Proceedings of the 1st International Conference on Philosophy: Yesterday, Today, and Tomorrow (PYTT 2013), Global Science and Technology Forum, pp. 91-101.

Remaking responsibility

Complexity and scattered causes in human agency

Joshua Fost & Angela Coventry
Philosophy Dept.
Portland State University
Portland, OR USA 97207-0751
jwfost@pdx.edu

***Abstract*—Contrary to intuitions that human beings are free to think and act with "buck-stopping" freedom, philosophers since Holbach and Hume have argued that universal causation makes free will nonsensical. Contemporary neuroscience has strengthened their case and begun to reveal subtle and counterintuitive mechanisms in the processes of conscious agency. Although some fear that determinism undermines moral responsibility, the opposite is true: free will, if it existed, would undermine coherent systems of justice. Moreover, deterministic views of human choice clarify the conditions in which we ought to protect people from themselves, for example when they cannot give informed consent to medical procedures. Some of the most unresolved questions in this domain are just now emerging; they include robot ethics and the responsibilities of groups. We propose a philosophical and scientific research program to apply complex systems science to these problems.**

Keywords-free will; determinism; group agency; neuroethics; complex systems

I. INTRODUCTION

In the 1770 work *The System of Nature* [1], the Baron d'Holbach presented a thoroughly naturalistic and deterministic picture of the universe in which humans are "purely physical" beings that are "connected to universal Nature" and "submitted to the necessary, to the immutable laws that she imposes on all the beings she contains." Nevertheless, in spite of the "shackles" that "bind" us he notes that humans like to pretend to be free agents, that we are somehow independent of the causes that move us, that we "determine" our own "will" and regulate our own condition.

A few centuries later, most people still cling to a dualistic view in which, through means unspecified, we somehow remain metaphysically autonomous agents. Charles Campbell notes that human beings "obstinately persist in believing that there is an indissoluble core of purely self-originated activity which even heredity and environment are powerless to affect" [2]. Galen Strawson [3] put the point this way (see also [4]):

> *Almost all human beings believe that they are free to choose what to do in such a way that they can be truly, genuinely responsible for their actions in the strongest possible sense; responsible period; responsible without any qualification; responsible sans phrase, responsible tout court, absolutely, radically, buck-stoppingly responsible; ultimately responsible, in a word – and so ultimately morally responsible when moral matters are at issue.*

Opposing these intuitions, however, recent research in neuroscience strongly suggests that we are not free agents in that way. But if that is so—if our brains are the purely biological machines they give every evidence of being, built by genes, sculpted by development and life history and compelling us to act in deterministic ways—then what are we doing when we assess a person's status as a rational agent, grade their competence to make important decisions, or hold them responsible for their actions? How does our understanding of the brain bear on how we treat people and deal with each other socially? In this article, we will review some of the evidence leading to the conclusion that humans are not free agents, more fully situate that conclusion in a historical context, and begin to address some of the questions this raises.

II. PHILOSOPHICAL PERSPECTIVES

The sort of freedom Strawson describes is called *libertarian free will*. On this view, agents have some kind of power of free will, specifically a kind of freedom pertaining specifically to the operation of the will itself that is required for moral responsibility so that either praise-and-reward or blame-and-punishment can be deserved.[1] One recent proponent of libertarianism is Robert Kane, who claims that an agent is "ultimately responsible" for free actions [6]. *Ultimate* responsibility requires the action not be causally determined. Kane calls these sorts of actions "self-forming actions." These self-forming actions are acts of will or "self-forming willings." The idea that we deserve praise-and-reward and blame-and-punishment for our actions follows

[1] For an overview of libertarian positions see [5].

directly from this: we are in some sense ultimately responsible for becoming who we are, not causally determined by circumstances beyond our control.

Recent evidence from experimental philosophy suggests that people think their choices are free in just that way [7]. Gregg Caruso [8] noted that "One of the strongest supports for the free choice thesis is the unmistakable intuition of virtually every human being that he is free to make the choices he does and that the deliberations leading to those choices are also free flowing."[2] This argument from intuition can be traced back at least as far as 1739 when David Hume in his *Treatise of Human Nature* remarked that "the false sensation or experience" of free will is "regarded as an argument for its real existence" (THN 2.3.2.2; SBN 408).[3] Indeed, in most people this intuition is so strong and so obvious that they would flatly reject any alternative as preposterous.

Of course, a lone intuition or feeling of freedom proves nothing: one might well be mistaken [15], [16]. At the very least, intuitions of freedom err if they take our control over our own thoughts to be absolute. Here is proof: do not think about zebras for the next sixty seconds. That you cannot help from doing so shows that external stimuli can override your preferences concerning the objects of your thought.

Consider Hume's classic take on the relation between free will and necessity. He claimed as a rule, an axiom of methodological naturalism—of science itself—that every effect in the universe has a cause, and that this must extend to human behavior. The common intuition is that choice causes behavior, but Hume argued that that is not enough. The choices themselves have causes.

People generally feel that such choices—"free" here meaning unforced, not coerced by other agents, circumstances or other brute external constraints—are deliberations based on desires, intentions, goals, prior experience, and so on. In short, free actions are those that are caused by the agent. But Hume realized that this is equivalent to saying that if a third party knew your desires, intentions, etc. and also knew that you were acting according to them, then your actions would be, to the same degree as the perfection of the third party's knowledge, predictable. So we "may imagine we feel a liberty within ourselves; but a spectator can commonly infer our actions from our motives and our character" (THN 2.3.2.2; SBN 408).

We learn from experience that there is a great regularity between the motives and actions of our fellows: the same motives produce the same actions, just as the same events follow from the same causes (THN 2.3.1.4; SBN 400-401/EHU 8.1.6; SBN 83). After observing a variety of conduct in different people and differing circumstances, we are able to form generalizations and make predictions about human behavior. While it may appear that sometimes people act in unpredictable ways, Hume thinks that even the most irregular actions are frequently accounted for by the person's character and situation. A person with a nice reputation lashes at you, but it turns out that they have not eaten all day resulting in a bad headache which causes them to act snappish (EHU 8.1.15; SBN 88). Thus an action which seems at first irregular can be subsumed under another regularity about human behavior. In cases when the spectator *cannot* infer our actions from our motive and character, he or she concludes that they *would* be able to make the inference if they were "perfectly acquainted with every circumstance of our situation and temper and the most secret springs and of our complexion and disposition" (THN 2.3.2.2; SBN 408-9).

The alternative to predictability is randomness, but if in order to be free one must make only random choices, those uncorrelated with our desires and beliefs, then freedom, it seems, is relegated to the basest form of chaos. The more arbitrarily one behaved, the more free one would be. Freedom would require irrationality, because the rational aim of *wanting something* and then undertaking to get it is a constraint—not only because it biases our behavior in predictable directions, but also because we do not seem to be free to choose what we find pleasing. As Schopenhauer [17] said, you may be free to do what you want, but you are not free to want what you want. Holbach's objector in the *System of Nature* tried to demonstrate freedom by arbitrarily moving his hand: "[I]f it be proposed to any one, to move or not to move his hand, an action in the number of those called indifferent, he evidently appears to be the master of choosing" [1]. Holbach's reply is simply that the objector's choice is not arbitrary at all, but determined by the motive to convince his opponent that he is free.

If, contrary to these claims of determinism, we were ultimately responsible in the libertarian sense, then we would stand, as it were, at the beginning of causality. We would be causeless causes (*causa sui*), origins unto ourselves and agents of *originative responsibility*. This kind of view is defended by Roderick Chisholm [18] who claims that "each of us, when we act, is a prime mover unmoved," recalling Aristotle's description of God as the 'Prime Mover Unmoved.' For Chisholm, the decisions of a particular Earth-bound primate are mysteriously exempt from the laws of causation that govern every other phenomenon in the natural universe. Surely this extraordinary claim is in demand of considerably more than intuition to justify it.

We now set philosophy aside for the moment to consider what light contemporary neuroscience has shed on this problem.

III. The Neurophysiology of Motor Intentions

Armed with the guarantee that causes precede their effects, Benjamin Libet [19] conducted a seminal experiment to try to uncover the neural events that preceded voluntary movements.

---

[2] See also [9] and [10].

[3] The following abbreviations will be used for Hume: THN refers to [11], [12] and EHU refers to [13], [14]. References cite the book, chapter, section, and paragraph to the most recent Oxford edition followed by page numbers from the Selby-Bigge/Nidditch editions, prefixed by 'SBN.'

What he found was that there was indeed a detectable signal in the brain which *preceded* the moment at which subjects said they had decided to move. The experiment was as follows.

Human subjects outfitted with an EEG cap were asked to watch a rapidly rotating clock hand and move a finger whenever they wished. Once they moved the clock stopped and the subjects indicated where the clock hand had been when they first became aware of their intention to move. This moment turned out to average about 200 milliseconds before motion onset. On this basis, a libertarian might surmise that the will to move somehow influenced the brain, and 200 milliseconds later, movement occurred. That possibility, however, is undercut by Libet's EEG recordings, which showed that a "readiness potential"—a discernible electrical signal in the cortex—had begun about 500 milliseconds before the movement, i.e. about 300 milliseconds before the subject reported a conscious intention to move. Libet concluded "that cerebral initiation of a spontaneous, freely voluntary act can begin unconsciously, that is, before there is any (at least recallable) subjective awareness that a 'decision' to act has already been initiated cerebrally" [19]. While the size of that 500 millisecond window depends on experimental particulars, the basic ordering of unconscious biological event, followed by conscious intention, has been independently replicated since Libet's original report (see for example [20], [21]. Future clarification notwithstanding, such studies clearly suggest that conscious intention is not the proximal cause of behavior.

The EEG signal does not permit very good spatial resolution regarding the part of the brain where a signal originates. Lau et al. [22] therefore extended Libet's results by using fMRI to image brains under free-choice movement tasks. In the time leading up to movement, neurons in the pre-supplementary motor area (pre-SMA), dorsal prefrontal cortex (DPFC), and intraparietal sulcus became active. The precise functional role of these areas has not been firmly established (see [23] for a review), but it appears that the pre-SMA may be more closely linked to actual execution, while the other two areas collaborate in generating forecasts of motor plans.

It is well known that brain injury can impair our ability to execute a wide variety of cognitive tasks. Discoveries of brain correlates of volitional movement, such as those described by Lau and others, raise the possibility that lesions to those areas could alter the phenomenology of volition and free will itself. To that end, Sirigu et al. [24] found that patients with parietal lobe lesions could accurately assess the time at which a movement had begun, but had impaired awareness of their own intention to move. Rather than the 200 millisecond "warning" found by Libet, subjects in Sirigu's study consciously perceived an intention to move only 50 milliseconds prior to movement. Presumably, the undamaged areas outside the parietal lobe, such as the pre-SMA and DPFC investigated by Lau, were still active in the moments before movement. This suggests that the parietal lobe may be an important locus of conscious intention. More on this below.

Desmurget et al. [25] found an even more intriguing anatomical-functional dissociation. Patients undergoing awake brain surgery had electrodes inserted into their inferior parietal lobes and frontal premotor cortex. In these patients, weak stimulation of the parietal lobe actually *induced* a conscious intention to move; stronger stimulation enhanced this feeling to the point that patients claimed that they actually *had* moved, when in fact they had not. Stimulation of the premotor cortex did the reverse: it induced movement without the patient's awareness.

In related studies of healthy subjects (and no surgical intervention), Ammon & Gandevia [26] and Brasil-Neto at al. [27] successfully altered "free choice" of which hand to move (amusingly, exactly the scenario imagined by Holbach) by interfering with the supplementary motor area via transcranial magnetic stimulation (TMS). The subjects' choices were strongly influenced by the experimental intervention, yet the subjects themselves disavowed any such influence and maintained that their decisions were freely made.

Roskies [23] reviews a number of primate studies in an experimental paradigm first described by Newsome et al. [28]. A monkey views a field of moving dots and moves a joystick in the direction the majority of the dots are moving. Correct answers are rewarded. Neurons in the lateral interparietal area (LIP) play a key role in this task. When active, they signal the subject's choice of an action plan (e.g. move the joystick to the right) and the expected hedonic value of that action (e.g. how much of a reward the subject anticipates as a consequence of the behavior). By stimulating LIP, experimenters can bias the response and cause subjects to (for example) select a rightward joystick movement when in fact the visual stimulus was not coherently to the right. Roskies notes that LIP is probably not unique in this regard and may be but one example of a whole class of modality-specific (in this case, visuomotor) decision-making circuits. There may be modality-independent circuits as well: neurons in dorsolateral prefrontal cortex (DLPFC) can also be used to predict behavior, and also have firing rates that vary with expected reward value.

A general though admittedly not universal pattern begins to emerge from these and related studies. With respect to bodily movement, neural ensembles in the dorsal prefrontal cortex (and perhaps parts of the parietal cortex) compete with each other to determine the action plan that will maximize expected hedonic return. During this phase, subjects are not aware of having formed an actual intention to take action and indeed are not committed to any particular behavioral course. Once this competition is complete, a single winning motor program, i.e. the one projected to result in maximal reward, is sent both to the supplementary motor area and to inferior parietal cortex. The motor area activation produces the movement but no awareness of intention. The parietal activation generates a prediction about what the body position (and perhaps also the spatial arrangement of the world around us) will be once the motor program is complete. *This prediction* of the way the world will be once the

action is complete is identical with the phenomenology of intention to move.

Without validating the use of intuition as support for an argument, we venture to say that this picture is in rather good registration with neurally-informed common sense. The parietal lobes provide a variety of spatial processing services. Inasmuch as an intention to move is a forward-looking model, an envisioning of a goal state, *and* inasmuch as we are talking about physical movement, it seems reasonable that a clear spatial picture of that goal state could be closer to intention than, say, the means by which we will achieve it. I *intend* to exit the room. I don't *intend* to lift my left knee, flex my quadriceps, etc. The intention is the goal state, and the goal state is an updated spatial arrangement of body- and world-state. Since such representations are the business of the parietal lobes, the model seems at least plausible.

All told, neuroscience provides both a powerful argument against the causal force of free will and a (albeit nascent and provisional) physical account of how decisions (at least decisions to move) are actually made.

IV. MELE'S CRITIQUE

The widespread influence of Libet's experiments makes them a target of many criticisms.[4] One prominent opponent is Alfred Mele, who defends a moderate libertarianism [33]. Mele argues that conceptual analysis of the terms typically used to describe volitional acts undermines the significance that Libet's experiments are supposed to have for free will [34]. In the context of more recent neuroscience, we think Libet's original report is probably only minimally persuasive, so it is unclear to us why Mele would focus on it if he wished to defend a causal role for conscious intention. Nevertheless, in this section we outline the core of Mele's critique and offer a rebuttal.

The centerpiece of Mele's critique is a conceptual distinction between an urge and an intention—language that Libet himself uses somewhat equivocally. An urge to move, Mele says, is rather like a preponderant motivation to act at some future time. An intention to move (actually, what Mele calls a "proximal intention") in contrast, is an "act now" command. To support this distinction, Mele discusses another of Libet's experiments, one in which subjects were told to plan to move at a prespecified time but then to "veto" the movement at the last possible moment. The results of this experiment showed that the readiness potential still appeared—this, even though the subjects knew that they were never going to move. Mele concludes that the readiness potential cannot then signify an intention to move. Rather, it is more like an urge. This makes room for conscious intention to remain the final arbiter of movement. In Mele's words, "[T]hat in certain settings… urges to do things arise unconsciously – urges on which the agent may or may not act about half a second after they arise – is no cause for worry about free will" [34]; see also [35]. (Note that in both of Libet's experiments the origin of the urge ought to be fully transparent to the subjects: they were given explicit instructions to feel it!)

In our view, Mele's analysis does shed light on the concepts of folk psychology as they pertain to free will. Even so, this conceptual analysis bears upon the massively more informed and rigorous scientific project only insofar as any of these folk psychological posits have causal force. Quite simply, the idea that free will causes behavior is a theory asserting that our commonsense feeling of intending to A is the proximal and sufficient cause of our behavior. If that feeling were shown to be perfectly associated with certain brain processes (and evidence in that direction is starting to accumulate—see above), then the feeling *qua* phenomenal experience would be redundant in the causal explanation. Behavior itself is a brute physical event; muscular contraction has a purely physical explanation in terms of the firing of motor neurons. Motor neuron activity is in turn explicable in terms of the firing of yet other neurons, and those yet others, until at last we appeal to the set of external physical circumstances in which the person finds himself. Superimposing onto this unbroken chain the psychological / phenomenal will to move creates a problem of causal overdetermination wherein a single effect has more than one sufficient cause [36]. Such an overdetermined scenario would be unprecedented in natural science. To avoid this problem, and for the free-will-as-cause theory to be vindicated, *the phenomenal state that is the will to move would have to cause brain processes*, and that is an empirical claim with dualistic entailments that most scientists and philosophers, including Mele, want to avoid.

Let us reframe this rebuttal with a non-psychological example to make the distinctions clearer. On a windy day, a beach umbrella is caught by a strong gust and tips over. What caused the umbrella to tip? Our folk meteorological impulse is that it was the gust, but we catch ourselves realizing that the broader pattern of windiness that day caused the gust. Mele might characterize the gust as the cause and the windiness as a general condition of preparedness in which the gust can happen. (Again, these distinctions are in line with his characterization of the *intention*, which he thinks is free, coming in the context of an *urge*, which is not free, and which he thinks is what Libet's readiness potential measured.)

We may wish to refine this conceptual analysis by saying that gusts are not discrete things, but rather have fuzzy spatiotemporal beginnings and ends, while the umbrella's tipping only happened because there was a particular instant at which a horizontal force moved the center of mass past a threshold. This conceptual refinement can continue indefinitely, but no matter how adroit our distinctions at the folk-meteorological level, causal authority there stands to be usurped by explanations at more fundamental levels. The fluid dynamicist will give an account more far-reaching and powerful than the folk meteorologist: volumes of gas at certain temperatures and pressures flow according to the Navier-Stokes equations, etc. At this level, "gusts" become so ill-defined that they cannot be sustained in a coherent ontology, much less given causal authority. The high-energy physicist, in

[4] See for example [29]–[32].

turn, will regard the talk of fluids and gasses as conversational conveniences and decent approximations for some purposes, but will argue that superstrings vibrating in particular ways offer an even better causal account.

There is no denying that talk in terms of high-level constructs like gusts is often good enough for everyday purposes, but the standards for a philosophical and scientific explanation of the causes of human behavior must meet a higher standard. Unless the scientific program of recent centuries fails in some unprecedented way (e.g. the conservation of energy is shown not to obtain in human brains), high-level accounts of human behavior have to be cashed out at some point in terms of physics. The attribution of causal power to mental events is a refusal to participate in that cashing out process, and therefore a refusal to participate in the most productive practices of modern empirical inquiry.

Our complaint with Mele's critique of Libet, in short, is that Mele seems to assert that the conceptual analysis attached to intuitionistic folk psychology not only sheds light on the neuroscientific data, but actually forecloses upon what those data imply for the causal origins of human behavior. While we agree that philosophical treatments might play important or even vital roles in the interpretation of neuroscientific experiments, Mele's approach is not a particularly helpful treatment, particularly because investigations of the neural basis of motor acts conducted since Libet, such as [24], suggest functional dissociations that probably would never be revealed by the conceptual analysis that Mele favors. These dissociations include phenomena like (a) movement without intention or awareness of having moved; (b) deciding to move, then believing that one has moved, without any movement having actually occurred; and (c) deciding to move, then moving, but not believing or understanding that the intention to move caused the movement. Phenomena like these show that folk psychologistic causal accounts maintain their appeal only as long we don't confront special cases, such as people with particular sorts of frontal and parietal lesions. The *neuroscientific* accounts of volitional motor acts—accounts which can perhaps explain some folk psychologistic posits, but need not include those posits as causal forces—have greater scope and coherence and are therefore the preferred explanations.

V. Psychological mechanisms: Why do we feel free?

As disconcerting as it may be to confront the possibility that we may not be as in charge of ourselves as we thought, there seems to be no getting around the fact that even if we are *not* in charge, we feel as if we are. Under normal conditions, we do not feel like passive observers of our own behavior (as we do with, for instance, the patellar reflex), coerced by external forces, or compelled to move on pain of internal discomfort. Rather, we feel that we are in the driver's seat of our minds and bodies. At each moment we can turn or accelerate in whichever direction we like. Having chosen, we feel in retrospect that we could just as easily have done otherwise.

Many thinkers past and present have acknowledged this feeling and concluded that it is illusory. Most recently Nichols argues against the intuition or sense that we are free [7]. He maintains that in "light of work in cognitive science, we are no longer justified in sustaining the presupposition that we know what influences our choices." Spinoza [37] claims that people mistakenly believe themselves to be free, "simply because they are conscious of their actions, and unconscious of the causes whereby those actions are determined."[5] D'Holbach [1] similarly held that the "multiplicity" and the "diversity of the causes which continually act upon man," often without our knowledge of them, "render it impossible, or at least extremely difficult" to accept "the true principles of his own peculiar actions, much less the actions of others." Hume thought that the problem was that people are generally ambivalent in their views on necessity and mistakenly suppose themselves exempt from the determinate connections between causes and effects in the natural world in part because we are hasty reasoners and also because the causal origins of our behavior are so diffuse (EHU 8.1.22; SBN 93)—this is what Clark [39] calls a "scattered cause" and what Hume more poetically described as "a vast variety of springs and principles" (EHU 8.1; SBN 87).

Hume left the particulars of this hypothesis to future natural sciences and anatomy (THN 1.1.2.1; SBN 7-8/2.1.1.2; SBN 275-6). Gratifyingly, a sketch of one such solution is available in the competing neural ensembles model described above. In that model, discrete populations of neurons in the frontal cortex simulate alternative behaviors. The ensembles then compete, one motor program wins, and is acted on. Haggard suggests that the brain, in the interests of computational efficiency, allows the representations of all losing motor programs to dissipate [40]. From an attentional perspective, it would be overwhelming to be aware of all those things that we might have done but didn't. While the neural ensembles representing unexecuted programs fade into quiescence and are therefore lost to conscious awareness, the winning program can be—and possibly *must* be—kept active so as to compare the predicted motion with the actual motion and thereby generate an error signal used in learning and online behavioral correction. The end result of this programmatic pruning is a reliable sequence of events: a *single* motor program followed by movement. We interpret such consistent temporal correspondences as causal chains and conclude that our decision to X was the cause. We fail to recall all of the non-conscious and unwilled forces that led to that particular motor program's victory in the frontal lobe competition.

In a related model, Wegner asks us to imagine a magical process by which we could always know when a particular tree branch was going to move, and in which direction [41]. Further assume that by the same magic, we would always happen to be thinking about the tree branch's motion just before the event. Observing the reliable sequence of events—our thought, followed soon after by the real motion of the branch—we could scarcely fail to conclude that our thinking of it was the cause of

[5] For an updated Spinozistic approach see [38].

the motion. And yet, we stipulated from the beginning that no such causal connection exists. Wegner proposes that *attending to one's model of oneself before a behavior gives rise to the sense of causal agency*. Of course, to unpack the metaphor, we need a source of "magic." This could spring from the brain's representation of time. We know from other examples that the brain plays tricks in this way. Given neuronal conduction velocity, for example, we should see our foot touch the ground before we feel it (the visual signal from the eye reaches the brain before the tactile signal from the foot), but we don't. The events seem simultaneous. This could be because our brain takes the sensory signal and, as it were, spoofs the timestamp.

Remarkably, even inferences of causal agency can be affected through simple interventions. Gibbons [42] asked subjects to "imagine you are rushing down a narrow hotel hallway and bump into a housekeeper who is backing out of a room." Many subjects blamed the housekeeper, but if they were asked the exact same question while facing a mirror or hearing their own voices on tape, they became more likely to say that they themselves were the cause of the collision.

The moments in which we are attending to ourselves as agents in the world are of special importance in evolutionary terms, because they are moments when the process of learning about our ability to influence our surroundings is in full force. Our attention is more engaged in these moments than it is when we execute a well-worn routine. This attentional asymmetry produces a textbook example of *confirmation bias*: the moments when we are most alert and therefore most likely to remember are exactly those moments in which we are most likely to experience the illusion of causing our own behavior. In hindsight, our own causal agency stands out as a key feature of conscious existence.

VI. Applications and Implications

A. Criminal responsibility

Surely most audiences outside the particular psychological and neurophysiological subfields discussed here are unaware of these findings, so it should not surprise us that they continue to see themselves as buck-stoppingly free. Suppose, however, that they were aware of these findings. What would be the result? One might expect lukewarm reactions, not least because weakening the causal role of intention seems to have disturbing implications for the law, criminal law in particular, and other areas of applied ethics. The worry is that if the true causes of behavior are not intentions that have their origins in an agent's consciousness, then we lack standing to hold the agent responsible for the behavior. As we will see, however, this concern is a little overanxious. Even in a deterministic universe, there are coherent and defensible bases for holding people responsible for their actions.[6]

To illustrate why determinism is necessary for coherent justice, consider for a start Hume's view (EHU 8.2.29; SBN 98). He claims that determinism is in fact required if we are to be genuinely morally responsible for our behavior. Reward and punishment, for example, provide motives for ethical and/or legal behavior that assume some necessity in human behavior, that people desire certain rewards and fear certain punishments. These motives are supposed to have an "influence" on the mind and "both produce the good and prevent the evil actions" (THN SBN 410/ EHU 8.2.28; SBN 97-8). Hume thinks that a libertarian free will that is uncaused, free from prior causal determination, actually undermines moral responsibility. On this view no person can become "an object of punishment" (EHU 8.2.29; SBN 98). This is because a condition of moral responsibility is that the action should be attributable to some cause in the motives, desires, character and disposition of the agent who performed the action (EHU 8.2.28; SBN 97-8). Without these causal conditions present we cannot attribute the action to the person and so will have to forgo holding persons responsible for their actions. This is a key point in understanding the legal implications of determinism, so we will use the next few paragraphs to extend Hume's argument against libertarianism and address the two dominant models of criminal justice*:* consequentialist and retributivist.

If libertarians were right, then a person's conscious choice would be the original cause of a bad act. Judicial intervention would proceed along one of two lines: *consequentialist* or *retributivist*. Consequentialist motivations include deterrence (dissuading would-be criminals), incapacitation (preventing future bad acts from a criminal already caught), rehabilitation (turning a criminal into a productive citizen), and restoration (restoring something to the victim(s) of a crime). A libertarian free will entails that consequentialist intervention would be pointless, because it requires that the person being punished react in predictable ways to the punishment. That would mean that the person would have to be a deterministic system, with thoughts and behaviors governed by intrinsic mechanisms involving beliefs and values—mechanisms that the punishment aims to reshape. But the libertarian denies this causal structure, so consequentialist intervention couldn't possibly succeed.

Retributivist punishment, too, is undercut by libertarianism. The libertarian believes that a criminal possesses originative responsibility for his actions. He cannot admit into that picture the notion that the criminal did what he did *because* he lived in oppressive poverty, was abused as a child, etc. The word *because* indicates that one is about to supply a set of causes that sufficiently explains an effect. Under originative responsibility, there are no antecedent causes, or at least not sufficient ones:

[6] The position we are defending has some broad similarities with Dennett's stance on the matter. He also claims that human actions are determined and allow room for moral responsibility [43]. But there are also key differences. Dennett seeks to eliminate what he sees as a threat to free will posed by Libet-style experiments and he provides a positive account of free will based on evolution: we are free in ways that matter for responsibility because of the abilities that we evolved [31]. It is beyond the scope of this paper to explore the intricacies of Dennett's position.

choice stands as the ultimate go/no-go arbiter of behavior. Now presumably *the libertarian himself*, like the criminal, possess originative responsibility, in the sense that external events are not sufficient to cause him to think in any particular way. What, then, is the origin for his wish that punishment be meted out to the criminal? It seems obvious that his desire to punish follows from the commission of the crime. But beware: that commission cannot be a sufficient cause for his wish. If it were, then his desire would be psychologically determined, and as a libertarian, he denies just that kind of determinism. The crime can push him so far, but ultimately, *his decision to punish or not punish would have no sufficient cause.* Its *ultimate* origin would have to be his own free-floating will. For his punishment to be justified, it must not be arbitrary. But if it is not arbitrary, it is determined.

Libertarian freedom, then, is incompatible with both consequentialist and retributivist systems of justice. The determinist view, in contrast, is quite compatible with consequentialism at least, as we now show.

The consequentialist motivations of deterrence, prevention, etc. are all practical concerns motivated by the desire to create a functioning society, and there is nothing in that motivation that precludes even a strongly mechanistic view of the individual. Deterrence, for example, requires nothing more or less than a careful study of behavioral conditioning: will promised punishment dissuade a would-be criminal from a bad act? How effective does law enforcement have to be before a would-be criminal decides that a bad act is not worth the risk? Rehabilitation is much the same. What are the optimal reinforcement schedules for the operant conditioning that would reshape the convict's habits? Neuroscience can help answer all these questions. Understanding the principles involved and being able to apply them to a desired effect is, to be sure, a complex problem, not least because the governments of free societies cannot control most of the influences on a given person. But these are practical, not conceptual problems. The model of a person as a bundle of deterministic mechanisms is well suited, as far as it goes, to a science and technology of criminal justice.

Retributivist motivations for punishment, on the other hand, are based on the idea of moral redress. In that sense, they look backward in time and dispense punishments proportional to the infraction. Greene & Cohen [44] sum up the retributive philosophy as one under which "we legitimately punish to give people what they deserve based on their past actions— in proportion to their 'internal wickedness', to use Kant's phrase—and not, primarily, to promote social welfare in the future." Despite its seemingly metaphysical basis, retributive justice is the dominant philosophy in many, if not most countries, including the United States, partially because rehabilitation has proven so difficult in practice. Unlike consequentialism, therefore, retributivism seems less easily reconciled with the deterministic picture of humans as a bundle of mechanisms. This irreconcilability, however, loses some significance when put into context with the way criminal prosecution actually operates. Morse summarizes:

> *The law does not treat people as non-intentional creatures or mechanical forces of nature. The law treats persons, including people with mental disorders, as intentional creatures, agents who form intentions based on their desires and beliefs. Mental health laws treat crazy people specially not because the behaviors of crazy people are mechanisms, but because people with mental disorder may lack sufficient rational capacity in the context at issue. In other words, they were or are not responsible for their legally relevant conduct.* [45]

The central question, in other words, is not whether a person is buck-stoppingly free, but whether they have a general rational capacity sufficient to determine the likely consequences of their actions. If a person has that capacity and knows that the forecasted consequences are censured, they have *mens rea*—a guilty mind—and can be held responsible. So although Morse explicitly disavows the mechanistic view of a person, the role he assigns to intentionality entails no commitment to originative responsibility. Indeed, the description of intentions as things based on desires and beliefs suggests a view in which intentions constitute plans that will plausibly bring about the realization of whatever goals the person may have.

If, on the other hand, a person lacks the capacity to accurately forecast the likely results of their actions, or to understand that those results are illegal, they become candidates for exclusion from legal responsibility. In the United States, many insanity defenses, including the widely used M'Naghten Rules, express logic of this form. The insanity defense reflects the notion that persons who cannot appreciate the consequences of their actions should not be punished for criminal acts. The same argument can be applied to cases in which the defendant is too young to have developed the requisite cognitive capacities. In the 2005 case *Roper v. Simmons*, it was deemed unconstitutional to impose capital punishment for crimes committed while under the age of eighteen [46]. In 2010, the Supreme Court ruled that juveniles could no longer be sentenced to life without parole, except in homicide cases. We are clearly in the early days of neuropsychologically informed determinations of whether someone can appreciate the consequences of their actions, but already some of this work has led courts to conclude that brain development during adolescence lessens the criminal culpability that we ought to assign to younger people.

Neuroscience also has been an increasingly important tool for criminal defense lawyers in cases involving adults. This sort of evidence may illuminate specific cases to prove that a brain injury or brain abnormality caused the criminal behavior.[7] Laurence Miller documents the number of felony murder cases where posttraumatic stress disorder (PTSD) has been used as a defense to plead not guilty by reason of insanity or to argue for diminished capacity [49]. A recent case in Italy has an

[7] According to [47], the number of cases when judges have mentioned evidence from neuroscience in their opinion increased from 112 in 2007 to more than 1,500 in 2011. See also [48].

experienced and well-respected doctor accused of making sexual advances towards young girls under his professional care. By all accounts, this uncharacteristic behavior emerged suddenly. The case has yet to go to trial, but neuroscientific evidence is expected to play a major part: MRI scans revealed a tumor that, as attested by a molecular geneticist and psychiatrist, raised intracranial pressure and "altered his behavior" [50].

Irrespective of the success of defense strategies like this, it's clear enough that neuroscientific understanding of deterministic behavioral flow is not only compatible with legal responsibility, but possibly necessary in understanding when responsibility is or is not attributable. Trying to ascertain by an oral interview whether a defendant can forecast the consequences of his actions is subject to, at the very least, the difficulty of knowing what someone understands and whether they are being truthful. Probing brain activity holds the promise of ameliorating those worries, and perhaps significantly so.

Finally, we must acknowledge the potential risks that come attached to investigative power of this sort. Consider, for instance, the possibility of predicting future acts. During sentencing and parole decisions for a person already convicted of a crime, one consideration is the likelihood of recidivism. Neuroscience may contribute in its potential ability to predict that likelihood. For example, in an ongoing study, Thomas Nadelhoffer et al. [51] are compiling MRI data on a group of prison inmates to make predictions about possible future violence. One might worry, of course, that neuroscience will one day be so advanced as to predict that someone will commit a crime before they have committed it and arrests made on that basis. On the other hand, failing to intervene in such a circumstance might be ethically even more problematic. These are clearly questions deserving of further work.

B. Competence in medical treatment

The implications of uncovering the causal origins of human thought and behavior extend beyond criminal responsibility. Consider the following case study: a 75 year-old woman is admitted to a hospital with a gangrenous infection and is told that her leg needs to be amputated. She declines, saying that she has "lived long enough and wants to die with her body intact" [52]. Should her physicians accede to her wish?

One intuitive assessment might be that she is depressed as a consequence of her generally frail condition, and that if the she decided to forgo the procedure, thereby allowing the infection to spread, this would be "the depression talking," rather than a woman in possession of her normal rational capacities. Our knowing a possible material cause for her feelings seems to lift some of her originative responsibility for them: these are not *her* feelings, but deterministic consequences of external events. Whether we ought to credit this intuition is a question we'll return to in a moment.

The principle of autonomy is a guiding principle in medical practice: patients should be able to make their own decisions. However, that principle must be counterbalanced by a need to protect those with cognitive impairments. It turns out that the woman in this particular case had been showing signs of dementia in recent years. Is that enough to overrule her decision? Not necessarily. Applebaum lists four criteria for determining competence: ability to communicate a choice (which itself ought to remain stable over time); demonstrated understanding of the relevant information; appreciation of the situation and its consequences; and capacity to reason about treatment options [52]. Assays of patient competence along these dimensions are available, but an "overwhelming majority" of physicians find these tools inadequate [53]. Although a life-and-death situation like this one calls for a high standard of competence, if the patient can pass muster in all four areas, her right to autonomy ought to prevail over the possible objections of her physicians. [8]

We suggested above that knowing a simple cause for a patient's feelings might make us reluctant to see them as her own, and therefore less likely to regard as free or autonomous the choices flowing from them. Nothing like this condition—call it the "no simple cause" criterion—appears in Applebaum's list, suggesting that physicians are not troubled by this intuition, and are prepared to regard patients' decisions as deterministic so long as the operational standards of competence are met. The broader case that we have been making in this paper is that *all* feelings and decisions have deterministic causes—it's just that most of the time, these causes are not perceived as simple because they are scattered over our personal histories and the microscopic details of our constitutions. If that's right, then all that is required for a patient to preserve medical autonomy is for the complex causal network that is their decision making process conform to the standards of rational information processing represented in the competence standards. Only in cases where that processing was significantly disrupted (e.g. dementia, psychosis) would a judgment of mental incompetence be made. This aspect of contemporary medical practice, in other words, would not be troubled by the disappearance of the libertarian free will concept.

VII. FUTURE DIRECTIONS

In closing, we propose five interrelated avenues for future research. These concern (1) rational agents *qua* computationally complex systems; (2) the relationships between agents, scattered causes, and groups; (3) folk psychological models of agency in experimental philosophy; (4) the ethics of non-biological agents; and (5) the application of determinism to itself. We discuss each in turn and end with some positive reflections on the deterministic / anti-libertarian view.

In our medical competence example above, we adumbrated an intuition about the reluctance to identify free will as the cause of choices that have overt, simple causes. Although that intuition doesn't seem to figure into actual medical practice, we suggest

[8] Raymont et al. [54] found that mental incapacity was highly underreported. In a sample of 302 inpatients, blinded evaluation of a competence assay led to an estimate of at least 121 cases of incompetence, but the actual clinical team only flagged 12 of these (less than 10%) as incompetent.

here that there may be grounds for crediting it, at least to some extent. The field of complexity science studies, among other things, the conditions under which systems can compute. A system composed of interconnected elements that remain utterly static, for example, no matter what inputs are provided, is clearly not capable of computing, no matter how those elements may be wired together. Contrarily, a system composed of elements with internal states so chaotic that they maintain no correlation with the past is also incapable of doing interesting calculations. Langton applied to such systems the language of statistical mechanics and phase transitions. The first system was to be described as "frozen," a solid [55]. Such systems are perfect for storing memories, since the internal states are so stable. The latter system was seen, metaphorically, as a gas, in which individual components are largely uncorrelated. In between are liquids: structured enough to display spatiotemporal regularities, but loose enough to permit changes to one element's state to propagate to other elements. That causal cascade is a flow of information.

The upshot of this metaphor is that "interesting" computational systems seem to exist at some kind of computational phase transition, balanced between the needs of stable memories one the one hand and information transmission on the other. Kauffman suggests that life itself exists at such a phase transition, and that evolution is tuned to maintain the balance [56]. Living things too computationally frozen—e.g. those with no mutation—cannot evolve, while those too volatile cannot maintain stable identities or store adaptive solutions they discover. It seems possible that our intuitions about the defining characteristics of consciousness obey some of these same rules: solid enough to be law-abiding, rational engines, but fluid enough to react to changing stimuli and defy perfect predictability. Minds that sway too far to either side are treated specially. It would be interesting to know whether the concepts and mathematical tools of complex information processing systems could add structure to our theories of competence and rational agency.

Our second avenue for future research extends this theme. We have taken the position that the behavior of a person arises from a collection of causes scattered across their brain and the world that impinges upon it. So too does the behavior of a group arise from the contributions of the individuals comprising the group and the external circumstances in which they find themselves. This suggests that some of the tools for thinking about the causes of individual behavior can be applied to thinking about the origins of group behavior. When we think of corporate responsibility, for example, perhaps we should compare the employees of that corporation with the neurons of an individual brain. When we think of national responsibility, perhaps we should regard the whole nation as a group agent, and individual governmental bodies as components whose collective action determines, in complex ways, the national policies and acts that we may find morally blame- or praiseworthy.

Phillip Pettit thinks that there are such things as group agents and principled reasons for holding them responsible, as wholes, for their behavior [57], [58]. His case is too rich to summarize in its entirety here, but there are two features that we find particularly useful in setting a forward-looking research agenda. The first is his criterion for the existence of a group agent: "[I]f a set of individuals are to constitute a group agent…then they cannot rely on the group attitudes being formed on the basis of corresponding attitudes among the members." In other words, if the actions of the group are merely translations of an attitude held by the members, then the group is not a group agent. The second interesting feature is Pettit's requirement of rational agency: briefly, the group must be in a position to make value attributions and undertake actions intended to bring about states of affairs in line with those values.

With respect to the first feature: Pettit focuses on majoritarianism, but that is just one of many ways that a group decision could be an uninteresting function of its components. Another example would be a weighted voting scheme in which each member's vote is weighted by their status in the group. That is not majoritarian because a minority view could prevail if the members of the minority had sufficiently high status. More generally, a group output that depends only on linear functions of member output will fail to yield the sort of computational power we normally require before attributing agency to the group.

The second of Pettit's conditions, the requirement of rational agency, entails a feedback mechanism in which the group can produce a behavior, observe the effects of that behavior, and use the observed effects to refine future acts. A strictly "feedforward" mechanism wherein the group blindly churns through a computational procedure over its members, generating behavior after behavior but never learning or adapting to achieve its ends, would fail to yield the sort of flexibility we normally require before attribute *rational* agency to the group.

Together, requirements for nonlinearity and feedback may be unfounded bias, or they may be grounded in some sort of more principled distinction regarding certain classes of computational mechanisms. We don't know which one is the case, but we do not think it is a coincidence that the field of complexity science mentioned above takes each of these systemic properties to be key properties of computationally complex systems (see, for example, [59]). We therefore foresee a potential convergence between the philosophical study of group agency and the computational study of complex systems. Such convergence may help clarify, among other things, the conditions under which a system ought to be regarded as autonomous and the ways in which the parts of that system should be held accountable for decisions of the whole.

As a third thread for future research, we note that many discussions of determinism and human behavior point to ethics and moral responsibility as the primary concern for the inquiry. Concerns here include those questions of application already raised—e.g. how shall we incorporate knowledge of the causes of behavior into coherent judgments of agency and holdings of moral responsibility?—as well as questions of impact: what effect, if any, does understanding the deterministic physical

origins of decision making have on the moral soundness of human behavior? Under the umbrella of experimental philosophy, Nahmias et al. probed the attitudes of non-philosophers on some of these questions and found a moderate majority (60-76%, depending) of them to be natural compatibilists: subjects held that even when an agent's actions were causally determined, the agent was still morally responsible for his actions [10]. One might interpret that result as meaning that (to take just one example) popular support for the moral foundations for retributive criminal justice would not be undermined by a belief in determinism. This interpretation, however, is muddied by a further finding by Nahmias et al., which is that these same subjects believed that even when an agent's actions were causally determined, the agent was still free to choose and could have done otherwise. Thus it is not clear that subjects even understood causal determinism in the way that philosophers do. If they didn't really understand the entailments of the question scenarios, their recorded attitudes regarding compatibilism or incompatibilism would be moot. The thread for this line of research, then, would be to clarify what models of agency are in use in the minds of the folk: how does it happen that many of them believe simultaneously in causal determinism and freedom to choose otherwise?

Our fourth proposal for future research concerns again moral responsibility, this time as it applies to rapidly emerging robot technology. Robot ethics is increasingly becoming an area of interest, particularly because robots play increasingly vital roles as domestic caregivers and even surgeons.[9] When a person is injured by a robot, the question naturally arises as to who is responsible. Depending on the situation, the responsible agent might be a human user of the robot, the robot's manufacturer, or perhaps—if the internal mechanisms are sufficiently complex—even the robot itself. These are just beginning to become live issues: In 2007, a surgical robot's arm broke off inside a patient's body during a prostate cancer procedure.[10] Some military robots are not remotely operated, but actually autonomous and able to move, select, and fire upon targets without any human intervention [62]. Clearly, as capacities for robot autonomy grow, a whole set of moral and legal issues arise (see [63]).

There is also much ongoing work that integrates computers and robots with biological brains. Recently, progress has been made towards "hybrid systems which integrate biological neurons and electronic components" [64]. A conscious human brain and its body are normally regarded as legal persons with both rights and responsibilities. If we replace parts of the brain with something else, without impairing its function, then we would seem to preserve at least some if not all of those rights and responsibilities. We may come to a point at which more than half of the brain or body is artificial, making the organism more robotic than human. William Lycan considers such a scenario involving (fictional) Henrietta, a normal human being who very gradually undergoes replacement of parts of her physiology until eventually her entire body is artificial [65]. Suffice to say in the future the assignment of responsibility might have to be distributed over a wide range of biological and non-biological arrangements or systems. This will require engagement with many difficult issues such as complexity, unpredictability, determinism, responsibility, personhood and free will.

Our fifth and final proposal for future research is motivated by the perplexing observation that the deterministic nature of human thought applies to everything we think, including determinism itself. The determinist paints herself into a corner and must in the end say that her arguments for determinism, whatever their persuasive power, are as inevitable as are the indeterminist's arguments for indeterminism. And yet the determinist clearly will make the case that she, and not the indeterminist, is right. If she is, *why* is she? Why would *that* human mind, but not its opponents, come to be organized so as to possess capacities making it inevitable that its conclusions about the universe were correct? What forces of evolution gave rise to this singular outcome? More generally, what is the most useful way of understanding goal-directed projects—scientific, philosophical, mathematical, etc.—given that the goals themselves, the methods chosen to achieve them, and the criteria applied in deciding among solutions, are all determined by natural events?

Doubtless there are there are many other avenues to pursue. Hume himself, at the end of his discussion of liberty and necessity, acknowledges that there was much more to be said on the topic:

> *I pretend not to have obviated or removed all objections to this theory, with regard to necessity and liberty. I can foresee other objections, derived from topics which have not here been treated of. It may be said, for instance, that, if voluntary actions be subjected to the same laws of necessity with the operations of matter, there is a continued chain of necessary causes, pre-ordained and pre-determined, reaching from the original cause of all to every single volition, of every human creature. No contingency anywhere in the universe; no indifference; no liberty. While we act, we are, at the same time, acted upon (EHU 8.2; SBN 99).*

Some may find the notion of a pre-determined universe that Hume describes depressing. William James, for one, believed that a thorough-going determinism led necessarily to a bleak and pessimistic world view [66]. On the contrary, we contend that it is a lack of understanding of determinism that leads to a pessimistic outlook. Some people are depressed by the idea that everything is predictable; but that is not an entailment of determinism because of the celebrated *butterfly effect:* nonlinear dynamical systems, including the brain, are sensitive to initial conditions that cannot, even in principle, be measured with infinite precision. This means that many details of the future are probably forever beyond our predictive powers.

---

[9] For more on the implications of robots as caregivers, see [60] and for robotic surgery see [61].

[10] http://www.newscientist.com/article/mg21729105.800-robosurgeon-da-vinci-faces-lawsuits.html

Other people are depressed by the idea that determinism means that we play no part in the future. This too is mistaken. A. J. Ayer makes the point well at the end of his of his classic essay, "Freedom and Necessity" [67]. He says that determinism entails that the future is in some sense "already decided" before we choose. But we wrongly infer from this that what we choose will make no difference to what occurs—in other words, we wrongly infer fatalism. In fact, the future is only "decided" in the sense that it is predictable (*pace* the butterfly effect) using true universal generalizations (laws of nature) and it is possible to deduce from a set of facts about the past together with general laws to predict future events. Even if this is true, it does not entail that we are somehow prisoners of fate. It is false that what we choose makes no difference to what occurs in the future; for our choices are causes as well as effects: if our choices were different then the consequences could be different. The predictable outcomes only occur because we predictably choose as we do.[11]

Without crediting the appeal to emotion that underwrites these reluctances to accept a set of empirical claims, we note that an individual might, by these lights, actually be encouraged by determinism.[12] By understanding and increasing our awareness of the many factors that influence our behavior as natural beings, we may become more skilled and proactive in our decision making. We think there is much to be said in favor of a standpoint that not only explains so much about our experiences and practices, but also gives us tools to improve them.

ACKNOWLEDGMENT

Preliminary forms of some of these ideas appear in an earlier work (Serbian language only) by JF—see [69]. Thanks to Alex Sager for helpful comments on an earlier draft.

REFERENCES

[1] B. d'Holbach, A System of Nature. 1770. [Online]. http://www.gutenberg.org/files/8909/8909-h/8909-h.htm#link2H_4_0004.
[2] C. Campbell, In Defence of Free Will. London: Allen and Unwin, 1967.
[3] G. Strawson, Interview in "The Believer," 2003. [Online]. http://www.believermag.com/issues/200303/?read=interview_strawson.
[4] G. Strawson, "The impossibility of moral responsibility," Philos. Stud., vol. 75, pp. 5–24, 1994.
[5] R. Clarke, Libertarian Accounts of Free Will. Oxford: Oxford University Press, 2003.
[6] R. Kane, The Significance of Free Will. New York: Oxford University Press, 1996.
[7] S. Nichols, "The indeterminist intuition: source and status," The Monist, vol. 95, no. 2, pp. 290–308, 2012.
[8] G. Caruso, Free Will and Consciousness: A determinist account of the illusion of free will. Lexington Books, 2012.
[9] C. Lamont, Freedom of Choice Affirmed. Beacon Press, 1969.
[10] E. Nahmias, S. Morris, T. Nadelhoffer, and J. Turner, "Surveying freedom: folk intuitions about free will and moral responsibility," Philos. Psychol., vol. 18, no. 5, pp. 561–584, 2005.

[11] For an updated take of Ayer's view see the chapter on free will in [68].

[12] See the end of Part 2 in [37] for an explanation of the benefits of accepting Spinoza's own brand of determinism.

[11] D. Hume, A Treatise of Human Nature, D. Norton and M. Norton, Eds. Oxford: Oxford University Press, 2007.
[12] D. Hume, A Treatise of Human Nature, 2nd ed., L. Selby-Bigge and P. Nidditch, Eds. Oxford: Clarendon Press, 1978.
[13] D. Hume, An Enquiry Concerning Human Understanding, T. Beauchamp, Ed. Oxford University Press, 1999.
[14] D. Hume, An Enquiry Concerning Human Understanding and Enquiry Concerning the Principles of Morals, 3rd ed., L. Selby-Bigge and P. Nidditch, Eds. Oxford: Clarendon Press, 1975.
[15] A. Shariff, J. Schooler, and K. Vohs, "The hazards of claiming to have solved the hard problem of free will," in Are we free? Psychology and free will, J. Baer, J. C. Kaufman, and R. F. Baumeister, Eds. Oxford University Press, 2008.
[16] T. Clark, "Fear of mechanism: A compatibilist critique of The Volitional Brain," J. Conscious. Stud., vol. 6, no. 8–9, pp. 279–283, 1999.
[17] A. Schopenhauer, On the Freedom of the Will. Oxford: Basil Blackwell, 1839.
[18] R. Chisholm, Human Freedom and the Self. Lawrence, KS: , 1964.
[19] B. Libet, C. Gleason, E. Wright, and D. Pearl, "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act.," Brain, vol. 106, pp. 623–664, 1983.
[20] M. Matsuhashi and M. Hallett, "The timing of the conscious intention to move," Eur. J. Neurosci., vol. 28, pp. 2344–2351, 2008.
[21] P. Haggard and M. Eimer, "On the relation between brain potentials and the awareness of voluntary movements," Exp. Brain Res., vol. 126, pp. 128–133, 1999.
[22] H. Lau, R. Rogers, P. Haggard, and R. Passingham, "Attention to intention," Science, vol. 303, no. 5661, pp. 1208–1210, 2004.
[23] A. Roskies, "How does neuroscience affect our concept of volition?," Annu. Rev. Neurosci., vol. 33, pp. 109–130, 2010.
[24] A. Sirigu, E. Daprati, S. Ciancia, P. Giraux, N. Nighoghossian, A. Posada, and P. Haggard, "Altered awareness of voluntary action after damage to the parietal cortex," Nat. Neurosci., vol. 7, no. 1, pp. 80–84, Nov. 2003.
[25] M. Desmurget, K. T. Reilly, N. Richard, A. Szathmari, C. Mottolese, and A. Sirigu, "Movement intention after parietal cortex stimulation in humans," Science, vol. 324, no. 5928, pp. 811–813, May 2009.
[26] K. Ammon and S. C. Gandevia, "Transcranial magnetic stimulation can influence the selection of motor programmes.," J. Neurol. Neurosurg. Psychiatry, vol. 53, no. 8, pp. 705–707, Aug. 1990.
[27] J. P. Brasil-Neto, A. Pascual-Leone, J. Valls-Sole, L. G. Cohen, and M. Hallett, "Focal transcranial magnetic stimulation and response bias in a forced-choice task.," J. Neurol. Neurosurg. Psychiatry, vol. 55, no. 10, pp. 964–966, Oct. 1992.
[28] W. Newsome, K. Britten, and J. Movshon, "Neuronal correlates of a perceptual decision," Nature, vol. 341, pp. 52–54, 1989.
[29] P. S. Churchland, "On the alleged backwards referral of experiences and its relevance to the mind-body problem," Philos. Sci., vol. 48, pp. 165–181, 1981.
[30] A. Batthyany, "Mental causation and free will after Libet and Soon: Reclaiming conscious agency," in Irreducibly Conscious: Selected papers on consciousness, Heidelberg: Universitätsverlag Winter, 2009.
[31] D. Dennett, Freedom Evolves. Allen Lane, 2003.
[32] W. Klemm, "Free will debates: Simple experiments are not so simple," Adv. Cogn. Psychol., vol. 6, no. 1, pp. 47–65, 2010.
[33] A. Mele, Autonomous Agents: from self-control to autonomy. Oxford University Press, 1995.
[34] A. Mele, "Free will: action theory meets neuroscience," in Intentionality, Deliberation, and Autonomy: The Action-theoretic Basis of Practical Philosophy, C. Lumer, Ed. Ashgate, 2007, pp. 257–272.
[35] A. Mele, "Libet on Free Will: Readiness Potentials, Decisions, and Awareness," in Conscious Will and Responsibility, W. Sinnott-Armstrong and L. Nadel, Eds. Oxford University Press, 2011.
[36] J. Kim, Mind in a physical world. Cambridge, Mass.: MIT Press, 1998.
[37] B. Spinoza, "Ethics," in Baruch Spinoza: The complete works, S. Shirley, Ed. Indianapolis: Hackett University Press, 2002.

[38] R. Holton, “The act of choice,” Philos. Impr., vol. 6, pp. 1–15, 2006.
[39] A. Clark, Mindware: An introduction to the philosophy of cognitive science. New York: Oxford University Press, 2001.
[40] P. Haggard, “Conscious intention and motor cognition,” Trends Cogn. Sci., vol. 9, no. 6, pp. 290–295, Jun. 2005.
[41] D. M. Wegner, “Précis of The illusion of conscious will,” Behav. Brain Sci., vol. 27, no. 05, Mar. 2005. [Online]. http://www.journals.cambridge.org/abstract_S0140525X04000159.
[42] F. Gibbons, “Self-attention and behavior: a review and theoretical update,” in Advances in Experimental Social Psychology, vol. 23, M. Zanna, Ed. Academic Press, 1990, pp. 249–303.
[43] D. C. Dennett, Elbow Room: The varieties of free will worth wanting. MIT Press, 1984.
[44] J. Greene and J. Cohen, “Neuroscience changes nothing and everything,” Philos. Trans. R. Soc. Lond. B, vol. 359, pp. 1775–1785, 2004.
[45] S. J. Morse, “The non-problem of free will in forensic psychiatry and psychology,” Behav. Sci. Law, vol. 25, no. 2, pp. 203–220, Mar. 2007.
[46] Roper v. Simmons, vol. 543. 2005, p. 551.
[47] N. Farahany, “Criminal minds: the use of neuroscience as a defense skyrockets,” U.S. News and World Report, 2012. [Online]. http://www.usnews.com/news/articles/2012/11/09/criminal-minds-use-of-neuroscience-as-a-defense-skyrockets.
[48] E. Aharoni, C. Funk, W. Sinnott-Armstrong, and M. Gazzaniga, “Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience.,” Ann. N. Y. Acad. Sci., vol. 1124, pp. 145–160, Mar. 2008.
[49] L. Miller, “Posttraumatic stress disorder and criminal violence: Basic concepts and clinical-forensic applications.,” Aggress. Violent Behav., vol. 17, pp. 354–364, 2012.
[50] K. Kelland, “Neuroscience in court: my brain made me do it,” Reuters, 2012. [Online]. http://www.reuters.com/article/2012/08/29/us-neuroscience-crime-idUSBRE87S07020120829.
[51] T. Nadelhoffer, S. Grafton, K. Kiehl, A. Mansfield, W. Sinnott-Armstrong, and M. Gazzaniga, “Neuroprediction, Violence, and the Law: Setting the Stage,” Neuroethics, vol. 5, p. 67, 2012.
[52] P. S. Applebaum, “Assessment of patients’ competence to consent to treatment.,” N. Engl. J. Med., vol. 357, no. 18, pp. 1834–1840, 2007.
[53] L. Rabin, M. Borgos, and A. Saykin, “A survey of neuropsychologists’ practices and perspectives regarding the assessment of judgment ability,” Appl. Neuropsychol., vol. 15, pp. 264–273, 2008.
[54] V. Raymont, W. Bingley, A. Buchanan, A. David, P. Hayward, S. Wessely, and M. Hotopf, “Prevalence of mental incapacity in medical inpatients and associated risk factors: cross-sectional study,” Lancet, vol. 364, pp. 1421–1427, 2004.
[55] C. Langton, “Computation at the edge of chaos,” in Emergent Computation, S. Forrest, Ed. Cambridge, Mass.: MIT Press, 1991.
[56] S. Kauffman, The Origins of Order. New York: Oxford University Press, 1993.
[57] P. Pettit, “Responsibility incorporated,” Ethics, vol. 117, pp. 171–201, 2007.
[58] P. Pettit, “The reality of group agents,” in Philosophy of the Social Sciences: Philosophical theory and scientific practice, C. Mantzavinos, Ed. Cambridge University Press, 2009.
[59] S. Strogatz, Nonlinear Dynamics and Chaos. Westview Press, 2001.
[60] J. Borenstein and Y. Pearson, “Robot caregivers: harbingers of expanded freedom for all?,” Ethics Inf. Technol., vol. 12, no. 3, pp. 277–288, Sep. 2010.
[61] S. Najarian and E. Afshari, “Evolution and future directions of surgical robotics,” Int. J. Clin. Med., vol. 3, no. 2, 2012.
[62] P. Finn, “A future for drones: automated killing,” The Washington Post, 2011. [Online]. http://articles.washingtonpost.com/2011-09-19/national/35273383_1_drones-human-target-military-base.
[63] P. Lin, K. Abney, and G. Bekey, Robot Ethics. MIT Press, 2012.
[64] K. Warwick, D. Xydas, S. Nasuto, V. Becerra, M. Hammond, J. Downes, M. Marshall, and B. Whalley, “Controlling a mobile robot with a biological brain,” Def. Sci. J., vol. 60, no. 1, pp. 5–14, 2010.
[65] W. Lycan, “Qualitative experience in machines,” in How Computers are Changing Philosophy, T. W. Bynum and J. H. Moore, Eds. Blackwell, 1998.
[66] W. James, “The dilemma of determinism,” in The Will to Believe and other essays in popular philosophy, Cambridge, Mass.: Harvard University Press, 1979.
[67] A. J. Ayer, “Freedom and necessity,” in Philosophical Essays, London: Macmillan, 1954.
[68] S. Blackburn, Think: A compelling introduction to philosophy. Oxford University Press, 1999.
[69] J. Fost, “Empirical and theoretical issues in the predictability of human behavior,” MD Explorer, 2011. [Online]. http://www.mdexplorer.rs/predvidljivost-ljudskog-ponasanja/.