupposes free will, it also seems to follow that we cannot be morally responsible for any of our actions if causal determinism is true.

There have been different philosophical responses to causal determinism. While some of the authors in this volume describe and discuss some of these responses, it will be helpful to summarize them here. Philosophers who believe that free will is incompatible with causal determinism can be divided into two groups. Hard incompatibilists, or as William James described them, "hard determinists" (1956, pp. 145ff.) argue that causal determinism is true and therefore we have no free will. Libertarian incompatibilists argue that we do have free will and that causal determinism is false. They can be divided further into event-causal libertarians and agent-causal libertarians. The first group maintains that the decisions and actions of human agents are physically caused, but in an undetermined way. The second group maintains that human agents can perform some actions that are not caused by any physical events. The agent is the sole cause or originator of the action. All incompatibilists claim that the control of actions identified with free will requires that we be their source and that alternative possibilities be open to us when we act (van Inwagen 1983; Kane 1996; Strawson 2010). This depends on whether certain features of the universe obtain or fail to obtain. To use a metaphor from Jorge Luis Borges' 1941 short story, "The Garden of Forking Paths," alternative possibilities are the forking paths extending from the present to the future and among which we choose when we act (Fischer 1994, p. 3). If causal determinism is true, however, then there are no forking paths but only one path to the future, a path we cannot create or choose. This is what the hard incompatibilist believes. Compatibilists argue that free will and responsibility are compatible with causal determinism because the relevant sort of control does not require these types of alternative possibilities (Dennett 1984, 2003; Frankfurt 1988; Fischer 1994; cf. Berofsky 2012) Instead, we can control our thought and behavior when we have the capacity to respond to reasons for acting or refraining from acting (Fischer and Ravizza 1998). It is not natural laws and events in the past that impair or undermine this capacity but instead different forms of coercion, compulsion, or constraint. The control associated with free will does not depend on metaphysical features of the universe but on mental capacities of agents. The alternative possibilities necessary for one to act or refrain from acting are not possible states of affairs waiting to be actualized by us. Instead, they are functions of different sets of desires, beliefs, reasons, and intentions leading to actions. Like the actual sequence of mental and physical events that results in action, any alternative sequence that might have resulted in a different action is not external but internal to the agent and a function of his or her own mental simulation of future possibilities in adapting to the environment.

The will is not a single faculty associated only with volition but a deliberative and executive process extending from beliefs and motivational states to actions. The will is free when one has the capacity to recognize and react appropriately to reasons in forming and executing intentions in actions that accord with one's considered desires and values. We have free will when we can control the sorts of mental states we have and whether and how they issue in our actions. A psychological conception of the will based on mental capacities rather than a metaphysical conception based on features of the universe is more in line with our normative practices and institutions, especially the criminal law's view of persons as agents with the capacity for rationality. Legal theorist Stephen Morse notes that "libertarian free will is not an element of any crime or of any affirmative defense. To establish prima facie guilt, the prosecution *never* needs to prove that the defendant had [libertarian] free will" (2013, p. 38). What is needed to establish guilt or excuse is whether the defendant was capable of responding to reasons. This may be described as a broadly compatibilist sense of free will. That is, even if causal determinism were true, persons would still have free will if they had the capacity for deliberation and rational reflection, to respond to reasons and translate them into appropriate actions. Morse further says that "compatibilism is the only metaphysical position that is consistent with both the criminal law's robust conception of responsibility and the contemporary scientific worldview" (2013, p. 41. See also Chapter 13 in this volume).

Neuroscientific determinism

Neuroscience has a critical role in the psychological conception of free will because of the brain's capacity to represent different courses and outcomes of action and to generate and sustain the mental capacities that enable decisions and actions. Neuroscientific findings in the last thirty years have motivated a gradual shift in the focus of free will from external factors associated with natural laws and events in the past to internal factors associated with the relation between our brains and minds (Mele 2014b). This shift has not defused all possible threats to free will. Neuroscience may show that the mental processes on the basis of which we explain our actions may be determined by or reducible to neurobiological processes. This could undermine conscious control of our thought and behavior. As neuroscientist and philosopher Adina Roskies puts it: "The underlying worry is that those things that once seemed to be forever beyond the reach of science might soon succumb to it: neuroscience will lead us to see the 'universe within' as just part and parcel of the law-bound machine that is the universe without" (2006, p.420). So,

moving the locus of agency inside the agent may just introduce another medium for determinism. There are differences between deterministic and mechanistic processes in the brain regarding their implications for agency, and I will discuss the implications of each for free will in turn. The main point here is that neuroscientific determinism challenges the psychological conception of free will because it implies that neural processes alone ensure that we perform particular actions at particular times.

There is insufficient evidence from neuroscience to prove a deterministic relation between neural processes and mental processes associated with behavior. Because of limitations in functional neuroimaging and other neurophysiological measures of brain activity, neuroscience at best can show correlations between brain activity and the mental activity involved in decision-making and action. Correlation is not causation, and accordingly there is no warrant for claiming that neural processes cause us to have particular mental states and to choose and act in particular ways. The blood-oxygenation-level-dependent (BOLD) signal in fMRI is an indication of blood flow and increased metabolic demand in the brain. A more active BOLD signal in certain regions of the brain indicates more activity in those regions. If a subject in an experiment using fMRI is asked to perform a cognitive task such as choosing between two options, then presumably there will be increased activity in the prefrontal cortex and anterior cingulate cortex, two structures associated with executive functions such as planning and conflict resolution. But there is no way of telling whether the activity is due to excitatory or inhibitory effects in these brain regions and whether these effects determine that the subject chooses one of these options over the other. Moreover, it cannot be known whether or to what extent the changes in blood flow and increased metabolic demand were caused by the subject's consciously forming and executing an intention to choose, or whether these mental acts were caused by the neural activity. We cannot conclude from the activity displayed on fMRI that it caused the subject to choose as he did. Choice cannot be reduced to activation in a particular area of the brain. The BOLD signal and images displayed on fMRI or other forms of functional imaging are at best approximations of brain activity. They are visualizations of statistical analyses based on a large number of images and are more accurately described as scientific constructs than "pictures" or "snapshots" of what is actually occurring in the brain. The signal-to-noise ratio in fMRI requires that images be averaged over many trials to yield statistically significant results. Functional imaging generates group data, and one cannot draw a direct inference from information about the brains of groups to information about the brains of individuals and how they affect their mental states (Poldrack

2011). Roskies points out that this "raises a major concern . . . about the degree to which functional data about individuals can be interpreted, and whether and how scientific generalizations about brain function can be rendered applicable to individual cases" (2013, p. 45). The epistemic gap between what we know about neuroimaging and what we know about individual brain function warrants caution in drawing inferences from our limited knowledge of the brain to ontological claims about events and processes occurring within it. This pertains especially to the relation between brain activity and human thought and behavior.

There is also insufficient evidence to prove that brain functions can be explained in terms of indeterministic or stochastic processes. Some event-causal libertarians (Kane 1996, pp. 129ff.) appeal to this idea to motivate their arguments for free will. Moreover, some neuroscientists have explained free will in terms of fluctuating background noise in the brain that generates an element of randomness in our decisions to act (Bengson et al. 2014). Conceptually, it is unclear whether quantum stochasticity or any sort of randomness could provide us with the control of our behavior required for free will, since randomness is at odds with the idea of control. Empirically, it is plausible to assume that quantum stochasticity operates at the micro-level of subcellular information transfer in neurons. It is also plausible to assume that determinism operates at the meso-level of neural ensembles and at the macro-level of neural networks. But it is not known whether presumed indeterminate micro-level processes influence or cancel out and become irrelevant at meso- and macro-levels of brain activity. There is no convincing empirical evidence for the claim that brain activity is either completely deterministic or completely indeterministic. This casts doubt on claims that deterministic neural processes rule out or that random neural processes support the idea of free will. It is an oversimplification to assume that external deterministic or indeterministic processes go straight through the brain to action in a direct input-output relation, as though the behavioral outputs were nothing more than a function of the inputs. Neuroscientists Harald Atmanspacher and Stefan Rotter assert: "The intricate relations between determinacy and stochasticity raise strong doubts concerning inferences from neurobiological descriptions to ontological statements about the extent of determinism in the brain" (2011, p. 98). They further state: "Our bottom line is that pretentious claims as to deterministic or indeterministic brain activity are unfounded, and so are the consequences drawn from them" (2011, p. 99. See also Roskies 2006, 2010; and Balaguer 2010).

There are conceptual and practical difficulties with neuroscientific determinism apart from the empirical difficulties just considered. If brain

events determined mental events in a way that ruled out free will and responsibility for all human agents, then there would be no normative difference between a person who commits premeditated murder and one who kills another because of an uncontrollable violent impulse. Nor would there be any difference between a drunk driver who kills a pedestrian and a driver who has an unforeseeable seizure resulting in the same outcome. Those who assassinate others in the name of religion could not be responsible for their actions because they "could not have done otherwise." Neuroscientific determinism is also at odds with the fact that some therapies can cause changes in the minds and brains of some criminal offenders and rehabilitate them. The different normative judgments we make of these individuals reflect the understanding that the control of thought and behavior associated with free will is not an all-or-nothing capacity but one that comes in degrees. There is a continuum of control corresponding to the extent to which one has the mental capacity to respond to reasons for or against actions. As some of the authors in this volume point out, if neuroscientific determinism were true, then these differences would be irrelevant to questions about control and responsibility. No one could control their behavior and be responsible for it. There would be no basis for distinguishing between and among full responsibility, mitigation, or excuse in the criminal law, differences based on the understanding that individuals possess and exercise cognitive, affective, and volitional control of their actions to varying degrees. But there is no need to revise our moral and legal concepts and practices in light of neuroscientific determinism because this thesis has yet to be – and may never be – proven.

Neuroscientific mechanism

Mechanism may pose a more serious threat to free will. This thesis says that mental phenomena can be explained entirely in terms of their component neurobiological parts and the organization of and interactions between these parts (Craver 2007). If reductive mechanism is true, then neural processes obviate the need for psychological explanations of our behavior (Nahmias 2006, 2010). This can be described as the "causal exclusion problem" (Bayne 2011, p. 39). Mental properties are excluded from having a causal role in agency because the causal efficacy of neural properties underlying them provides a complete explanation of agency. If free will requires that our mental states play a causal role in our decisions and actions, and if mechanism shows that these states, *qua* mental, play no such role, then it seems that we lack free will.

In the early 1980s, neuroscientist Benjamin Libet conducted a series of experiments in which he used electroencephalography (EEG) to detect and measure activity in motor, premotor, and prefrontal cortices and supplementary motor areas of the brain when subjects were asked to flex their fingers or wrists (1985). These were a further development of similar experiments conducted in the 1960s (Kornhuber and Deecke 1965). Libet's experiments demonstrated that neural activity in the form of readiness potentials in these motor regions preceded the subjects' conscious awareness of their intention to act by several hundred milliseconds. The results of Libet's experiments suggested the epiphenomenal view that conscious mental states are the effects of neural mechanisms but have no causal influence on these mechanisms, which provide a complete account of our actions (Gallagher 2006, p. 110). Libet did not explicitly say that his experiments demonstrated that free will is an illusion, but he did say that they would have "a profound impact on how we view the nature of free will" (2004, p. 201) The challenge posed by the results of these experiments is how we could have free will if unconscious events in the brain rather than conscious mental states initiate actions. Psychiatrist Sean Spence spells out the apparent upshot: "If this is the case, then what space is left for freedom?" (2009, p. 6) It is difficult to sustain the view that we can be free and responsible agents if we are "merely victims of neuronal circumstances" (Greene and Cohen 2004, p. 1785).

But the idea that our conscious mental states or events do not initiate our actions does not mean that they have no causal role in our performing them. The fact that an unconscious neural event initiates an action does not imply that it can account for all the events in the process that extends from the formation of an intention to act to its execution in action. Philosopher Alfred Mele distinguishes proximal intentions from distal intentions. While the first type refers to intentions occurring just before action, the second refers to intentions to act at later times. These intentions may be either conscious or unconscious. Regarding bodily movements, Mele explains that from the datum that some neural events "begin before a conscious proximal intention emerges, one cannot legitimately infer that any of the following play no role in producing the movement: the acquisition of the proximal intention, the agent's consciousness of the intention, or the physical correlates of either of these items" (2009, p. 71). Events occurring unconsciously in motor or other areas of the brain may be the basis of an urge or inclination to act; but these might not lead to a decision to act (Mele 2009; Bayne 2011). The formation and execution of a conscious intention may be necessary to complete this process. Many of our actions result not only from proximal

intentions to perform them but also from distal intentions involving long-range planning. Even if neural events initiate particular actions, conscious distal intentions may influence neural events and their causal role in these actions. Libet's experiments at most show that certain neural events are necessary, not sufficient, for the occurrence of decisions to act (Mele 2014b). This is one example of the limitations of Libet's experiments. The actions the subjects performed involved a very narrow time frame and did not reflect the fact that many of our actions are the result of a broader temporal process of planning and decision-making. This process may involve many hours, days, weeks, or even years. Moreover, finger- and wrist-flexing are relatively trivial motor tasks that have little bearing on the types of actions for which free will is worth wanting. They are not the types of actions that make us candidates for praise, blame, responsibility, or punishment (Roskies 2006, 2010; Bayne 2011). These actions and the deliberation that precedes them do not occur as isolated neurological or mental events but are "embedded in a pragmatically or socially contextualized situation" (Gallagher 2006, p. 120. See Gillett's discussion of Chapter 3 in this volume).

The conclusions drawn from Libet-type studies cannot be generalized to the broader class of actions that figure in our normative judgments. This is at least partly because these studies involved only basic cognitive and motor tasks and focused mainly on prefrontal, premotor, and motor cortices and the supplementary motor area mediating these tasks. While the basal ganglia are also mentioned by some of the neuroscientists conducting these studies, there is little discussion of the role of these nuclei in planning and motor tasks (Haggard 2008, 2011, p. 11). The capacity for reasoning and decision-making is mediated by a broader network of interacting neural circuits in cortical, limbic, and subcortical regions of the brain (See Lipsman and Lozano's discussion in Chapter 10 of this volume). This network consists of re-entrant loops projecting from the prefrontal and parietal cortices to the thalamus, basal ganglia, and cerebellum, which then project back to the cortex (Spence 2009, p. 154). The basal ganglia and cerebellum are not only critical for motor control but also for the cognitive and affective processes necessary to coordinate action plans. While the prefrontal cortex plays a critical role in the formation and execution of intentions, higher- and lower-level neural circuits projecting to and from each other are all engaged in a normal healthy brain when a person makes a decision. Still, it seems plausible to say that neural events alone underdetermine our decisions and actions, and that both neural and mental events and states are necessary to explain thought and behavior and assess whether or to what extent a person has free will and can be responsible for what he or she does or fails to do.

Many neuroscientists who have expressed skepticism about free will base their claims on dualism about neural and mental properties. Some assume that free will presupposes substance dualism, the thesis that the brain and mind are distinct material and immaterial substances. These neuroscientists question the causal efficacy of the mental because an immaterial mind or soul cannot have any causal power of its own. But the assumption that the mind is independent of the brain may beg the question in favor of the causal irrelevance of the mind and the idea that free will is an illusion. Seizing upon this idea, biologist Anthony Cashmore says that "if we no longer entertain the luxury of a belief in the 'magic of the soul,' then there is little else to offer in support of the concept of free will" (2010, p. 4499). One philosophical response to this challenge relies on a very different interpretation of substance dualism. For example, philosopher Richard Swinburne claims that it is metaphysically possible that we could exist without a brain, which presumably supports the idea that we are essentially souls or minds. He adopts an agent-causal view of free will purportedly immune to the constraints imposed on agency by any form of materialism (Swinburne, 2013). Given current knowledge about the brain–mind relation, these claims against all forms of materialism and for free will are questionable because substance dualism is not a tenable thesis. It may be metaphysically possible for the mind to exist independently of the brain, but not empirically possible.

A number of philosophers of mind accept some form of property dualism. This says that there is only one kind of substance, the physical kind, but two distinct kinds of properties, physical and mental (Chalmers 1996, 2010). Mental properties depend on physical properties of the body or brain. Some of these philosophers argue that property dualism implies epiphenomenalism. They concede the existence of conscious mental properties but claim that they have no causal effects in the brain (Jackson 1982). Consistent with this interpretation of property dualism, some neuroscientists seem to hold that a mechanistic explanation of behavior in terms of neural processes implies epiphenomenalism. Neuroscientist Patrick Haggard suggests this idea in his comment that "Although consciousness may be part of brain activity, consciousness cannot cause brain activity, nor can it cause actions" (2011, p. 18).

Yet studies in psychiatry and neurology show that conscious mental states can cause brain activity and can at least partly explain behavior. Cognitive behavioral therapy (CBT) can rewire circuits in cortical-limbic pathways of the brains of patients with major depressive disorder (Goldapple et al. 2004). Reframing one's conscious and unconscious beliefs about external stimuli can cause changes in brain regions

mediating thought and mood. Neurofeedback may also enable subjects to modulate brain activity and symptoms in a range of disorders through their visualization of and response to this activity as displayed on EEG or fMRI. Patients with intractable pain can attenuate their perception of pain by modulating perceived activity in brain regions mediating pain, such as the rostral anterior cingulate cortex (de Charms et al. 2005). Producing effects similar to those in CBT, patients with major depression can alter brain activity regulating mood through neurofeedback as well (Linden et al. 2012). In addition, at least one study has shown that patients with Parkinson's disease who expected to receive a dopamine agonist but instead received a placebo produced endogenous dopamine in the basal ganglia (de la Fuente-Fernandez et al. 2001). In a similar study, Parkinson's patients receiving a placebo modulated neural activity in the subthalamic nucleus, a component of the basal ganglia, which in turn relieved some of their motor symptoms (Benedetti et al. 2004; Benedetti et al. 2011). While unconscious conditioning mechanisms based on neural processing partly explain these phenomena, they do not provide a complete explanation of them. Conscious expectation in the type of top-down processing involved in placebo responses also plays a causal explanatory role in the effects in the brain. These examples illustrate that psychological properties can have physical effects in the brain and that these properties have an important role in behavior control. Insofar as the content of the relevant mental states is shaped by interaction between the patient and physician, or the research subject and investigator, the examples also show that a biopsychosocial model including factors internal and external to the brain may provide the most satisfactory explanation of behavior.

If neuroscientific skepticism about free will is driven by a mechanistic model of human agency, and if this model cannot provide a satisfactory explanation of agency, then such skepticism is not on solid ground. Like neuroscientific determinism, the failure of neuroscientific mechanism to rule out free will and responsibility leaves our normative practices and institutions such as the criminal law unscathed. This lends support to the psychological account of free will. Morse implicitly appeals to this account and its application to the law when he states: "The law's view of the person is a creature capable of practical reason, an agent who forms and acts on intentions that are the product of the person's desires and beliefs. The law does not treat persons generally as non-intentional creatures or mechanical forces of nature" (2007, p. 205). Our increasing knowledge of the neurobiological underpinning of mental disorders may lead to a more humane criminal justice system in which deterrence and rehabilitation are given more weight than retribution in the treatment of