Received: 29 May 2020 | Revised: 16 July 2020 | Accepted: 14 August 2020

DOI: 10.1111/jtsb.12250

ORIGINAL ARTICLE

WILEY

Addiction science and the perception of freewill

John R. Monterosso[1,2] | **Barry Schwartz**[3]

[1]Department of Psychology, University of Southern California, Los Angeles, California, USA

[2]Neuroscience Graduate Program, University of Southern California, Los Angeles, California, USA

[3]Haas School of Business, Berkeley, California, USA

Correspondence
John R. Monterosso, SGM 501, 3620 McClintock Ave, Los Angeles, CA 90089, USA.
Email: johnrmon@usc.edu

ABSTRACT

Do people think individuals with Substance Use Disorder (SUD) have freewill? How does addiction science inform views on the issue? We distinguish between two senses of freewill: 1) **libertarian freewill**, in which freedom turns on a particular metaphysical conception of action (sometimes operationalized as "*could have done otherwise*"), and 2) **compatibilist freewill**, in which freedom depends on the relation between the actor's psychology and her actions (e.g., "*was the act what she wanted to do?*"). We argue that, in different ways, scientific accounts can impact conceived freewill by linking addictive behavior to mechanisms that observers view as peripheral to the actor (**motivation modularity)**. While a variety of impacts on conceived compatibilist freewill are plausible, we argue that contemporary addiction science has no direct bearing on conceived libertarian freewill. Addiction science may, however, indirectly impact conceived libertarian freewill by priming an explanatory framework in which intention is superfluous (especially materialism).

KEYWORDS
addiction, compatibilism, determinism, freewill, naïve dualism, substance use disorder (SUD)

Do people view addiction as entailing loss of freewill? The term "addiction" already suggests some collective belief of this sort. The root of the term is linked to the Greek folktale of Addictus, a slave who kept his chains on even after his master unlocked them to set him free. However, in the English language from at least the time of Shakespeare (McQuain, Malless, &

Shakespeare, 1998) through the latter part of the nineteenth century, its usage was more a designation of commitment to something, without a particularly pejorative connotation. For example, "Thomas James of Oxford, a learned man and a great lover of books, who, wholly addicted to learning..." (Wood, 1796, p. 296). According to the NGrams tool for analyzing word frequency in the "GoogleBooks" database, this adjective form ("addicted") was used at a far higher frequency throughout the 17th and 18th centuries than it is today. However, the noun form "addiction" did not begin to enter common use until early in the 20th century. This shift in language may represent a partial medicalization. Whereas the adjective form describes the person's behavior, the noun form "addiction" allows separability from the individual. It facilitates (though does not necessitate) talking about the undesirable behavior as a symptom – a result of a disorder, rather than a description of behavior. There is no similar linguistic form in common usage for, say, murder or pedophilia. Nevertheless, survey questions probing conception of addiction indicate that people are conflicted, physicians included (Lawrence, Rasinski, Yoon, & Curlin, 2013), about whether or not addiction is a disease and whether it entails lack of freewill (Vohs & Baumeister, 2009). Here we examine the latter issue in relation to addiction science. In what way does addiction science inform views about addiction and freewill? We begin by identifying distinctions in how people think about freewill generally, and then use those distinctions to offer suggestions regarding the impact of addiction science on conceptions of freewill.

1 | VARIETIES OF FREEWILL

A good starting point in summarizing conceptions of freewill is the recognition of the categorically distinct way in which we understand the behavior of animate entities in the world. Animate objects act with purpose. We intuitively see one another as agents with, "beliefs and desires and other mental states... and whose actions can be explained or predicted on the basis of the content of these states." (Dennett, 1993, p. 76) Intentional action is directed at a goal (e.g., 'Jack is grabbing the apple', or 'Mary is trying to hurt Jane'). The observer is said to understand a behavior within this "intentional framework" when she recognizes both the goal being pursued, as well as beliefs and desires that are consistent with that goal. This is wholly different from the way non-intentional events are understood.

Viewing behavior from the intentional framework is connected with the conception of behavior as an act of "freewill," but the nature of the connection depends on what precisely is meant by freewill. It is helpful here to distinguish two senses of freewill that are captured by the philosophical traditions of, on the one hand, **libertarianism**, and the other, **compatibilism**. Although we are ultimately concerned with everyday understanding of freewill (particularly in the context of addiction) these philosophical traditions are useful because each appears to match some portion of everyday intuitions (Deery, Davis, & Carey, 2015; Feltz & Cokely, 2009; Feltz, Cokely, & Nadelhoffer, 2009; Nahmias, 2011). Our discussion is not intended to present a full context of relevant philosophical positions on freewill, only to highlight two that we think are useful for thinking about the impact of contemporary addiction science. According to libertarianism, freewill turns on a particular metaphysical conception of intentional behavior that sets it apart from other occurrences in the material world. It is typically operationalized either by the standard of, '*could have done otherwise*'—free actions as occurring at choice points from which multiple directions could have been traveled (sometimes likened to "forking paths", a reference to Borges' character Ts'ui Pên's worlds endlessly dividing; Borges, 1962), or else by the

actor's role as the *originating source* of her action—a "mover unmoved", rather than just a link in an infinite causal chain (Horgan & Timmons, 2011). Libertarian freewill is particular to intentional-framework explanations. Events of the world, human or otherwise, that are not understood with reference to goals and mental states (falling rocks, dilating pupils, etc.) are not conceived as free in the libertarian sense.

Libertarian freewill is incompatible with determinism, which is, "the idea that every event is necessitated by antecedent events and conditions together with the laws of nature" (Hoefer, 2016, p. 1). A universe in which all events, including human behavior, follow by necessity from antecedents admits no forking paths. However, some philosophers with a deterministic metaphysics, rather than dismissing the notion of freewill, have instead argued for a conception of it that is compatible with determinism (e.g., Hobbes, Spinoza, and many modern philosophers including Peter Strawson, Gary Watson, and Daniel Dennett). According to *compatibilism*, what makes an act free is its relation to the intentions, beliefs and motivations of the actor (the elements of intention-framework explanation). If Mary is hurting Jane, but doing so is in complete conflict with her mental state (maybe she wishes nothing more than *not* to hurt Jane), then she does not do so freely. Compatibilism shifts freewill from a metaphysical issue to a psychological one: *Is the action consistent with (something about) the mental states, beliefs and goals of the actor?*

A primary challenge for compatibilists is specifying precisely what it is about the relation between action and mental states upon which freedom hinges. Are all desires equivalent, as suggested by Hobbes's general freewill-standard (Hobbes, 1968) of the actor finding "no stop, in doing what he has the will, desire, or inclination to do"? Or, as suggested by Frankfurt, should we consider a taxonomy that prioritizes "higher-order desires" (such as the alcoholic's wish to stop wanting alcohol) over "first-order desires" (wanting alcohol), such that engagement in some desired acts can be unfree (Frankfurt, 1971)? Should it depend on sensitivity of the behavior to good reasons, in the sense that the actor would behave differently if the right sort of reason were introduced (a substitution of the libertarian metaphysical test "could have done otherwise" with the determinism-compatible "would have done otherwise") as suggested by Dennett (1984)? Should it be grounded, as P.F. Strawson proposed, in the "reactive attitudes" that observers have to intentional-framework inferences, such as resentment and gratitude (Strawson, 1974)? This remains an active debate among compatibilists (McKenna, 2012). But the primary proposals share a reliance on the inferred psychology of the actor as the critical basis for judging freedom. Thus, the intentional framework is as central to compatibilist freewill as it is to libertarian freewill.

2 | EVERYDAY INTUITIONS OF LIBERTARIANISM AND COMPATIBILISM

How do these notions of freewill compare to everyday intuitions? One of us once brought his daughter near tears through explication of determinism (denial of libertarian freewill) as captured by Laplace's famous formulation:

An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace, 1814).

How could I say something so outrageous, she wondered, particularly with regard to human behavior? It seemed to her as if I were revealing secret membership in a cult with a bizarre worldview. Although it is difficult to articulate precisely what it means, people may intuitively feel their actions have a quality along the lines of libertarian freewill. As philosopher Galen Strawson put it, libertarian freewill is, "the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess" (Strawson, 1994, p. 30). Determinism's explicit denial of it can be unsettling.

However, even granting the claim that most people view action as sometimes free in the libertarian sense, it does not follow that this is the only, or even the primary issue people have in mind when they talk about freewill and responsibility. A compatibilist sense of freewill -- some variant of the psychological question of whether the action was what the actor wanted to do—may also be important, or even central. Indeed, this has been borne out in studies in "experimental philosophy" that explore intuitions about freewill. For example, Nahmias and colleagues presented scenarios designed to establish a particular human action (e.g., keeping vs. returning a found wallet) was fully determined by causal antecedents. While about a quarter of participants judged the determined actor *not* to have had freewill, the majority of participants judged the act to be free despite it being stipulated to be deterministic (Nahmias, Morris, Nadelhoffer, & Turner, 2006). This suggests that some other sense of freewill, likely along the lines of some variant of compatibilism, was important in their intuitive judgments of freedom. Along similar lines, Chapman and Neidermayer (2001) explored the conditions under which people judged an actor to have "decided" to act in some way. For example, in one pair of scenarios, the actor, on a diet, passes by the reception desk at work in the morning and either takes or refuses to take a donut. What Chapman and Neidermayer found, with a series of scenarios like this, is that respondents took evidence of "reflection" as indications that a decision had been made. If one thinks about it, it is a freely chosen decision; if not, it is some other sort of action, perhaps habitual, but certainly thoughtless. We will return to these issues in the context of addiction science after considering the more general issue of how scientific explanation can impact notions of freewill.

3 | SCIENCE OF BEHAVIOR WITHOUT THE INTENTIONAL FRAMEWORK

Intentional framework explanations of behavior may be innate, but they are not the only way that behavior can be understood. In particular, science provides approaches to explaining behavior, even at the level of the individual, without reliance on the intentional framework. Two such approaches where this is particularly evident are behaviorism and cellular neuroscience. Behaviorism dominated American psychology for more than three decades and it continues to be highly influential within addiction science. Among behaviorism's most influential theorists was B.F. Skinner (e.g., Skinner, 1954; Skinner, 1971; see Schwartz, 1978). When Skinner began his research on conditioning, the dominant figure in the field was Pavlov (1927), the Russian physiologist who discovered that through so-called laws of association, previously neutral, insignificant stimuli could come to trigger reflexes (e.g., salivation or eye blink) that were previously triggered by stimuli (food in the mouth; a puff of air to the eye) the causal role of which was wired into the organism. These "conditioned reflexes" extended the range of stimuli that might trigger them. Motivations and intentions had no place in explaining conditioned salivation or eye blinks. Indeed a test that you were studying an involuntary reflex

was that if you made it worth the organism's while not to salivate or blink its eye it would do so anyway (e.g., Sheffield, 1965; Williams & Williams, 1969).

What Skinner added to Pavlov's edifice was a set of principles of behavior that was not triggered by antecedent stimuli, but instead triggered by consequences. What E.L. Thorndike had called "the law of effect," (Thorndike, 1898) Skinner transformed into laws of operant or instrumental conditioning. The significance of this extension cannot be overestimated. Skinner was trying to explain prototypical goal-directed behavior by appealing **not to the mental states of the intentional-framework** but instead to past experiences engaging in that behavior and the positive or negative consequences that had followed. For Skinner, so-called "voluntary" behavior was as automatic as the salivation of Pavlov's dogs. Indeed, Skinner's most popular book, *Beyond Freedom and Dignity* (1971) was an extended argument that the status granted intentional behavior, including presumed personal responsibility, were just figments of our collective imagination, to be replaced by scientific laws when psychology made sufficient progress. And descendants of Skinner endeavored to show that even when organisms could "freely" choose among a set of different behaviors, their choices could be explained mechanistically, and described mathematically, without reference to mental states and intention. Skinner's work was perhaps the most sustained and deliberate effort within psychology to displace the intentional framework. From his point of view, people could supply the "reasons" for their actions, but such reasons were essentially inventions, or as Skinner called them, "explanatory fictions," themselves in need of causal explanations. Indeed, people could provide reasons for their pets' behavior (e.g., "our dog sat on command because she knew that doing so would get her a treat and she wanted the treat.") For Skinner, these intentional-framework accounts were causally inert, epiphenomenal manifestations of how people learn to talk about their behavior and the behavior of at least some other animate entities. Real explanations were there, and to find them, researchers had to ignore intentional framework accounts. Although behaviorism is no longer a dominant force in psychology, it continues to be an influential perspective within addiction science, where its technologies and experimental methods (e.g., Skinner-boxes and reinforcement schedules) remain central.

A second scientific approach to explaining behavior that does not include the intentional framework is what we refer to as cellular neuroscience. By "cellular" we mean to exclude areas within neuroscience that incorporate psychological-level constructs, such as cognitive neuroscience. Cellular neuroscience explains the workings of neurons, which, in coordination with other cells and signaling molecules of the body, accomplish the inconceivably complex work of orchestrating human action. The activity of the neuron is explained by the biochemistry of ions flowing through porous channels, not by intentions. Cellular neuroscience implies the theoretical possibility of explaining *all* behavior with no appeal to the intentional framework. However, the nervous system is too complex by orders of magnitude for this to be anything more than a theoretical possibility.

4 | SCIENTIFIC EXPLANATIONS AND HIDDEN MOTIVES

Not all scientific explanation of behavior dispenses with the intentional framework; indeed most does not. More often scientific explanation changes the way action is understood with respect to psychological constructs. One theme of the past few decades of behavioral science is that the individual's intentional framework explanations for her own actions are sometimes demonstrably wrong. Perhaps the most remarkable demonstration of this comes from work

with patients that have had the white-matter tracts severed that would otherwise connect the left and right cortical hemispheres. This condition allows the possibility to present information to a single hemisphere (Gazzaniga, 1967). Suppose information driving a response is presented to the right hemisphere (e.g., a visual instruction to grab a cup). When the patient is asked to explain her behavior, the explanation she gives will be controlled by the left hemisphere since in most individuals, the left hemisphere is dominant in language production. But without the pathway connecting the cortical hemispheres, the left hemisphere has no representation of the key antecedent (the presented instruction). How does the left hemisphere explain a behavior for which it has no access to the true cause? One might expect the patient to express being dumbstruck, "I have no idea why I did that." But instead, the usual response is an explanation like, "I am a little thirsty." Such confabulations appear to come naturally to the patient, so much so that it suggests the reasons people give for behavior in general are less direct readouts of true reasons than they are *interpretations*. Indeed, Gazzaniga and colleagues suggest the intentional framework explanations of our own behavior are the product of an interpreter module with incomplete access to relevant information (Gazzaniga & LeDoux, 1978). Notice here that the true explanation is still understood within the intentional framework. The individual (or perhaps more correctly, her right cortical hemisphere) believed the experimenter was telling her to grab the cup and she wanted to do what was asked of her. However, the mental states that best explain the behavior are hidden.

A great deal of psychology, and especially social psychology, can be viewed as a series of attempts to illuminate the mental states that truly drive our behavior, as distinct from those that we think drive our behavior (Nisbett & Wilson, 1977). For example, in the domain of moral judgment, Haidt and other "moral intuitionists" argue that people have no trouble providing reasons to explain their moral judgments, but those reasons are not the real source of the judgments (Haidt, 2001, 2006; Haidt & Joseph, 2004). Instead, the source is usually a set of deeply held intuitions, perhaps rooted in evolutionary biology, of which people are often unaware. "Reason," Haidt says, "is a lawyer not a judge." It is used to justify or defend moral judgments rather than arrive at them. In a similar vein, a half century of research in behavioral decision making has made clear that people's "rational" decisions are often influenced, if not completely determined, by factors that are quite distinct from those that people themselves identify as the reasons behind their decisions (e.g., Kahneman, 2011). This kind of thinking and theorizing has led to a wide range of applications, often called "nudges," designed to help people make better decisions in the face of convincing evidence that giving people good reasons to make those decisions will not do the job (Thaler & Sunstein, 2003). Again, none of this work imagines a way of understanding behavior that makes the elements of the intentional framework—beliefs, motivation and goals—irrelevant. But each places space between the actor's experience of her own intentional behavior, and the true motivations that drive the behavior. How such "hidden motive" intentional-framework explanations impact conceptions of freewill is, we think, an open question.

One reason it is an open question is the extent to which freewill as the default for understanding human action depends on culture. Foucault (e.g., 1973; 1973; 1978; see also Rose, 1996) argued that modern notions of freedom depend upon the conception of human action as the product of rational self-interest, which came to prominence in Western thinking during the enlightenment. According to Ferentzy (2002), Foucault had identified a critical cultural development for contemporary notions of addiction and freewill—a shift away from conceiving the relevant struggle as between external forces of good and evil, but rather as the struggle between psychological forces within the individual (Ferentzy, 2002). The critical role of culture

in informing conceptions of freedom is supported by empirical work in social psychology. In a series of studies, Savani, Markus, Naidu, Kumar, and Neha (2010) showed that when asked to reflect on their own recent mundane behavior or the mundane behavior of another person observed on videotape coming into the house, changing clothes, and making a snack, and then judge how many choices they (or the observed other) had made, Americans interpreted many more actions as "chosen" than did people of Indian descent. Whereas "chosen" acts are presumably the product of freewill, unchosen acts may not be. The point of both the philosophical argument and the psychology evidence is that intuitions about freewill may depend on deep cultural background factors.

5 | SCIENTIFIC EXPLANATIONS AND ANTECEDENTS TO PSYCHOLOGICAL STATES

Another way that behavioral science advances without completely displacing intentional framework explanations is by illuminating antecedents of intentional-framework elements (beliefs, motives and goals). For example, work in attachment theory suggests that the behavior of the primary caretaker towards the infant and child will impact the child's development, affecting the quality of her other relationships into her adult life. An unreliably available primary caretaker makes it more likely that the child, and perhaps eventual adult, will be excessively clingy in her relationships. Far from displacing the intentional framework, these antecedent causes are understood by their inferred effect on the individual's mental states.

Increasingly, antecedents identified in the behavioral sciences are biological, but this does not mean they involve displacement of the intentional framework (unlike cellular neuroscience). For example, there is some evidence that suicide is associated with lower levels of serotonin, as inferred from its metabolites in cerebrospinal fluid (Samuelsson, Jokinen, Nordstrom, & Nordstrom, 2006). If true, the plausible interpretation of this statistical relationship is not that it makes the feelings and intentions of the actor irrelevant, but rather that the low availability of serotonin is an antecedent that explains something about suicidal feelings or suicidal decision-making. Or consider work on phobias. Molecular genetics appears to have statistically linked propensity for phobias to a particular genetic variant (the val158met polymorphism of the COMT gene (Howe et al., 2016)). But of course, if true, this does not imply a new explanation of phobic behavior that does not invoke the mental state of fear. Instead it represents the identification of an antecedent associated with fear.

While explanations in the behavioral sciences that illuminate antecedents to psychological states do not displace the intentional framework, they are nevertheless in logical conflict with libertarian freewill *if they are complete*. If a behavior is completely explained by antecedents, then what might have looked like a forking path—*the person could do X or Y*—is shown to instead be a glide rail—*given antecedent A, the person was bound to do Y*. Any action that yields to a complete causal account, logically, cannot satisfy the requirements for freedom in the libertarian sense. How people *actually* respond to full causal accounts may differ, but the logic of how it should impact conceived libertarian freewill is clear.

Of course, the determinants of beliefs, motives and goals are so complex that complete accounts are essentially non-existent. Human behavior is almost never fully predictable. Genetic and brain-based predictors like those noted above generally explain only a small fraction of variance in behavior. They are probabilistic. Complete causal accounts of behavior are inconsistent with libertarian freewill, but what are the logical implication for libertarian freewill

of a partial causal account? We think almost none at all. Imputing freewill in the libertarian sense is made no less logically coherent by the identification of antecedents if those antecedents only partially account for/predict behavior. Indeed, it is uncontroversial to anyone, libertarians included, that causal antecedents outside the actor's control exert *some* influence over her (e.g., the effect of a toothache or bad night's sleep on how patient a parent is with her children). So, why should partial causal accounts pose any challenge? Freewill, it might even be maintained, is the reason behavior is ultimately only probabilistic with respect to antecedents.

However, we think as an *empirical matter*, it is possible that some scientific accounts do have an impact on conceived freewill in the libertarian sense, despite being probabilistic. In part, this intuition is informed by previous work we conducted in which such probabilistic causal explanations for blameworthy behavior (e.g., violent crime) some of which were biological, were presented to participants (Monterosso et al., 2005). Subsequent to hearing about a scenario and answering questions about responsibility and volition, participants were asked several questions like the following, "[The scenario] said that 80% of people with this chemical imbalance behaved in this way. That means that 20% did not. What do you think is different about this group?" Some participants rejected the possibility that the antecedent was actually equally true of the set of individuals who had not behaved like the actor in the vignette. This was the case for one participant who reconciled these individuals by explicitly conjecturing that for them, "... it wasn't as extreme as hers." Another participant appeared to have had the same thought, but first checked out this possibility with the interviewer in the following dialog:

> Interviewer: "It said that 20% of the people with this chemical imbalance behave in this way. That means that 80% do not. What do you think is different about the 80%?" Participant: "They have the same chemistry, to the same degree?" Interviewer: "Right." Participant: [long pause] "Geez, I would say that it would be an unexplainable scientific fluke."

This type of response was significantly more prevalent when the antecedent was biological rather than psychological (Monterosso et al., 2005). We suggested that the basis for this pattern was a "naïve dualism" (also referred to as "commonsense dualism"; Bloom, 2004) among observers, in which the physical world was conceived to be deterministic, but not the mental (psychological) world. While behavior is generally viewed as controlled by mental forces, people are also objects in the physical world. Highlighting a biological cause for a behavior may have primed observers to view it as an event of the material world—a **material-framework priming** effect. Or perhaps biological explanations have a more seductive appeal for some other reason. Whatever the basis, we take participants' responses as suggestive evidence that in some cases even partial causal accounts are sufficient to induce participants to treat behavior as deterministic, and so to treat it as unfree in the libertarian sense. We will return to this in the context of addiction.

Scientific explanations that illuminate antecedents (without bypassing the intentional framework) may also impact conceived freewill in the compatibilist sense, though here the issue does not turn on completeness of the account, but rather on how they inform intentional-framework understanding. For instance, if the behavior in question is cowardice, maybe identification of the biology of a panic attack makes the behavior seem less free by illuminating motivational forces that seem difficult to overcome. Importantly, it is not the mere existence of identified causal antecedents that makes the difference here, but rather the particular way the

account impacts the conceived relationship between the action and the mental states of the actor.

5.1 | Freewill and the Science of addiction

What about addiction, and the impact of addiction science on the conception of addictive behavior as free or unfree? Nora Volkow, director of the National Institute on Drug Addiction, is prominent among those who argue both that addiction undermines freewill, and that this conclusion is supported by science. Speaking at the 2015 annual convention of the American Psychiatric Associations, she argued that addiction entails "disruption of the areas of the circuits that enable us to exert freewill, that enable us to exert free determination."(Volkow, 2015) That position is aligned with that of previous NIDA director Alan Leshner, who urged the scientific community and public to "face the fact that even if the condition initially comes about because of a voluntary behavior (drug use), an addict's brain is different from a nonaddict's brain, and the addicted individual must be dealt with as if he or she is in a different brain state" (Leshner, 1997, p. 278). Presumably, by specifying that drug taking was *initially* voluntary, his implication is that once someone is addicted, that is no longer the case. And like Volkow, he implies here that neurobiological findings are important to the claim. NIDA has a research budget of over a billion dollars and thus can shape the majority of the world's scientific research on addiction. So, this shared position has some claim to "official position" for addiction science. We will refer to it as such, but it should be noted that the view is nevertheless subject to considerable dissent (Heim, 2014; Heyman, 2009; Lewis, 2017; Peele, 1989).

The basic contours of the official position are not new. Benjamin Rush, a signatory of the Declaration of Independence and Surgeon General of the Confederate Army expressed a similar view in 1812. Discussing what he saw as a growing problem with distilled alcohol consumption (mostly whisky), he wrote, "The use of strong drink is at first the effect of free agency. From habit it takes place and from necessity." (Rush, 1812, p. 264). By this Rush meant to express the same position as Volkow and Leshner; that drug use starts out as voluntary, but that it progresses to something different. And like Leshner and Volkow, he considered the disease concept to have much to overcome with respect to public acceptance, "I am aware that the efforts of science and humanity, in applying their resources to the cure of a disease induced by a vice, will meet with a cold reception from many people" (Rush, 1784, p. 29).

The fact that, nearly a century before the discovery of neurons, a physician could stake out a position that in some ways mirrors that of contemporary addiction scientists suggests the actual basis of the modern perspective on addiction and freewill is not exclusively the product of what we have learned through scientific advance. What led Rush, without the benefit of addiction science, to his generally modern position? He explained his position regarding the compromised freewill of the individual this way, "That this is the case, I infer from persons who are inordinately devoted to the use of ardent spirits being irreclaimable, by all the considerations which domestic obligations, friendship, reputation, property, and sometimes even by those which religion and the love of life, can suggest to them." (Rush, 1812, p. 264) In other words, the fact that the considerations that normally move people no longer do is evidence that will is compromised. This is, broadly speaking, both a good fit with compatibilism (especially as formulated by Dennett, 1987), as well as with the criteria for diagnosing substance use disorder (American Psychiatric Association, 2013). We take this to be an important point—one driver of conceived impairment of freewill is the degree to which the individual's motivation diverges

from those that are typical and that are healthy, even in the absence of any scientific explanation of mechanism. Even someone who knows very little about addiction and its causes may think, "Something abnormal is at work here—this is not behavior any sane person would choose" and so infer a loss of freewill. Science may significantly refine the picture, or provide support or refutation for ideas about addiction and compromised freewill, but it certainly not the unique basis for them. This dependence of judgments of freewill on background conceptions of what is "normal" or "rational" may have informed Foucault's (1973; 1973; 1978) position regarding the dependence of ideas about freewill on background conceptions on Enlightenment rationality.

5.2 | Addiction Science and libertarian vs. compatibilist freewill

Before we examine the likely impact of *specific* contemporary addiction science theories of addiction on conceived freewill, it is useful to consider in general terms, what sort of impact is plausible. We noted above that partial/probabilistic theories of behavior do not conflict with libertarian freewill, since the observer can suppose that freewill is at least part of the reason theories are only probabilistic. So, there is a *prima facie* case for hypothesizing that addiction science would not impact conceived libertarian freewill. On the other hand, as we also discussed above, there may be some tendency for observers to treat even probabilistic biological antecedents as though they were deterministic (material-framework priming). We suspect an effect like this would be limited in scope since the experience of individuals with addiction, and the language used in discussing it (cravings, withdrawal discomfort, feelings of hopelessness, shame, etc.) are within the intentional framework.

With regard to compatibilist freewill, the potential impact of addiction science is more straightforward, and, we suspect, more significant. Addiction science can shed light on the mechanisms that drive behavior, and so holds the potential to impact observer's understanding of how addictive behavior relates to the actor's mental states (the key issue for compatibilist theories generally). Broadly, we think scientific accounts can impact conceived compatibilist freewill by, in different ways, linking addictive behavior to mechanisms that observers view as peripheral to the actor.

We now turn to three important contemporary scientific theories of addiction and consider how each may impact perceived freewill. Because, as discussed above, theories of behavior from the biological sciences appear to have greater impact on conceptions of freewill (Monterosso et al., 2005) we focus first on three leading proposed mechanisms: 1) incentive-salience caused by dopaminergic signaling in the ventral striatum, and which cause intense spikes in motivation, 2) impaired executive function, caused by both structural and functional impairment in the prefrontal cortex, and which are thought to undermine capacity to pursue long-term interest, and 3) habit-control, linked to dorsolateral striatal circuit that drives automatic behavior, potentially leading to action that is selected independent of the actor's preferences.

5.3 | Craving/ Incentive-Salience research can suggest addiction motivation overpowers freewill

An important early breakthrough in the neuroscience of motivation came in the 1950s through the work of James Olds and Peter Milner. Neural activity is electrochemical, and sending a

small current through a population of neurons causes those neurons to fire repeatedly, releasing their neurotransmitter payloads *en masse*. Olds and Milner placed electrodes in different locations of the rat brain, and they yoked the initiation of current to a lever the rat could press. They were looking for (and eventually found) locations in which the rat would lever press in order to receive the electrical stimulation (Olds, 1956; Olds & Milner, 1954). Effective brain areas included portions of the midbrain and the ventral striatum. When the electrodes were implanted in these regions, rats lever pressed without satiation, often forgoing all else and sometimes dying. The work was a major step forward in the neuroscience of incentive, and of addiction.

Electrical current is of course a blunt manipulation. Subsequent work exploited different techniques to record from and to stimulate individual neurons, and to measure and manipulate specific neurotransmitters. The picture that emerged linked strong desire states to the release of dopamine within subcortical regions including the ventral striatum. The role of striatal dopamine in motivation is evident from manipulation and observational studies. When dopamine neurons are selectively destroyed experimentally, the result is an animal with very little interest in anything (Moy, 1995). It looks lazy in the extreme, engaging in less exploration, mating, eating and fighting (Berridge, Venier, & Robinson, 1989). Conversely when dopamine activity is raised through, for example, infusion of stimulants, animals evidence heightened sensitivity to other reinforcers including brain stimulation (Maldonado-Irizarry, Stellar, & Kelley, 1994). In cell recording of midbrain neurons that release dopamine into the ventral striatum, firing rates elevate dramatically when the animal is pursuing or anticipating a valued reward (Schultz, 2002).

Although many of the techniques used in rodent experiments are not applicable to human work, advances in neuroimaging do allow observation of neural correlates of craving for drugs, as well as other rewards like food and sex. In a typical experiment, the participant looks at cues such as images of injecting heroin (Langleben et al., 2008; Li et al., 2012) and reports craving. Simultaneously, changes in the activity of neurons throughout their brain are indirectly measured through a growing list of imaging methodologies that includes functional Magnetic Resonance Imaging (fMRI) which is indirectly sensitive to general neuronal activity and Positron Emission Tomography (PET) which can be sensitive to the activity of particular neurotransmitters. The results of human imaging work, typically visualized as color 'brain maps' of craving, are broadly consistent with the animal work. Subcortical structures implicated in incentive signaling in rat work are reliably activated by drug, food, and sex cues (Noori, Cosa Linan, & Spanagel, 2016). Moreover, activity in the nucleus accumbens in response to cocaine cues predicts relapse among individuals in treatment for cocaine substance use disorder (MacNiven et al., 2018).

What does this say about the question on addiction and freewill? On its face, identifying the neural substrates of desire might seem irrelevant to the question of whether addiction compromises freewill. Don't we already think the brain is the substrate of craving? Why should localizing craving to something more specific than "the brain" make any difference? We consider two paths whereby it might: 1) it might inform views on the potency of motivation driving drug use, and 2) it might prime a non-intentional framework conception on drug use.

Regarding the "potency of motivation" path to perceived loss of freewill, intuitively, behavior that runs counter to an actor's previous plans (e.g., she said she was going to quit) and other values, and that could be linked to a strong transient craving may seem less free. The observer may view the incentive state as something that overwhelms and takes over. Of course, the conceived strength of the incentive is critical. Just how strong is it? Is it similar to the urge

to, say, break one's diet with late night ice cream? Or is it something far more extreme? Is it, as one clinician put it, "[akin to] the strongest drive which is the drive to breath" (Sucher, 2013)? If the latter is a fair characterization, she may be seen as having been "overcome" by the craving (Loewenstein, 1996). That is, the craving may be seen as a force that temporarily takes over—a Dr. Jekyll to Mr. Hyde transformation. Such a view comports well with Aristotle's idea of *akrasia*, or weakness of will, as well as Mischel's classic demonstrations of childrens' failed struggles to resist eating one marshmallow with the promise that if they do, they will earn a second one (e.g., Mischel, Shoda, & Rodriguez, 1989). So, motives may be strong, but how can we know how strong addictive motivations are?

One way to assess how strong addictive motivations are is to ask people with direct experience. However, a downside of this direct approach is that the observer may be skeptical of these self-reports. Maybe the addicted or formerly addicted individual is motivated to exaggerate the strength of the motivation to herself or to others in order to mitigate reproach. Or maybe the experience is so state dependent that she *under*estimates it when outside the 'heat of the moment' (Loewenstein, 1996). An alternative means of gauging strength of desire is by inference from "revealed preference." What price does the addict pay, not just in money, but in the broader sense of what is traded away in addiction? If a mother relapses even though she is being monitored by child protective services, doesn't it suggest the motivation to use was very high? If the 20-year old is willing to inject with a used needle, risking contracting a life-compromising disease, can't we conclude that the craving was extremely powerful? The primary problem with this approach is that revealed preference relates to the ordering of motivations, rather than to their magnitude in an absolute sense. Inferring the magnitude of craving to be high might fit the destructive behavior of the addict, but an alternative is to infer that the individual lacks the normal amount of all other motivations, or the normal amount of "willpower." We think this kind of inference is common, as when the observer makes sense of the addicted individual's destructive behavior by attributing it to a kind of extreme selfishness or global weakness. Indeed, the phrase "addicts are selfish" returns 738,000 hits on Google, while the phrase "addicts are not selfish" returns six. And the phrase "addicts have no willpower" returns 693,000 hits, while "addicts have good willpower" or simply, "addicts have willpower" both return no hits. Maybe craving is not extraordinarily strong, but rather the addicted individual is extraordinarily weak.

It is possible that delineation of neural substrates could provide objective evidence that bears on the question, "How strong is drug craving?" If the neural mechanisms that correspond to incentive are fully delineated, then it might be possible to objectively gauge the general strength of motivation associated with addiction. This might even provide a basis for addressing whether a behavior problem is a true addiction (e.g., is the incentive to use social media ever comparable in magnitude to the incentive to use heroin?) However, there are many practical challenges to this. Can a lab model elicit the incentive experienced in the real world by people attempting to quit? Is an experimental setting, in which a participant knows she is in a fishbowl, inherently different in ways that limit its value as a model? We know of no significant effort to overcome such challenges. Nevertheless, observers may infer something about the strength of drug-seeking motivation from the neuroscience, warranted or not. They may, for instance, see the colored images from fMRI studies of drug craving and interpret them as evidence that the strength of the state must be extraordinary to be visible. And that may in turn add to the evidence that the addicted individual has lost her freewill. Her actions may appear more like the product of an overpowering motivation, a psychological *force majeure,* than the product of her will. Importantly, this inference of a potent transient motivation would still maintain the

intentional framework (the individual overcome with craving wants to get high and acts accordingly). But it may be seen as *not really her* – a kind of **motivation modularity** in which control is temporarily hijacked by one motivation to the exclusion of others. If the individual experiencing the craving state did manage to resist it, we expect the observer would view it as laudable, rather than inexplicable. This suggests that it is the conception of compatibilist freewill that is impacted by this perspective, and not the conception of libertarian freewill.

Alternatively, even though incentive is inherently an intentional framework construct, we think it is conceivable that illumination of its neural mechanisms could prime observers to think of behavior in a way that bypasses intention. Perhaps the delineation and discussion of physical activity in the brain as the driving force invites observers to imagine a clockwork deterministic control—the material-framework priming effect noted above. If observers are naïve dualists, as seems to be commonly the case (Bering, 2006; Bloom, 2004; Monterosso et al., 2005), then linking behavior to something specific about the brain may invite adoption of a material framework when understanding the behavior. And since libertarian freewill does not apply to the material world, this pathway would undermine both the compatibilist and libertarian senses of freewill. It should be noted though, to the extent to which this really is how observers respond to incentive-salience neuroscience, it is likely similarly true of any neural mechanism associated with drug use.

5.4 | Dysfunction in circuitry critical for decision-making can suggest weakened freewill

There is an extensive literature suggesting neurocognitive deficits in "executive function" which broadly refer to the capacities that support controlled goal directed action, and which are highly sensitive to dysfunction in the prefrontal cortex. For example, in non-human primate models, chronic stimulant exposure can induce deficits in basic capacities such as adjusting behavior to meet the demands of changes in the environment (Jentsch & Taylor, 1999). In humans, chronic methamphetamine use is associated with low efficiency in inhibiting inappropriate responses (Monterosso et al., 2005; Salo et al., 2002). Large studies report some differences in prefrontal gray matter, with deficits associated with addicted populations (Xiao et al., 2015; Yuan et al., 2010). They also suggest deficits in the white matter tracts that provide the pathway whereby the prefrontal cortex exerts influence over subcortical signaling (Hampton, Hanik, & Olson, 2019). Other work suggests relevant neurotransmitter deficits in addicted populations, including a deficiency in dopamine projecting to the prefrontal cortex, which supports executive function (Koob & Volkow, 2010; Martinez et al., 2012; Volkow, Fowler, Wang, Baler, & Telang, 2009).

Noting findings like those above, in her aforementioned address arguing that addiction is "a disease of freewill," NIDA Director Nora Volkow asserted that the prefrontal cortex is critical in, "our ability to change our behavior when the environment changes so that we can optimize our actions. To sustain effort. To resist immediate rewards. To delay gratification. To be able to consider a goal for our future, and to carry it through. And drugs disrupt that" (Volkow, 2015). The focus on ability is important. It implies the addicted individual may have motivations and intentions to quit that go unrealized because of a deficiency. This interpretation is relevant to compatibilist freewill since it implies the actor's behavior is not in line with her desires. The freewill she lacks turns on how her motivations are understood rather than on the metaphysical question of whether she could have done otherwise, or was a "mover-unmoved." However, as with the neuroscience on incentive salience, it is possible that the neuroscience of executive

dysfunction could lead to material-framework priming, and thereby impact libertarian freewill, though for reasons we discuss below, we think any such impact would be quite limited.

As an aside, while we think deficits in executive function associated with addiction may be clinically significant (Dean, Kohno, Morales, Ghahremani, & London, 2015), they are relatively modest (Dean, Groman, Morales, & London, 2013; Martinez et al., 2012)—more modest than the impairment of normal aging (Klimova, Valis, & Kuca, 2017). And of course, anecdotally, there is no shortage of "high functioning addicts" who maintain exemplary behavior in other domains of life. Nevertheless, the findings of neurocognitive deficits in executive function may impact how people view addiction.

5.5 | Control by habit circuit can suggest involuntary drug use

Another prominent theory relates addiction to the neural processes that support habits. In operant conditioning, a behavior is made more likely if it is followed by a reinforcer. Relatively early in shaping a behavior, it can be shown that animals engage in the operant based on valuation of the reinforcer. Suppose that after, say, twenty trials in which a lever press was reinforced with a particular food, the animal is removed from the cage and given the same food paired with something to make it sick. The animal will develop a taste aversion to that food, so that it will not consume it in the future. What if that animal is then returned then to the cage in which a lever press yields that food -- will it press the lever? Generally, given the relatively short training history of twenty reinforcement trials, it will not. This is revealing as it implies that the animal has a representation of the food reward, and that its valuation of the reward determines whether or not it presses the lever (goal-directed responses). However, what if instead of twenty learning trials, the animal has experienced 2000 learning trials, followed by the same "devaluation" procedure. Now, the animal, when returned to the operant cage will typically press the lever that yields the food it no longer wants. Something evidently changes with excessive learning such that control of the behavior is no longer linked to valuation of its consequences (Dickinson, 1985). This is how "habit" is operationalized in the behavioral neuroscience literature (Robbins & Costa, 2017). Dramatic progress has been made in delineating and dissociating the neural circuits that control goal-directed vs. habit-driven responses (Belin, Jonkman, Dickinson, Robbins, & Everitt, 2009). In general terms, the ventral-striatal – cortical loops that control behavior early in learning (when behavior is goal directed) give way to control by dorsolateral striatal control as behavior becomes habitual. This shift results in faster and more efficient action control, but at the expense of flexibility to changes in conditions, including goal devaluation.

Addiction neuroscientists have, naturally, wondered if something like habit overcontrol could explain why people with addictions are prone to continuing drug use even after they seem to want to quit. Perhaps the addicted individual no longer wants to engage in the behavior, but despite that, continues to do so because of a habit response that is too powerful to override. This is broadly the argument made by Everitt and Robins: "habit learning occurs in parallel with instrumental action-outcome learning but, with extended training, eventually dominates behavioral output. Crucial to drug addiction is the persisting quality of these habits, which has been likened to the subjective state of 'wanting', but which we would suggest corresponds more obviously to the subjective state of 'must do!'—although this subjective response could arise post hoc as a rationalization of the 'out-of control' habitual behavior rather than being the driving influence" (Everitt & Robbins, 2005, p. 1485). The suggestion here is that the subjective

reports of craving that suggest goal directed action (like the rat pressing the lever for a food because it values it) are actually misleading, as the actual behavior is under habit control (like the rat pressing a lever for food it does not value).

For the above explanation to make sense, it has to be the case that the habit is so strong that it resists being overridden by countervailing goals. It is habit control that leads one, when distracted, to make a wrong turn that is consistent with a frequent driving path, but the moment the goal is brought to attention, the habit yields to the goal-directed action. If habits are to work as an explanation for addiction, they would obviously have to be able to at least sometimes overpower goals. Whether this is plausible or not, if convincing to observers, the habit explanation makes addictive behavior non-intentional (outside intentional framework explanation). Thus, habit explanation could impact both the conceived libertarian and compatibilist freewill, since understanding the behavior through the intentional framework is a necessary condition for each. However, a more plausible conception, given that cessation is often eventually successful, would be to view goal-directed action as still potentially able to win out, in which case, as with the previous theories, the impact would primarily be limited to the compatibilist sense of freewill. The behavior is counter to what the individual wants (so not free in the compatibilist sense). But given the indeterminacy of the behavior, the force-of-habit account poses no direct challenge to conceived libertarian freewill. However, as was also true of other neuroscience theories, it is possible that the identification of relevant habit neural circuitry could prime a material framework that is incompatible with libertarian freewill.

5.6 | Limitations on science's impact on conceptions of addictions

Across the theories considered, their impact on conceived freewill is likely larger for the compatibilist sense of freedom than for the libertarian sense. In everyday person perception, we expect the two senses are not sharply distinguished and instead may blend in observer's thinking and language. Moreover, we think the impact of addiction science on how observers view the issue of even compatibilist freewill is likely to be modest. As noted above, some observers take the suffering and self-destructive behavior of the addicted individual as evidence of compromised freewill, even absent any understanding of it. For such an observer a scientific explanation merely fills in details, it does not change the basic conclusion that the addicted individual's will is compromised. And for individual's less sympathetic, the implications of the science are not so clear as to compel a particular conclusion regarding freewill. We don't think it is plausible that an individual who holds a strong view on the issue will dramatically change that view based on encountering a particular scientific explanation of addiction. To the extent the science of addiction has an immediate impact, it is likely limited to those whose inclination is neither strongly sympathetic nor unsympathetic. And perhaps more importantly, scientific theories of addiction may exert a slow but continuous influence that, over time, shifts the way individuals with addiction are viewed.

ORCID

John R. Monterosso https://orcid.org/0000-0002-5613-3182

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Association.

Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W., & Everitt, B. J. (2009). Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. *Behavioural Brain Research*, *199*(1), 89–102.

Bering, J. M. (2006). The folk psychology of souls. *Behavioral and Brain Sciences*, *29*(5), 453–462.

Berridge, K. C., Venier, I. L., & Robinson, T. E. (1989). Taste reactivity analysis of 6-hydroxydopamine-induced aphagia: Implications for arousal and anhedonia hypotheses of dopamine function. *Behavioral Neuroscience*, *103*(1), 36.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Borges, J. L. (1962). *The garden of forking paths*. New York: Ficciones Grove Press.

Dean, A. C., Groman, S. M., Morales, A. M., & London, E. D. (2013). An evaluation of the evidence that methamphetamine abuse causes cognitive decline in humans. *Neuropsychopharmacology*, *38*(2), 259–274.

Dean, A. C., Kohno, M., Morales, A. M., Ghahremani, D. G., & London, E. D. (2015). Denial in methamphetamine users: Association with cognition and functional connectivity in brain. *Drug and Alcohol Dependence*, *151*, 84–91.

Deery, O., Davis, T., & Carey, J. (2015). The Free-Will Intuitions Scale and the question of natural compatibilism. *Philosophical Psychology*, *28*(6), 776–801.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Dennett, D. C. (1993). *Consciousness explained*, Boston. MA: Brown: Little.

Dennett, D. C. (1987). The Intentional Stance. Cambridge, MA: MIT Press.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1135), 67–78.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*, *18*(1), 342–350.

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language*, *24*(1), 1–23.

Ferentzy, R. (2002). Foucault and addiction. *Telos*, 167–191.

Foucault, M. (1973). *Madness and Civilization: A History of Madness in the Age of Reason*. New York: Vintage.

Foucault, M. (1973). *The Order oft Things: An Archeology of the Human Sciences*, New York: Vintage.

Foucault, M. (1978). The History of Sexuality. *An introduction* (Vol. 1), New York: Vintage.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, *68*(1), 5–20.

Gazzaniga, M. S. (1967). The split brain in man. *Scientific American*, *217*(2), 24–29.

Gazzaniga, M. S., & LeDoux, J. E. (1978). *The integrated mind*. New York: Plenum Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814.

Haidt, J. (2006). *The happiness hypothesis: Finding modern truth in ancient wisdom*. New York: Basic books.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66.

Hampton, W. H., Hanik, I. M., & Olson, I. R. (2019). Substance abuse and white matter: Findings, limitations, and future of diffusion tensor imaging research. *Drug and Alcohol Dependence*, *197*, 288–298.

Heim, D. (2014). Addiction: Not just brain malfunction. *Nature*, *507*(7490), 40.

Heyman, G. M. (2009). *Addiction: A disorder of choice*. Cambridge, MA: Harvard University Press.

Hobbes, T. (1968). *Leviathan*. Baltimore: Penguin Books.

Hoefer, C. (2016). Causal determinism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (spring 2016). Stanford, CA: Metaphysics Research Lab, Stanford University.

Horgan, T., & Timmons, M. (2011). Introspection and the phenomenology of free will: Problems and prospects. *Journal of Consciousness Studies*, *18*(1), 180–205.

Howe, A. S., Buttenschøn, H. N., Bani-Fatemi, A., Maron, E., Otowa, T., Erhardt, A., ... De Luca, V. (2016). Candidate genes in panic disorder: Meta-analyses of 23 common variants in major anxiogenic pathways. *Molecular Psychiatry*, *21*(5), 665–679.

Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug abuse: Implications for the control of behavior by reward-related stimuli. *Psychopharmacology*, *146*(4), 373–390.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Klimova, B., Valis, M., & Kuca, K. (2017). Cognitive decline in normal aging and its prevention: A review on non-pharmacological lifestyle strategies. *Clinical Interventions in Aging*, *12*, 903–910.

Koob, G. F., & Volkow, N. D. (2010). Neurocircuitry of addiction. *Neuropsychopharmacology*, *35*(1), 217–238.

Langleben, D. D., Ruparel, K., Elman, I., Busch-Winokur, S., Pratiwadi, R., Loughead, J., ... Childress, A. R. (2008). Acute effect of methadone maintenance dose on brain fMRI response to heroin-related cues. *American Journal of Psychiatry*, *165*(3), 390–394.

Laplace, P. S. (1814). *A philosophical essay on probabilities*. New York: Dover Publications.

Lawrence, R. E., Rasinski, K. A., Yoon, J. D., & Curlin, F. A. (2013). Physicians' beliefs about the nature of addiction: A survey of primary care physicians and psychiatrists. *The American Journal on Addictions*, *22*(3), 255–260.

Leshner, A. (1997). Addiction is a brain disease, and it matters. *Science*, *278*(5335), 45–47.

Lewis, M. (2017). Addiction and the brain: Development, not disease. *Neuroethics*, *10*(1), 7–18.

Li, Q., Wang, Y., Zhang, Y., Li, W., Yang, W., Zhu, J., ... Tian, J. (2012). Craving correlates with mesolimbic responses to heroin-related cues in short-term abstinence from heroin: An event-related fMRI study. *Brain Research*, *1469*, 63–72.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

MacNiven, K. H., Jensen, E. L. S., Borg, N., Padula, C. B., Humphreys, K., & Knutson, B. (2018). Association of neural responses to drug cues with subsequent relapse to stimulant use. *JAMA Network Open*, *1*(8), e186466.

Maldonado-Irizarry, C. S., Stellar, J. R., & Kelley, A. E. (1994). Effects of cocaine and GBR-12909 on brain stimulation reward. *Pharmacology Biochemistry and Behavior*, *48*(4), 915–920.

Martinez, D., Saccone, P. A., Liu, F., Slifstein, M., Orlowska, D., Grassetti, A., ... Comer, S. D. (2012). Deficits in dopamine D2 receptors and presynaptic dopamine in heroin dependence: Commonalities and differences with other types of addiction. *Biological Psychiatry*, *71*(3), 192–198.

McKenna, M. S. (2012). Contemporary compatibilism: Mesh theories and reasons-responsive theories. In R. Kane (Ed.), *The Oxford handbook of free will* (2nd ed.). Oxford: Oxford University Press.

McQuain, J., Malless, S., & Shakespeare, W. (1998). *Coined by Shakespeare: Words and meanings first used by the bard*. Springfield, MA: Merriam Webster.

Mischel, W., Shoda, Y., & Rodriguzez, M. L. (1989). Delay of gratification in children. *Science*, *244*, 933–938.

Monterosso, J. R., Aron, A. R., Cordova, X., Xu, J., & London, E. D. (2005). Deficits in response inhibition associated with chronic methamphetamine abuse. *Drug and Alcohol Dependence*, *79*(2), 273–277.

Monterosso, J., Royzman, E. B., & Schwartz, B. (2005). Explaining away responsibility: Effects of scientific explanation on perceived culpability. *Ethics & Behavior*, *15*(2), 139–158.

Moy, S. S. (1995). Impaired acquisition and operant responding after neonatal dopamine depletion in rats. *Pharmacology, Biochemistry and Behavior*, *52*(2), 433–441.

Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford Handbook on Free Will*. (2nd ed., pp. 555–575). New York, NY: Oxford University Press.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive?. *Philosophy and Phenomenological Research*, *73*(1), 28–53.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

Noori, H. R., Cosa Linan, A., & Spanagel, R. (2016). Largely overlapping neuronal substrates of reactivity to drug, gambling, food and sexual cues: A comprehensive meta-analysis. *European Neuropsychopharmacology*, *26*(9), 1419–1430.

Olds, J. (1956). Pleasure centers in the brain. *Scientific American*, *195*(4), 105–117.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, *47*(6), 419.

Pavlov, I. P. (1927). In G. V. Anrep (Ed.), *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*, London, England: Oxford University Press.

Peele, S. (1989). *Diseasing of America: Addiction treatment out of control*. Lexington, MA: Lexington Books.

Robbins, T. W., & Costa, R. M. (2017). Habits. *Current Biology*, *27*(22), R1200–R1206.
Rose, N. (1996). *Inventing Our Selves: Psychology, Power, and Personhood*. New York: Cambridge University Press.
Rush, B. (1784). *An Inquiry into the effects of Spirituous Liquors upon the Human Body, and their influence upon the happiness of society*. Philadelphia: Thomas Bradford in Front-Street.
Rush, B. (1812). *Medical inquiries and observations, upon the diseases of the mind*. Philadelphia: Kimber & Richardson.
Salo, R., Nordahl, T. E., Possin, K., Leamon, M., Gibson, D. R., Galloway, G. P., ... Sullivan, E. V. (2002). Preliminary evidence of reduced cognitive inhibition in methamphetamine-dependent individuals. *Psychiatry Research*, *111*(1), 65–74.
Samuelsson, M., Jokinen, J., Nordstrom, A.-L., & Nordstrom, P. (2006). CSF 5-HIAA, suicide intent and hopelessness in the prediction of early suicide in male high-risk suicide attempters. *Acta Psychiatrica Scandinavica*, *113*(1), 44–47.
Savani, K., Markus, H. R., Naidu, N. V. R., Kumar, S., & Neha, B. (2010). What counts as a choice? : U.S. Americans are more likely than Indians to construe actions as choices. *Psychological Science*, *21*, 391–398.
Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241–263.
Schwartz, B. (1978). *Psychology of Learning and Behavior*. New York: Norton.
Sheffield, F. (1965). Relation between classical conditioning and instrumental learning. In W. F. Prokasy (Ed.), *Classical Conditioning* (pp. 302–322). New York, NY: Appleton-Century-Crofts.
Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, *24*, 86–97.
Skinner, B. F. (1971). *Beyond Freedom and Dignity*. New York, NY: Knopf/Random House.
Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *75*(1/2), 5–24.
Strawson, P. F. (1974). *Freedom and resentment, and other essays*. London: Egmont Books.
Sucher, M. (2013). *Addiction is A brain disease—YouTube*. Retrieved from https://www.youtube.com/watch?v=61bPxV-y1qY
Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, *93*(2), 175–179.
Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals*. New York: Macmillan. https://www.biodiversitylibrary.org/bibliography/25848
Vohs, K. D., & Baumeister, R. F. (2009). Addiction and free will. *Addiction Research and Theory*, *17*(3), 231–235.
Volkow, N. D., Fowler, J. S., Wang, G. J., Baler, R., & Telang, F. (2009). Imaging dopamine's role in drug abuse and addiction. *Neuropharmacology*, *56*(Suppl 1), 3–8.
Volkow, N. D. (2015). *Addiction is a disease of free will*. Retrieved from https://archives.drugabuse.gov/about-nida/noras-blog/2015/06/addiction-disease-free-will
Williams, D. R., & Williams, H. (1969). Auto-Maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior*, *12*(4), 511–520.
Wood, A. à. (1796). The history and antiquities of the university of Oxford. In *Two books*. Oxford: Oxford University Press.
Xiao, P., Dai, Z., Zhong, J., Zhu, Y., Shi, H., & Pan, P. (2015). Regional gray matter deficits in alcohol dependence: A meta-analysis of voxel-based morphometry studies. *Drug and Alcohol Dependence*, *153*, 22–28.
Yuan, K., Qin, W., Dong, M., Liu, J., Sun, J., Liu, P., ... Tian, J. (2010). Gray matter deficits and resting-state abnormalities in abstinent heroin-dependent individuals. *Neuroscience Letters*, *482*(2), 101–105.

How to cite this article: Monterosso JR, Schwartz B. Addiction science and the perception of freewill. *J Theory Soc Behav*. 2020;50:373–390. https://doi.org/10.1111/jtsb.12250