The Addict in Us All

Brendan Dill and Richard Holton

Abstract: In this paper, we contend that the psychology of addiction is similar to the psychology of ordinary, non-addictive temptation in important respects, and explore the ways in which these parallels can illuminate both addiction and ordinary action. The incentive salience account of addiction proposed by Robinson and Berridge (1993; 2001; 2008) entails that addictive desires are not in their nature different from many of the desires had by non-addicts; what is different is rather the way that addictive desires are acquired, which in turn affects their strength. We examine these 'incentive salience' desires, both in addicts and non-addicts, contrasting them with more cognitive desires. On this account the self-control challenge faced by addicted agents is not different in kind from that faced by non-addicted agents—though the two may, of course, differ greatly in degree of difficulty. We explore a general model of self-control for both the addict and the non-addict, stressing that self-control may be employed at three different stages, and examining the ways that it might be strengthened. This helps elucidate a general model of intentional action.

1. Introduction

On a common conception, addicts and non-addicts are very different. Addicts' compulsions drive them to act in ways that are quite foreign to the non-addicted. They consume drugs in the full knowledge that they are harmful, and in the face of a desire to stop, something that the normal agent does not do.

We argue here that this picture is quite misleading. Non-addicts, like addicts, have to contend with desires that are quite insensitive to their reflective judgments about what is good. And addicts, like non-addicts, have at their disposal a capacity for self-control that can enable them to resist and overcome these desires.

The situation faced by the addicted agent is thus parallel to that faced by the non-addicted agent. It is an extreme example of the same kind of thing. Both will have desires that persist even in the belief that their objects are worthless, or even actively harmful. And so both will be faced with the self-control problem of resisting these troublesome desires in the light of these beliefs. This self-control challenge, faced by both addicted and non-addicted agents, is the focus of this paper.

We begin by briefly outlining the empirical support for our first claim, that addictive desires are instances of a kind of desire common to all agents (§2). They result from a system—the 'incentive salience' system—that has evolved to create desires, for foods and other things, that are independent from the agent's evaluations of the worth of those things. What is different in the addict is not the intrinsic nature of these desires, but their origin. Addictive drugs cause the desire-formation process to malfunction, with the result that they come to be desired with an intensity and permanence that is quite out of proportion to any pleasure they have given. However, the same problematic features of addictive desires arise, albeit less intensely, even

when the incentive salience system does not malfunction. We see this in more mundane desires such as the craving for chocolate. We characterize the common features of these 'incentive salience' desires, and contrast them with the more reasons-sensitive desires, which we call 'cognitive desires', on the basis of which agents reflectively deliberate about what to do. The competition between these two kinds of desire for control over behavior poses the problem with which we are concerned throughout the remainder of the paper: the problem of self-control.

We begin our discussion of self-control by arguing that an agent's course of action is not solely determined by the relative strengths of her desires; it also matters whether, and how, she exerts *self-control* on behalf of some desires over others. Our argument centers on two subject populations whose behaviors are, we think, best explained as resulting from selective deficits in self-control capacity: subjects with lesions in the *ventromedial prefrontal cortex* (vmPFC), and subjects experiencing *ego depletion* (§3).

The picture that emerges from these first two sections portrays intentional action as the result of a competition between two systems: the incentive salience system, which automatically guides behavior on the basis of appetitive desires, and the self-control system, by means of which an agent can, with effort, bring her actions in accordance with her more reflective desires. Though the conflict between these systems is more dramatic in addicts, it pervades ordinary action as well.

Though we offer some new arguments in its support, this two-systems picture is far from novel. The basic outlines of the approach date back to Plato (1997, *Republic* Book IV) and the more contemporary version of this picture we present here has been defended before (Watson 1975; Holton 2009; Sripada 2014). What we hope to add to this literature is a more detailed picture of how these two systems interact to produce behavior (§4). We propose that there are three distinct loci of self-control conflict—at the point of deliberation, of formation of intention, and of execution of action—which we call the *deliberative, volitional,* and *implemental* stages of self-control. Distinguishing between these stages brings into focus the nature of the self-control challenge faced by addicts and non-addicted agents alike. Drawing on a large body of empirical work, we articulate the nature of the conflict between the self-control and incentive salience systems at each stage, and suggest ways in which each kind of self-control might be improved. What emerges is a single model of human motivational psychology that explains the predicaments of addiction and ordinary temptation with equal aptitude.

2. Desire

Let us start with the question of how we form desires. One might think—many have thought—that we are hedonists at heart. On such a view all of our desires stem from a fundamental intrinsic desire for pleasure. When we desire things other than pleasure we desire them *instrumentally*: that is, we desire them derivatively, because we believe that they will give us pleasure.

Many have objected that such an account makes us seem far too selfish: sometimes we want things because of the benefits that they will bring other people, independently of any

benefits they may bring to us. We think that this point is probably right,[1] but it is not our primary concern here. Our argument is rather that such a picture is wrong even when we consider such simple self-regarding desires as those we have for different foods. Suppose that an agent were to sample many different foods. Some they would like, others not, and they would then go on to regulate their future desires for them accordingly. We might expect these to be instrumental desires, formed in the service of the desire for pleasure. But the empirical evidence suggests not. It suggests instead that pleasure typically causes us to have intrinsic desires for the foods themselves, which then motivate independently of any beliefs about the pleasure that such foods will bring.

The crucial evidence for this is that our desires for different foods are not always directly responsive to our explicit beliefs about how pleasurable they are to eat. The desires do not need such beliefs to bring them into existence; and they can persist in their absence. We sometimes get a sense of this in our direct experience—many of us experience a desire to eat more of a thing (chocolates? over-rich desserts? peanuts? potato chips?) even when we know that we won't enjoy it and that it may leave us feeling somewhat nauseated. However, the best evidence for this phenomenon comes from studies, not of normal foods, but of addictive drugs, and moreover, of how they work on rats. So let us start there, and then return to the case of how more normal foods work on us. Our account will follow the 'incentive salience' theory developed by Robinson and Berridge (1993; 2001; 2008).[2]

Addictive drugs artificially increase the levels of the neurotransmitter dopamine in the brain. Different drugs do this in different ways: nicotine stimulates the production of dopamine directly, opiates decrease the production of substances that inhibit the production of dopamine, cocaine reduces the activity of the system that reabsorbs dopamine after it has been released, and so on (Holton and Berridge 2013: 245-246). What is remarkable is that these various substances with otherwise disparate biological and neurological effects have this single common feature: they all boost dopamine.

It is reasonable to infer that this shared neurobiological quirk must play a role in explaining these substances' more obvious common feature: that they all cause addiction. This idea is borne out by the evidence. By boosting dopamine levels, addictive drugs artificially stimulate the mesolimbic dopamine system, which has long been known to play an important role in motivation. That is, they stimulate it directly, and not in the normal way via experience. (Compare getting someone to see stars by banging them on the head, rather than by showing them stars.) So to understand how drug addiction works, we need to understand what role dopamine plays in motivation.

For many years dopamine was thought to be a pleasure signal. But it isn't. Whilst it is typically accompanied by pleasure, that isn't what it is causing or registering (for a detailed defense of this claim, see Berridge 2007). Separate the indicators of a rat's pleasure (its facial movements) from the indicators of its desire (the effort it will expend to attain the thing), and you find that dopamine is linked to desire and not to pleasure. Artificially increase a rat's

[1] See Batson and Shaw (1991) for a classic empirical response.

[2] The particular interpretation here follows that given in Holton and Berridge (2013); readers should look there for much more detail on what is here treated far too swiftly.

dopamine levels by giving it amphetamines, and it will work much harder to get something even if that thing gives it no pleasure, and it knows it (Wyvell & Berridge 2000). Reduce the rat's dopamine levels via genetic modification and it will fail to work for a thing even if that thing will give it great pleasure, and it knows it (Robinson, Sandstrom, Denenberg, & Palmiter 2005). Moreover—and this is crucial given the implications for addiction—if you increase the dopamine levels when a rat is sampling a foodstuff, what you bring about is not just an immediate desire for that foodstuff, but also a long-term dispositional desire for it (Wyvell & Berridge 2001). Show the rat the foodstuff again later, and it will still want it strongly.

What is happening here? Rats are opportunistic creatures, who need to be able to accommodate their tastes to a new environment. It makes sense for them to be able to regulate their desires in proportion to the pleasure that they get from various foodstuffs. Dopamine is clearly involved in this process. But it looks as though dopamine works directly on desires, without the need for the involvement of pleasure or beliefs about pleasure. It may be that dopamine release is typically *caused* by pleasure: in the case of most non-addictive foodstuffs, the most pleasurable ones will give the greatest dopamine release. But if dopamine is artificially increased, as it is by addictive drugs, then this leads to the production of desire independently of pleasure.

In fact, given what we have said, we need to identify two roles that dopamine plays in the production of desire. One, the *triggering role*, involves the triggering of occurrent desire: dopamine has a role in actually getting the rat to move towards the food in the moment. The other, the *formation role*, involves the formation of dispositional desire: dopamine works to set up a long-term disposition to want the food in the future.[3] Stimulate a rat's dopamine levels at the same time that it is consuming a certain food, and it will form a dispositional desire for that food (Wyvell & Berridge 2001). This is a focused desire: it is focused on the food that was being consumed when the dopamine was released. Present the food again, or present other cues that were associated with it, and the rat will want it, even if its dopamine levels are not then being stimulated. Dopamine thus creates a dispositional desire that, when cued by the relevant food or other associated cues, triggers an occurrent desire for that food.

The formation role that dopamine plays has often been described as a learning role. But that is misleading, since learning is typically taken to involve a change in belief. It is not that the rat comes to believe that the food is going to bring it some advantage, and so forms an instrumental desire conditional on that belief. Rather, what is happening is that an *intrinsic* long-term desire for the substance is being created. If the desire is not reinforced, it will fade in time. But with desires put in place by addictive substances, this can take a very long time indeed—they may last for much of a rat's life.

On the basis of this evidence, Robinson and Berridge (1993; 2001; 2008) posited a motivational system, the 'incentive salience' system, which has the following features. The incentive salience system creates dispositional desires for objects on the basis of those objects' past association with reward. These dispositional desires, which we will call *incentive salience desires*, are activated—become occurrent—when the rat encounters the desired thing, or cues

[3] Holton and Berridge (2013), in an attempt to avoid prejudicing the case, called these 'A-signals' and 'B-signals'. We have replaced this terminology with something a little more memorable.

that have been associated with it. Once an incentive salience desire is active, it leads automatically to behavior in pursuit of the desired object. Crucially, the neural reward signal on the basis of which the incentive salience system acquires its desires is a dopamine signal. Thus addictive substances, by artificially boosting dopamine levels in the brain, produce a disproportionately large reward signal, which in turn causes the formation of a disproportionately strong incentive salience desire for the substance.

We have good reason to think this incentive salience system is present in humans as well as rats. The argument for this claim is an inference to the best explanation: the puzzling features of human addiction are best explained by the hypothesis that addictive desires are incentive salience desires. It explains the craving that is typically prompted by cues associated with the drugs: because of the artificial dopamine boost addictive drugs provide, subjects who consume these drugs acquire a long-term intrinsic desire for them, which is then triggered by the drug-associated cues. This account explains relapse, even after withdrawal: for the dispositional desire remains, ready to be triggered by the relevant cues.[4] Finally, in human subjects, the account explains why the desires for drugs are so horribly independent from beliefs about their worth. For the incentive salience system is working quite independently of belief. The addict can know perfectly well that continued consumption will destroy everything that they hold dear. That does nothing to stop the rush of desire that is triggered by the thought or sight of the drug, or, more broadly, of the people, places or paraphernalia that have surrounded its consumption.

In addiction, the process whereby incentive salience desires are acquired malfunctions. When the system is functioning normally, the dopamine signal is proportional to the reward that the subject is experiencing, and thus the desire it produces is similarly proportionate. When a subject consumes an addictive substance, however, the artificial boost in dopamine that results severs this link between 'wanting' and 'liking,' leading to a desire for the substance that is way out of proportion with the pleasure it brings.

But of course, much of what we have said about the incentive salience system still holds when it is not malfunctioning in this way. When it works well it still lays down long-term dispositional desires for things that have previously given pleasure; and these desires will be triggered by the relevant cues. If the things fail to give pleasure, then in time, the desire will diminish, though it will not evaporate straightaway. And if the thing continues to give pleasure, then the desire will be reinforced, even if the agent comes to believe that it is harmful.

To see this, consider the case of sugar. As far as we know, sugar has no direct effect on the dopamine system: it doesn't imitate dopamine, or inhibit re-uptake, or do any of the things that addictive drugs do. Nevertheless, rats that have been exposed to a sugar solution are strongly motivated to get it, just as they are motivated to get addictive drugs. In fact, if they have a choice between cocaine and sugar, around 90% of rats will take the sugar (Ahmed 2010).[5] It is possible that there is something special about sugar that causes the formation of long-term dispositional

[4] Indeed, withdrawal, horrible though it can be, plays a minor role in addiction; consumption is not primarily motivated by a desire to avoid it.

[5] Similarly, under 10% of human beings who are exposed to most addictive drugs will become addicted. Most people in the West drink alcohol, but most do not become alcoholics. Nicotine is a clear exception, but it is anomalous in many ways: for instance, it does not give rise to euphoria.

desires in this way. But it is equally possible that sugar is simply highly pleasurable.[6] Certainly there is no reason to think that the rats' desire-formation systems are somehow malfunctioning when they develop desires for foods that are rich in it.

Nor is there reason to think that things are any different for human beings. It has become commonplace to speak of sugar addiction; and it is correct that for many people desires for sugar have much in common with addicts' desires for drugs. They too manifest in cravings that are highly cue-dependent, and that persist in the face of the conviction that it would be better to eat less sugar. As with the consumption of sugar, so with many other pleasurable behaviors. Gambling, sex, surfing the web, watching daytime television—all of these have been alleged to give rise to addiction.

But we need to distinguish two things here: the 'hijacking' of the desire-formation process that occurs with addictive drugs; and the nature of incentive salience desires themselves. The first of these features is unique to drugs: only substances that lead to artificial dopamine stimulation will hijack the desire formation process in this way. We have no reason to think that sugar 'addiction' results from a hijacking: there is no evidence that sugar leads to artificial boosts in dopamine. It is even more obvious that web-surfing and gambling do not stimulate dopamine in this way (since they are not *ingested*). So in none of these cases is there reason to think that the dopamine system has malfunctioned. Yet in every case there is reason to think that the motives to engage in these behaviors are insulated from the agent's beliefs about what would be good. Incentive salience desires have this feature regardless of how they are acquired. It is exactly this feature that leads to the talk of addiction, since it is what substance and non-substance 'addictions' have in common. Agents genuinely want to stop; and yet still they feel the pull of the desire.[7]

It is important to realize that there is a contrast here with many of our desires. While incentive salience desires are by nature insensitive to our judgments about what is good, not all desires share this feature. In many cases a desire is bound up with a reason or a justification: to want something is to want it for some reason.[8] As one's confidence in the reason diminishes, so the desire diminishes. Suppose that one of your favorite companies is bringing out a new model of some device that you particularly like; moved by the advance publicity you start to develop a hankering for it. But now the reviews come out, and without fail they are dismissive. The thing is clunky, ill-conceived, badly engineered, a definite step backwards. Your desire withers. You do not need to resist or overcome it. Once your beliefs have changed so that you see no reason to continue, the desire is no longer there. We do not have to think that these reason-based desires are always instrumental, i.e. that we only have them in order to get some other thing. But they

[6] For a review of the evidence that there is more going on in the formations of desires for sugar than simply the activity of the dopamine system see DiLeone, Taylor, & Picciotto (2012) and Ahmed, Guillem, & Vandaele (2013).
[7] We are thus faced with a terminological choice: do we reserve the term 'addiction' for desires formed by means of the dopamine hijacking process, and thus say that sugar and gambling 'addictions' are not addictions proper? Or do we use the term 'addiction' to refer more generally to the predicament an agent faces when she has sufficiently strong incentive salience desires, whatever their origin—and thus say that sugar and gambling addictions can be genuine addictions after all? We are torn on this question: RH is inclined to take the first option; BD leans toward the second.
[8] Such an approach has been advocated, in rather different ways, by Scanlon (1998) and Railton (2012). We agree that some desires have this feature, but deny that this is the only kind of desire.

are bound up with their reasons in such a way that they do not have a life of their own: they cannot live on without them, unlike the incentive salience desires, which can. We will call such desires *cognitive desires*, since they are sensitive to our cognition in a way incentive salience desires are not. Of course agents may simultaneously have both cognitive and incentive salience desires for the same object. Indeed, if the incentive salience system has not been hijacked by addictive substances that is what we would expect.

We should also distinguish incentive salience desires from another class of motivational states, namely habits. These clearly often play a role in addiction: it is not for nothing that we speak of an addict's 'habit'. Like incentive salience desires they are cued by circumstance, and often result in behavior that the agent rejects. Yet in so far as we have a good behavioral grip on them—behaviors like thumb-sucking, nail-biting, hair-pulling and muscle tics—they differ in one crucial respect. The most effective treatment for them is *habit reversal therapy*, which involves monitoring the habit, and then learning an alternative response (Bate, Malouff, Thorsteinsson, & Bhullar 2011).[9] And it seems that the most important part of this is simply the monitoring (Ladouceur 1979; see also Quinn, Pascoe, Wood, & Neal 2010). Habits work automatically, but once they are monitored, the agent can override them. In contrast, while incentive salience desires are sometimes combined with an automatic element (reaching unawares for a cigarette), becoming aware of that element is not enough to remove their force. If they are to be resisted, they need to be overcome.

Let us summarize this section so far. We have contended that there are at least two distinct kinds of desire at play in human motivation. First, there are incentive salience desires, which are formed for objects on the basis of their previous association with either rewarding experience (when the system is functioning well) or artificial dopamine stimulation (when the system is hijacked by addictive drugs). These desires form the motivational basis of addiction, but also play an ever-present role in non-addicted agents' motivation, encompassing at least the sphere of motives we normally call 'appetites' (desires for food, drink, sex, and other pleasurable stimuli). Crucially, incentive salience desires motivate independently from an agent's reflective judgments about what is valuable or even pleasurable. This distinguishes incentive salience desires from a second kind of desires, cognitive desires, which *are* sensitive to and based upon an agent's reflective beliefs about what is valuable; e.g. the desire to read a certain book or pursue a certain career.[10]

How do these two kinds of desire interact to produce intentional action? A simple model, traditional in both psychology and philosophy, sees the efficacy of desires as simply a function of their *strength* (or of their strength together with the subject's belief in how likely they are to be realized). On such a model what an agent does is simply determined by what she most wants to do. Incentive salience desires and cognitive desires will fight it out on the basis of their strength,

[9] It is very effective.

[10] We do not take this distinction to be exhaustive. There could be desires that are not cognitive, in the sense that they are not sensitive to our judgments about reasons, but are not incentive salience desires either, since they are not produced by the incentive salience system. The desires involved in emotional reactions such as fear or guilt, for example, do not seem to fall neatly into either category. We should also point out that dopamine's role is not limited to the incentive salience system. A total lack of dopamine, as one finds genetically engineered dopamine deficient mice, leads to a complete lack of motivation. So dopamine is clearly involved in the triggering of cognitive desires as well as incentive salience desires.

and the stronger desire will control behavior.

There is a great deal of empirical evidence that tells against such a model, evidence that suggests that action is not simply dictated by the strongest desire. In particular, agents are not passive spectators of the competition between their desires for domination over behavior. Rather, the agent herself plays a much more active role in determining which desire triumphs, employing self-control to resist some desires and to act on others. What determines an agent's behavior, then, is not merely how strong her desires are, but also whether and how she exerts self-control.

Self-control is hard work. In the case of addiction self-control is standardly employed to try to restrain incentive salience desires in the light of cognitive desires. Of course this attempt may not succeed. The addict may be aware that she (cognitively) prefers keeping her job to taking drugs, and be aware that taking drugs will cause her to lose her job, on that basis judge that she ought not to use, and yet *still* succumb to her desire for the drug. As R. Jay Wallace puts the point: "even if one succeeds, in the face of [an addictive] desire, in reasoning correctly to the conclusion that it should not be acted on, its continued presence and urgency will make it comparatively difficult to choose to comply with the deliberated verdict one has arrived at" (Wallace 1999: 648). Moreover, even if one chooses to comply, it is hard work to convert that resolution into action.

Our contention here is that these points apply equally to *ordinary* action. For the features of addictive desire that pose self-control problems are features of incentive salience desires in general, and thus are shared by a wide range of non-addictive desires as well. Just as the motivational force of an addict's incentive salience desire for heroin persists despite her judgment that she shouldn't take it, the motivational force of an ordinary agent's incentive salience desire for a cake will persist despite her judgment that she ought to have something more healthy instead. Whether the agent's judgment or craving prevails is a matter of self-control.

We have already elucidated the essential features of the incentive salience system, and presented empirical evidence for its existence. However, we have so far said little about the nature of self-control, and have given no empirical argument for the existence of this phenomenon. We now turn to this task (§3). Then we will be in a position to see how the different kinds of desires are mediated by the self-control system to produce intentional action (§4).

3. The reality of self-control

There are various reasons for believing in the existence of self-control as an independent system that is not reducible to strength of desire.[11] Here we present just one argument. The existence of a psychological system dedicated to a particular function is frequently accepted on the basis of evidence of a selective impairment in that function. For instance, autistic persons' selective impairment in social cognition has been taken as strong evidence for the existence of a

[11] See Holton (2009: 112-136).

psychological system dedicated to social cognition (Baron-Cohen, Leslie, & Frith 2003), and prosopagnosic persons' selective impairment in identifying faces has been taken as strong evidence for the existence of a perceptual system dedicated to face identification (Duchaine, Yovel, Butterworth, & Nakayama 2006; Kanwisher & Yovel 2006). In general, a functionally specific impairment that shows up across multiple subjects seems best explained by positing the existence of a functionally specialized psychological system that is impaired or damaged in that subject population. Furthermore, by comparing these impaired subjects to healthy controls, we can uncover the causal-functional roles of posited system.

Here we follow this broad strategy, contending that the behavioral abnormalities of two different populations are best explained by a selective impairment in self-control: patients with lesions in the *ventromedial prefrontal cortex* (vmPFC), and (healthy) subjects who have undergone *ego depletion*. However, our claim here is more limited than those that have been made about social cognition or face recognition. We are not arguing that the system involved in self-control is *exclusively* dedicated to the task: that would require showing that *only* self-control is affected in these subjects, which is far from obvious (not least because we are not yet clear on what counts as an exercise of self-control and what doesn't). Our point is rather that the subjects in the two groups show a systematic loss of self-control even though there is no reason to think that their desires and beliefs have been affected; and hence that we have good reason for positing some kind of system that is responsible for self-control, whether or not that system is also responsible for other, unrelated processes as well.

Our pairing of vmPFC lesions and ego depletion may seem surprising, given that the two subject groups have been studied separately and in different subdisciplines (neuropsychology and social psychology). However, these two groups have an important common feature: they both behave as we would expect people to behave who are motivated over-whelmingly by incentive salience desires. This seems to show that the motivational system that counteracts incentive salience desires' effects on behavior is selectively impaired in these subject groups. As we will argue, these subjects' deficits are best explained by appeal to the impairment of a psychological system that serves the function of governing behavior on the basis of cognitive desires. That is, these subjects seem to be suffering from selective impairment of the self-control system as we have described it.

This raises the question: how *should* we expect a person to behave who is motivated solely by incentive salience desires? We can make important predictions based on a single observation about how incentive salience desires are acquired: a dispositional incentive salience desire for an end state E is formed on the basis of past associations between E and a simultaneous dopamine reward signal (usually caused by pleasure, though sometimes caused by artificial dopamine stimulation, as with addictive drugs). The strength of a dispositional incentive salience desire for any end state E is proportional to a (recency-weighted) average of the past reward signals that have been associated with E (Holton and Berridge 2013).

Thus we can predict that incentive salience desires will only motivate agents to pursue ends that have been previously associated with co-occurrent reward. This means that agents will be unable to form incentive salience desires for ends that are not *immediately* rewarding, or not rewarding *to the agent*, since accomplishing these ends will not bring about a co-occurrent reward. This rules out two important kinds of ends. First, incentive salience desires will not

motivate agents to pursue *long-term* goals, which produce valuable or rewarding consequences only long after their end states have been attained. Examples of such goals include the goal to pass an examination, the goal to lower one's cholesterol, and (notably) the goal to quit an addictive drug: the benefits of achieving each of these goals accrue to the agent only long after the goal has been achieved. Second, incentive salience desires will not motivate agents to pursue *other-regarding* goals that, while they produce good consequences for others, are not immediately rewarding to the agent. Many moral and altruistic goals are likely to fall under this category: e.g. the goal to be honest when there is a prudential incentive to lie, the goal to avoid socially inappropriate or offensive behavior, and the goal to help others with whom one does not empathically identify. (This last qualification is necessary since there is some evidence that helping those with whom one does empathize can be rewarding in itself. In general our argument applies only to moral behavior that is not intrinsically pleasurable; and quite where the boundaries of that lie is not yet clear.) So we can predict that a person who is motivated solely by incentive salience desires will pursue predominantly *self-regarding* and *immediately rewarding* goals. In other words, they will be nasty, brutish, and shortsighted.[12]

Both vmPFC lesion patients and ego depleted subjects fit this prediction well. We'll start with the vmPFC lesion patients, as their deficit is more dramatic.

Since Phineas Gage, the first recorded and most famous case of vmPFC lesioning, the two most salient features of vmPFC-lesioned patients have been their severe deficits in socially appropriate behavior and long-term planning (Damasio 1994). vmPFC lesion patients usually display "acquired sociopathy," a disorder characterized by dampened and poorly regulated emotions as well as disturbed social decision-making. This typically causes vmPFC lesion patients, post-trauma, to be unable to maintain healthy social relationships or gainful employment (Damasio, Tranel, & Damasio 1990; Tranel, Damasio, Denburg, & Bechara 2005).

In addition to their sociopathic behavior, vmPFC lesion patients seem unable to base their behavior on the long-term consequences of their actions. The most famous demonstration of this deficit comes from the Iowa Gambling Task (IGT; Bechara, Damasio, & Damasio 1994). The IGT presents subjects with four decks of cards, which give differing monetary rewards when subjects draw from them. Two high-risk decks give large immediate rewards, but result in a long-term loss by giving even larger punishments; two low-risk decks present the long-term optimal option, yielding small but consistent rewards. Healthy control subjects will start by sampling all decks, temporarily favor the high-risk decks, and then learn to choose the low-risk decks after receiving punishment. vmPFC lesion patients, on the other hand, will continue to favor the high-risk decks throughout the task. The best explanation for this pattern seems to be that the vmPFC lesion patients are motivated by the short-term rewards offered by the high-risk decks, and cannot change their behavior on the basis of the cognitive desire to maximize their long-term payoff and the judgment that those decks have a suboptimal long-term predicted payoff.[13]

[12] Thanks to Matthias Jenny for the apt allusion.

[13] For more evidence beyond the IGT supporting the idea that vmPFC lesion patients are insensitive to long-term consequences, see Schoenbaum & Roesch (2005) and Moretti, Dragone, & di Pellegrino (2009).

As has been noted since the first studies, however, vmPFC lesion patients typically display normal intelligence, intact knowledge of social norms, and the ability to make accurate predictions about future social and non-social consequences (Saver & Damasio 1991, Leland & Grafman 2005). This indicates that these patients' psychological impairment is motivational rather than cognitive.

We submit that the best explanation for these results is that vmPFC lesion patients' behavior is guided over-whelmingly by the incentive salience system, which activates self-regarding and short-term goals. This is why vmPFC lesion patients show deficits in the two otherwise unrelated domains of moral behavior and long-term goal pursuit: both kinds of behavior require the capacity to set and pursue goals to achieve end states that are not immediately associated with rewarding experience.[14] However, these patients have normal explicit beliefs and evaluative judgments about what is good. So vmPFC lesion patients seem to be selectively impaired in their ability to act on their cognitive desires. This indicates that there is a psychological system, instantiated in or dependent upon the ventromedial prefrontal cortex, that (among other things) serves the function of controlling behavior on the basis of cognitive desires – i.e. the self-control system.

The self-control system can be impaired in healthy subjects as well, as is shown by studies on *ego depletion*. The ego depletion finding is that healthy (non-lesioned) subjects who exert self-control on one task will subsequently perform less well than control subjects on a second, unrelated task that also requires self-control.[15] The large literature on ego depletion has demonstrated that a wide range of tasks are ego depleting, from attention regulation (Gailliot & Baumeister 2007) to making choices (Vohs et al. 2008) to analytical thought (Schmeichel, Vohs, & Baumeister 2003). However, for our purposes, the most important ego depleting tasks are the motivational tasks, where subjects must exert self-control in order to override some desires in favor of others. On these tasks, ego depleted subjects show a similar pattern to vmPFC patients: they are selectively impaired in the pursuit of other-regarding and long-term goals.

(a) *Other-regarding goals*. The following results all support the claim that ego depleted subjects are less able to suppress selfish desires for the sake of other people:

- Ego depleted subjects are less likely to volunteer to help a victim of a tragedy (DeWall, Baumeister, Gailliot, & Maner 2008).
- Ego depleted subjects are more likely to lie about their performance for monetary gain (Mead, Baumeister, Gino, Schweitzer, & Ariely 2009).

[14] To return to an earlier point: we are not claiming that this is the *only* deficit that occurs in vmPFC lesion patients. Naturally occurring brain lesions are messy by nature and will rarely selectively impair a single psychological process without disrupting others. For instance, vmPFC lesion patients' reported abnormalities in moral judgment (Ciaramelli, Muccioli, Ladavas, & di Pellegrino 2007), social cognition (Shamay-Tsoory & Aharon-Peretz 2007), and affective experience (Damasio et al. 1990) are not straightforwardly explained by our hypothesis that they suffer from impaired self-control. However, we think our hypothesis provides a better explanation for vmPFC lesion patients' deficits in social behavior and long-term planning than the emotion-based explanation given by Damasio (1994), though we do not have the space to argue this point here.

[15] This finding has been replicated about 100 times (according to Inzlicht & Schmeichel 2012) and has been shown in a recent meta-analysis to be both highly significant ($p < .001$) and of medium-to-large size (Cohen's $d = 0.62$; Hagger, Wood, Stiff, & Chatzisarantis 2010).

- Ego depleted subjects express more interest in sleeping with someone other than their romantic partner, are less able to suppress sexually inappropriate thoughts, and are more likely to inappropriately engage in sexual behavior with their dating partner in the laboratory when given an opportunity to do so (Gailliot & Baumeister 2007).
- Ego depleted subjects are less effective at social self-presentation—for example, they are more likely to speak or disclose an inappropriate amount in conversation (Vohs, Baumeister, & Ciarocco 2005).
- Ego depleted subjects are more likely to respond destructively than constructively when their relationship partner behaves destructively (Finkel & Campbell 2001).
- Ego depleted subjects are more likely to respond with aggression after an insult (DeWall, Baumeister, Stillman, & Gailliot 2007).

(b) *Long-term goals*. The following results all support the claim that ego depleted subjects are less able to suppress short-term desires for the sake of long-term gain:

- Ego depleted subjects are less likely to choose to eat radishes rather than chocolates, or to restrain themselves from eating cookies when on a diet (Baumeister, Bratslavsky, Muraven, & Tice 1998).
- Ego depleted subjects' consumption of M&M's candies is better predicted by their implicit evaluations of M&M's than by their explicitly stated desires to eat healthy, while non-depleted control subjects' consumption of M&M's is better predicted by their explicit desires to diet than by their implicit evaluations (Hoffman, Rauch, & Gawronski 2007).
- Ego depleted subjects are less likely to restrain themselves from drinking too much beer when they expect to take a driving test afterward (Muraven, Collins, & Neinhaus 2002).
- Ego depleted subjects are less likely to choose to study for a test rather than procrastinate by reading magazines or playing video games (Vohs et al. 2008).
- Ego depleted subjects will drink less of a healthy but bad-tasting beverage (Vohs et al. 2008).
- Ego depleted subjects are more likely to spend money impulsively when given the chance (Vohs & Faber 2007).

All these seem to be cases where the long-term value of a future outcome (e.g. health, sobriety in a driving test, achievement, savings) needs to override a craving to pursue some immediately rewarding end (cookies, chocolate, beer, video games, impulse spending).

Like vmPFC patients, ego depleted subjects show a selective impairment that results in the relative domination of their behavior by incentive salience desires. Non-depleted subjects are better able to pursue long-term and other-regarding goals that cannot be activated by incentive salience desires. We think this data should be explained in the same way that we have explained the motivational deficits of vmPFC lesioned patients. Healthy, non-depleted human agents are different from vmPFC lesion patients and ego depleted subjects in that they have a fully functioning self-control system, which is impaired or absent in these other populations. The self-control system enables healthy agents to override their incentive salience desires and control their behavior in accordance with their cognitive desires. This allows their motivational

repertoire to include moral considerations, altruistic concern, and the long-term consequences of their actions. The powerful explanation of these two disparate bodies of data that we attain by positing the self-control system is, we submit, sufficient reason to accept its existence.

Intentional action, then, is the product of a competition between two different sorts of desires that is mediated by the self-control system. This thesis holds for both addicted and non-addicted agents. Both addicts and others have incentive salience desires, as we have already argued. In addition, both addicts and others have cognitive desires, and have self-control systems. One might be tempted to explain addiction as the result of an impairment of the self-control system, but we think this idea is a non-starter. If addicts had an impaired self-control system, we would expect them to show behavioral impairments across the board: they would not only have trouble controlling their addictive desires, but would be selfish and shortsighted across all other domains as well. But this is clearly not the case: addiction does not lead to the domain-general deficits characteristic of vmPFC lesion patients and ego depleted subjects. Unlike vmPFC lesion patients, addicts do not act like sociopaths; unlike ego depleted subjects, addicts do not seem to be impaired in *all* tasks that require self-control, such as attention regulation or analytic thought. Moreover, given the right incentives addicts do succeed in controlling even their addictive desires.[16]

Instead, we propose that the difference between addicted and non-addicted agents lies in the strength of their incentive salience desires. Due to the artificial stimulation of dopamine caused by addictive substances, the incentive salience desires involved in addiction are far stronger than any of the incentive salience desires typically experienced by non-addicted agents. Thus it is a far greater challenge for addicts to override their incentive salience desires, due to the abnormal motivational force of their addictive desire. Though this self-control challenge is far more difficult for addicts than it is for others, the *structure* of the challenge is the same for both, as we will now show.

4. Three stages of self-control

We have claimed that intentional action results from the competitive interaction amongst desires mediated by the self-control system. This is to see self-control as in the business of regulating which of the subject's desires gets to determine their behavior. But how does this work? Does self-control regulate which intentions the agent forms on the basis of their desires, or does it rather regulate whether they stick to their intentions? And mightn't it instead regulate which desires the agent has in the first place, or which judgments they form? We need to get clearer on what it is that self-control is controlling (or failing to control).

The philosophical literature on addiction presents several different, seemingly incompatible, answers to the question of where self-control breaks down in the case of addiction. Gary Watson argues that addictive desires bias addicts' *evaluative judgments* themselves, skewing deliberation so that they come to see taking the drug as the most attractive option. "One who is defeated by appetite is more like a collaborationist than an unsuccessful freedom fighter,"

[16] See Holton and Berridge (2013) for discussion of this.

Watson declares colorfully (1999, 7); and reiterates later: "We are not so much overpowered by brute force as seduced" (10). On this account, self-control works to control one's *judgments*.

In contrast, R. Jay Wallace argues that addictive desires make it difficult to *motivate* oneself to act on one's evaluative judgments once they have been formed. He emphasizes this in the passage we quoted earlier: "Even if one succeeds, in the face of such a desire, in reasoning correctly to the conclusion that it should not be acted on, its continued presence and urgency will make it comparatively difficult to choose to comply with the deliberated verdict one has arrived at" (648). On this account, self-control works to turn one's judgments into a commitment to action: in other words, to form an *intention* to act.

Finally, Timothy Schroeder and Nomy Arpaly emphasize the power of *habits* in producing addictive behavior, observing that these automatic behavioral dispositions may place addicts who are trying to get sober in tempting situations, situations that tend to undercut their intentions:

> The abstinent addict will do things without thinking about them at the time, only to find a difficult situation arising. 'Why did I agree to go to that party where everyone will be using?' 'Why did I turn down this street that leads me close to the dealers, and not down the next street?' 'Why did I end up calling my old drug buddy when I was bored?' Questions like these are often answered by an addict's unconscious behavioral tendencies (Schroeder & Arpaly 2013: 228).

On this account, self-control works even after one has formed an intention, to *implement* that intention in the face of the obstacles posed by one's bad habits.

Though each of these philosophers puts their favored locus of self-control conflict at center stage, we think there is no genuine disagreement between their claims. Instead, we favor a pluralist view: there are several distinct loci of self-control conflict. This view is advocated by Amelie Rorty in her classic article "Where Does the Akratic Break Take Place?" (Rorty 1980). Rorty begins by identifying several "stages on thought's way to action," and observes that "these distinctions allow us to locate the junctures where psychological akrasia can occur, in ways that explain the occurrence of behavioral akrasia" (334). These 'junctures' at which self-control failure can occur are also the places where self-control might be improved: "the place where the akratic break takes place also locates the place where the self-reforming akrates can best intervene to remedy his condition" (334).

In this section, we follow Rorty's strategy: first, distinguishing different stages by which thought leads to action; second, showing how self-control conflict arises at each of these stages; and third, showing how intervention at each of these stages can help an agent win the struggle to govern her own behavior.

We propose that there are at least three distinct stages by which thought leads to action, which we will call the *deliberative* stage, the *volitional* stage, and the *implemental* stage.[17,18]

[17] The stages we propose are inspired by Peter Gollwitzer's highly influential *Rubicon model of action phases* (Gollwitzer 1990). Though our division of stages does not correspond exactly with Gollwitzer's, we doubt this

(1) In the deliberative stage, the agent forms a *judgment* as to what action is best. This is the locus of self-control conflict Watson identified: the deliberative challenge of coming to a clear-eyed evaluative judgment of the consequences of potential actions in spite of the biasing influence of incentive salience craving.
(2) In the volitional stage, the agent chooses an *intention* to pursue. This is the locus of self-control conflict Wallace identified: the volitional challenge of willing yourself to pursue the end you have already judged to be best.
(3) In the implemental stage, the agent selects *actions* that implement her chosen intention. This is the locus of self-control conflict Schroeder and Arpaly identified: since habits are brute, unmotivated behavioral dispositions, they can cause goal-discrepant behaviors even when one is fully committed to a goal pursuit. So, it is during the implemental stage that one must grapple with and overcome one's habits.

We now proceed to discuss these stages in detail, with the aim of showing how self-control at each stage works similarly in addicted and non-addicted agents alike. For each stage we then briefly outline the ways in which self-control might be improved, again for both addicts and non-addicts alike.

4.1. The deliberative stage

4.1.1. The locus of deliberative self-control conflict: attention

As we have said (§3), a central function of the self-control system is to control behavior on the basis of an agent's all-things-considered judgments of the values of potential actions and their outcomes. But in order to do this, an agent must first form the evaluative judgments on the basis of which she aims to control her behavior. This involves creating mental simulations of various potential actions and their consequences, and then comparing them against one another on the basis of relevant evaluative criteria. This task of practical deliberation requires the agent to keep several different detailed simulations of actions in working memory simultaneously, attend to the evaluatively relevant features of each, and then compare them against one another. Since the capacity of working memory is limited, an agent will only be able to focus on a subset of the potentially relevant features of her different options. Thus what judgment she ultimately forms will depend to a large extent on what evaluatively relevant considerations capture her attention.

---

reveals a substantive disagreement, but rather reflects a difference in focus. Along similar lines, our stages are not the same as Rorty's proposed stages, but we think this is only because Rorty makes more fine-grained distinctions between stages than we do. Though we have limited ourselves to only those distinctions between stages for which we have empirical evidence, we are open to the possibility that there may be more useful distinctions between stages than we have made here.

[18] It is important to note that these stages are *goal-relative*: an agent might be in one stage relative to one goal while in a different stage relative to another. For example, an agent may have decided to take a trip to New York; having formed this intention, she is now in the implemental stage of this goal pursuit. However, in the process of implementing her intention, she will need to deliberate about further matters: should she take the train or a plane? Thus she might be in the deliberative stage regarding the question of *how to get to* New York even while she is in the implemental stage regarding her intention to *go* to New York. So the question to ask isn't: what stage of self-control is this agent in *full stop*; but rather: what stage of self-control is this agent in *for this particular goal pursuit*?

Consider, for instance, an alcoholic deliberating about whether to have another drink at a business dinner with a client. What choice she judges best will depend on what features of her options she attends to while deliberating. If she focuses exclusively on the features she finds attractive about the drink – the refreshing, pine-tree taste of a gin and tonic, the loose euphoria of inebriation – she will judge that having another drink is the thing to do. However, if she attends to the longer-term consequences of having another drink – the resulting drunkenness rendering her unable to comport herself appropriately in front of her client, her potentially losing business as a result, and the negative consequences of this for her professional reputation and career – she will likely judge that she ought to order a soda water instead. The judgment she makes about what is best to do will depend upon how she directs her attention during the process of deliberation.

The self-control and incentive salience systems will pull an agent's attention in different directions as she deliberates. An active incentive salience desire pulls an agent's attention to the attractive features of its object, thereby biasing the agent's deliberation in its favor. Only by exerting self-control can an agent attend to the reasons not to act in accordance with her incentive salience desires – i.e. the long-term consequences of her actions for things she reflectively values. It is thus over the control of attention that the deliberative stage of the competition between self-control and incentive salience is waged.

4.1.2. The role of the incentive salience and self-control systems in deliberation

If we are correct that the self-control system is the system that is impaired by ego depletion, then we can infer its functions from the capacities that are impaired in ego depleted subjects. It is thus instructive that ego depleted subjects show impairments in both analytic thought (Schmeichel et al. 2003; Wheeler, Briñol, & Hermann 2007; Masicampo & Baumeister 2008) and selective attention (Schmeichel et al. 2003; Gailliot & Baumeister 2007). Since practical deliberation requires both selective attention and analytic thought, we should expect ego depleted subjects to be impaired in this capacity as well. This means that the self-control system not only serves the function of controlling behavior on the basis of evaluative judgments already made, but is also deployed in the formation of evaluative judgments themselves.

However, the self-control system does not have complete sovereignty over attention. An active incentive salience desire exerts powerful influence over attention, drawing it toward the desired object and its most attractive features. This involuntary attentional pull has a significant biasing effect on practical deliberation. By automatically directing an agent's attention to the most attractive features of the desired object, an incentive salience desire can lead an agent to form evaluative judgments that give disproportionate weight to these features. This can lead agents subject to incentive salience cravings to form evaluative judgments that treat the desired object as much more valuable than they would judge it to be in the absence of craving.

This biasing effect has been demonstrated in empirical studies on both addicts and non-addicts alike. The most vivid display of this effect in non-addicts comes from a study in which the experimenters asked male subjects to answer survey questions while looking at pornography and masturbating (Ariely and Loewenstein 2006). The sexually aroused subjects, when

compared with non-aroused controls, reported being significantly more willing to engage in sexual behaviors they considered deviant (e.g. bisexual group sex) and to act immorally in order to have sex (e.g. slipping a woman a drug to get her to have sex). The influence of these subjects' active sexual desire went beyond their overt behavior, biasing even their *judgments* about what it would be pleasurable or morally acceptable to do. Less dramatically, some studies have shown that occurrent cravings for food make people overestimate how much they will enjoy foods in the future (Gilbert, Gill, & Wilson 2002; see also Van Boven & Loewenstein 2003). Hence why it is dangerous to go grocery shopping while hungry.

Addictive desires have the same kind of biasing influence on evaluative judgment as sexual desire and hunger, as demonstrated by Badger et al. (2007). Badger et al. studied a set of heroin addicts undertaking rehabilitation treatment who were receiving daily a heroin substitute medication Buprenorphine (BUP) to alleviate withdrawal symptoms. The experimental task asked these subjects to choose between receiving different amounts of money and receiving an extra dose of BUP, to be administered five days later.[19] The crucial manipulation was that one group of subjects was asked to make this choice while in a current state of craving, before they had received that day's dose of BUP, while a second group of subjects was asked to make the same choices while satiated, immediately after receiving their dose of BUP. The satiated subjects placed a substantially lower dollar value on the extra dose of BUP ($35) than the craving subjects, who valued the extra dose almost twice as much ($60). Notice that the difference in value here is for a dose to be received five days later—so subjects had no reason to think their current state of craving would have any influence on their enjoyment of the extra dose. And yet the currently craving addicts still judged receiving an extra dose five days later to be a more valuable outcome than the satiated addicts did. This seems best explained by the attention-biasing effect of active incentive salience desires: by drawing the craving subjects' attention to the attractive features of the extra BUP, their desire led them to judge it more valuable than they would have in the absence of craving.

4.1.3. How to improve deliberative self-control: mindfulness meditation

So, active incentive salience desires bias attention in both addicts and non-addicts, leading agents to disproportionately value the object of their current craving in their deliberative judgments about what is best. But agents can overcome this bias by exerting self-control, directing their own attention rather than letting it be guided by their current desire. This account yields a testable prediction: deliberative self-control can be aided by improving agents' selective attention. In other words, the better an agent's capacity to control her attention, the better she will be able to overcome the biasing influence of incentive salience-based temptation.

This prediction is confirmed by research on *mindfulness meditation*. Mindfulness is a traditional meditative practice that involves actively focusing one's attention on some aspect of one's present experience for an extended period of time. (Paradigmatically, one focuses on the experience of breathing). Among the many psychological benefits of training in mindfulness

[19] Subjects who chose the extra dose would receive two doses of BUP rather than one on the appointed day. This was a significant incentive: "Although a single dose of BUP is sufficient to eliminate addicts' acute cravings, a double dose produces a longer, more satisfying high" (Badger et al. 2007, 869).

meditation is an improvement in selective attention: both brief and long-term mindfulness training improve subjects' ability to selectively control their attention, as measured by many classic tests of attention regulation (Moore & Malinowski 2009; Jha, Stanley, Kiyonaga, Wong, & Gelfand 2010; Zeidan, Johnson, Diamond, David, & Goolkasian 2010). If our picture of deliberative self-control is correct, then these improvements in selective attention should help subjects to better resist incentive salience desires. And this is exactly what the data shows.

This prediction has been robustly confirmed in studies of addicts (for a review, see Brewer, Elwafi, & Davis 2012). Randomized and controlled studies testing a mindfulness training intervention for addiction have shown that mindfulness training leads to a significant reduction in use of the addictive substance and a significantly lower chance of relapse, both when compared to a no-treatment baseline (Zgierska et al. 2008; Bowen & Marlatt 2009) and when compared to conventional addiction treatments (Witkiewitz, Marlatt, & Walker 2005; Bowen et al. 2009; Brewer, Elwafi, & Davis 2011). One study found that smokers high on dispositional mindfulness measures are less likely to relapse after quitting than smokers lower in dispositional mindfulness (Vidrine et al. 2009). Finally, at least two studies have found that addicts who undergo mindfulness training not only use the addictive substance less, but also experience less intense *cravings* for the substance (Bowen et al., 2009; Westbrook et al. 2011).

Mindfulness-based interventions help non-addicts to overcome incentive salience temptations as well. In particular, several studies have shown mindfulness training to help obese or overweight subjects to achieve their weight-loss goals (Forman, Herbert, Moitra, Yeomans, & Geller 2007; Lillis, Hayes, Bunting, & Masuda 2009; Tapper et al. 2009; see also Alberts, Mulkens, Smeets, & Thewissen 2010; Marchiori & Papies 2014). In a recent review, O'Reilly Cook, Spruijt-Metz, & Black (2014) found that 18 out of 21 reviewed studies of mindfulness-based interventions for obesity-related behaviors reported significant decreases in the targeted behaviors.

One study directly supports our hypothesis that the mechanism behind these successful interventions is an improvement in deliberative self-control (Hendrickson & Rasmussen 2013). This study investigated the temporal discounting of food rewards in obese and healthy-weight individuals by offering them a choice between a large, delayed food reward and a small, immediate food reward. In an initial test, obese subjects showed a much steeper discounting curve than controls—that is, they were willing to give up a larger delayed reward for a smaller immediate reward. This is what we would expect, given that the obese subjects are experiencing a stronger incentive salience craving for food, which draws their attention disproportionately to the attractive features of the immediate reward. After the initial test, some of the obese subjects undertook a 50 minute training session in mindful eating, while others just watched an educational video on nutrition. These subjects then completed the temporal discounting test again. Obese subjects who underwent mindfulness training subsequently showed a significantly less steep discounting curve than they had in the initial test: they were more willing than before to give up a smaller immediate reward for the sake of a larger delayed reward. (Subjects who watched the educational video showed no such improvement). What this suggests is that the brief mindfulness training session helped the obese subjects to overcome the biasing effect of their food cravings and form more normal judgments about the relative values of immediate and delayed rewards. In other words, mindfulness training improved these subjects' deliberative self-control.

We submit that our model of deliberative self-control provides the best explanation for the above results. An important first step in overcoming an active incentive salience desire is to form a clear-eyed evaluative judgment that indulging one's craving will lead to worse consequences than refraining from doing so. An active incentive salience desire automatically biases one's attention to the positive features of the object desired, leading agents to overestimate the value of satisfying their current desire. Mindfulness meditation training makes agents more skilled at self-controlled attention regulation, and thereby improves their ability to resist the biasing effect of active incentive salience desires on evaluative judgment. It is thus by improving agents' capacities for deliberative self-control that mindfulness meditation helps addicts and non-addicts alike resist the influence of their incentive salience desires.

4.2. The volitional stage

4.2.1. The locus of volitional self-control conflict: goals

The second stage of self-control is the volitional stage: after one judges what is best (deliberative stage), one must choose a goal to pursue (volitional stage) before one begins implementing that goal pursuit in one's behavior (implemental stage). In other words, between judgment (deliberative) and action (implemental) lies *choice* (volitional), and to exert volitional self-control is to exert self-control in choosing a course of action. This was the self-control task identified by R. Jay Wallace, of "choos[ing] to comply with the deliberated verdict one has arrived at" (648).

Some readers may be skeptical that the act of making a choice is really distinct from the act of forming an evaluative judgment. Our first response would be to note that the possibility of *akrasia*, choosing against one's own best judgment, seems to require such a distinction. But, of course, people who are skeptical about the judgment/choice distinction will be skeptical about the existence of *akrasia* as well, and so this line of argument will seem to be begging the question.[20]

However, we think there is empirical evidence demonstrating that making a choice is psychologically distinct from forming an evaluative judgment. A study by Kathleen Vohs and colleagues (2008, Study 6) shows that choosing to act on one's evaluative judgments (volition) requires more self-control than merely forming evaluative judgments (deliberation). In this study, all subjects were presented with a webpage that gave various options for customizing a desktop computer for purchase. Some subjects were asked to choose between the customizations (the choice condition), while others were merely asked to consider the customization options and "form an opinion of the information, thinking about what [they] would prefer" (892), but importantly, were *not* asked to implement their judgments by selecting their preferred options on the website (the deliberation condition).

[20] It might also be the case that evaluative judgments are formed *subsequently* to the intentions: in the light of what they have decided to do, cognitive dissonance motivations might lead them to form judgments that present those decisions in the good light. But we still contend that the intentions and the judgments have genuine independent existence. See Holton (2009: 1-19).

The dependent measure of this study was subjects' subsequent persistence on an impossible anagram task, a task that has been shown to measure self-control capacity (Baumeister et al. 1998). What Vohs et al. found was that subjects in the choice condition, who had made a series of active choices, persisted significantly less on this task than subjects in the deliberation condition. This shows that the act of *choosing* involves an exertion of self-control that goes beyond the self-control required to form an evaluative judgment. These results not only dissociate choice from evaluative judgment, but also show that choice involves the exertion of self-control. In other words, this study establishes the existence of volitional self-control as a psychological task that is distinct from deliberation to a judgment.

So let us take as given the existence of volitional self-control and now ask what it involves. What is the psychological process involved in making a choice, and why might it require self-control?

We suggest that the exercise of choice involves the selection and activation of a kind of motivational mental state that psychologists call a 'goal'. A *goal*, in the technical sense used by psychologists, is a mental representation of a desired end that directs behavior in pursuit of that end.[21] We take it that such states often constitute *intentions*, as philosophers understand this term. The large research literature on goals, which we do not have the space to review here, has shown them to be a robust psychological natural kind with a distinctive suite of cognitive and behavioral signatures (for a review, see Förster, Liberman, & Friedman 2007). Active goals direct attention, cognition, and behavior in a flexible and instrumentally rational way in order to bring about the end state that is their content. One primary way for a goal to be activated is simply for subjects to form a conscious, deliberate intention to pursue a certain end. We thus submit that volitional choice is best understood as the self-controlled act of activating a goal with a certain end.

4.2.2. The role of the incentive salience and self-control systems in volition

We have already seen how self-control plays a role in volition: Vohs et al.'s subjects had to exert self-control to go beyond forming a judgment and activate a goal to act in accordance with that judgment. Crucially for our purposes, however, self-controlled choice is not the *only* route by which goals can be activated. Goals are also activated automatically by incentive salience desires, as we shall now explain.

A series of experiments by Henk Aarts and Rudd Custers have demonstrated that a goal to pursue a certain end state can be nonconsciously activated by subliminally associating positive affect with that end state (Custers & Aarts 2005b; 2005a; 2007a; Aarts, Custers, & Veltkamp 2008; Bijleveld, Custers, & Aarts 2010; 2011). Aarts and Custers first demonstrated that subliminally associating positive affect with a goal caused subjects to report greater *wanting* to pursue the goal (Custers & Aarts 2005b: Study 1), and then showed in subsequent studies (cited above) that this greater wanting leads subjects to behave in the ways characteristic of goal

[21] We are thus using the term 'goal' to refer not to the state of affairs one is pursuing (as 'goal' does when used colloquially, e.g. "my goal is to lose 5 pounds"), but rather to the mental state that guides one's behavior towards bringing about that state of affairs.

activation. These results seem best explained by appeal to incentive salience desires. We have already seen (§2) that incentive salience desires are proportional in strength to the previous association of the desired object with reward, and are automatically activated by encounters with desire-associated stimuli. Thus we should expect that Aarts and Custers' intervention to associate positive affect with an end state would activate an incentive salience desire to attain that end state. And as we would predict, this association leads subjects to *want* to attain the goal. This gives us good reason to think that Aarts and Custers have activated goals in their subjects *by means of* creating and triggering incentive salience desires. Thus their findings strongly indicate that an active incentive salience desire for an object automatically and nonconsciously activates a *goal* to attain that object, which then directs behavior in pursuit of its attainment.

On reflection, this is exactly what we should expect. The incentive salience cravings that addicts feel for heroin or non-addicts feel for sugar or sex do not merely influence behavior by biasing deliberative judgment. These desires seem to have *direct* motivational power, *pushing* the addict to shoot up or the non-addict to bite into the cake before either has a chance to even consider whether this is a good idea. Incentive salience desires seem to directly guide behavior in the absence of counteractive self-control, and now we can see why: cravings activate *goals*, which automatically guide action toward the attainment of the thing that is craved.

Thus the challenge of volitional self-control in the face of an active incentive salience desire is to resist the automatic activation of the goal to attain the desired object, and instead activate an alternative goal that accords with one's deliberative judgments about what is best. Only one goal can guide behavior at a time; in fact, a dominant goal actively *suppresses* the accessibility of the most attractive alternative goals (Shah, Friedman, & Kruglanski 2002). Thus the self-control system and the incentive salience system can be seen as competing in a 'horse race' of goal activation, where the winning system is the one whose favored goal is made most active and thereby comes to dominate downstream behavior. The stronger the incentive salience desire, the more activation it will give to its favored goal, and thus the greater exertion of self-control will be required to activate an alternative goal enough to override it. Hence why restraining yourself from acting on an addictive desire is a far more difficult task than restraining yourself from eating a chocolate cake.

4.2.3. How to improve volitional self-control: mental contrasting

If volitional self-control is a matter of giving sufficient activation to one's deliberatively chosen goal, then we should expect that any procedure that leads to greater activation of a consciously chosen goal will help agents to overcome temptation by incentive salience desires. The 'mental contrasting' procedure, created and researched by Gabriele Oettingen, is an intervention of this kind. In this procedure, subjects who wish to attain a goal are asked to undertake two imaginative steps: first, imagine a 'positive fantasy' of the goal's being attained, and all the beneficial consequences that would follow goal attainment; second, *mentally contrast* this positive fantasy with the 'negative reality' of one's present distance from achieving the goal and the obstacles lying in the way of goal attainment. Several studies have shown that this mental contrasting procedure powerfully increases subjects' motivation to attain the goal, causing them to expend much more effort in pursuit of the goal (Oettingen, Pak, & Schnetter 2001; Oettingen

et al. 2009; Oettingen, Mayer, & Thorpe 2010a; Oettingen, Stephens, Mayer, & Brinkmann 2010b; Oettingen 2012). What explains this effect?

We offer the following explanation. Goal pursuit research has independently shown that the activation level of a goal is automatically modulated based on three major factors: (a) *value*, the perceived value of achieving the goal (Aarts, Gollwitzer, & Hassin 2004; Förster, Liberman, & Higgins 2005; Cesario, Plaks, & Higgins 2006; Custers & Aarts 2007a); (b) e*xpectancy*, the perceived probability of attaining the goal (Förster et al. 2005); and (c) *discrepancy*, the perceived effort required to attain the goal (Rothermund 2003; Kawada, Oettingen, Gollwitzer, & Bargh 2004; Custers & Aarts 2007b). Goal activation is strongest when expectancy, value, and discrepancy are all high.

We propose that the mental contrasting procedure activates goals by means of boosting value and discrepancy: the 'positive fantasy' increases the perceived value of attaining the goal, while the 'negative reality' increases the perceived effort required to attain the goal. In line with this explanation is the finding that subjects who only complete the 'positive fantasy' component of the procedure become *less* motivated to attain the goal (Kappes & Oettingen 2011). Though this might seem initially surprising, it is easily explained by noting that the positive fantasy on its own will sharply *decrease* the discrepancy attributed to the goal, as subjects imagine the goal to already be completed; it is this decrease in discrepancy that demotivates these subjects.[22] This is why the 'negative reality' contrast, which counteracts the adverse effects of the 'positive fantasy' component on discrepancy while maintaining its positive effects on value, is necessary for the mental contrasting procedure to work.

Thus the mental contrasting procedure is well-designed to increase the activation of a consciously chosen goal. So, given our characterization of volitional self-control, we should expect the mental contrasting procedure to help agents overcome temptation by incentive salience desires. And this is what we find. Oettingen et al. (2010a) found that the mental contrasting intervention caused smokers who wanted to quit to take more immediate action towards quitting than subjects who underwent a control intervention. And for non-addicted subjects, Johannessen et al. (Johannessen, Oettingen, & Mayer 2012) found that dieters who performed the mental contrasting procedure were significantly more successful than control subjects at reducing their caloric intake over a two-week period.

We have portrayed volitional self-control as involving a competition between the self-control and incentive salience systems over the activation of goals. We take this picture to be nicely confirmed by the fact that the mental contrasting procedure, which increases the activation of deliberatively chosen goals, helps agents to overcome temptation by both addictive and non-addictive incentive salience desires. Mental contrasting helps agents succeed in *motivating* themselves to act in accordance with their deliberative judgment—which, as we have seen, is not a trivial task.

[22] In fact, the act of imagining goal completion has been shown in one study to lead to 'goal turnoff,' the suppression of goal accessibility that usually occurs after the goal has *actually* been completed (Denzler, Förster, & Liberman 2009).

4.3. The implemental stage

4.3.1. The locus of implemental self-control conflict: habits

As we have said, a goal, once activated, will automatically guide behavior toward its own fulfillment. Thus one might think that choosing the right goal in the face of temptation is sufficient for controlling one's behavior. However, goal implementation—the process of executing one's chosen goal pursuit in action—itself poses nontrivial self-control challenges.

This is because goals are not the only mental states that directly influence behavior. There are also *habits*, which Neal, Wood, & Quinn (2006) define as "response dispositions that are activated automatically by the context cues that co-occurred with responses during past performance" (198). In other words, habits are associations between contexts and behaviors that lead agents to produce a certain behavior when they encounter a certain contextual cue.

For our purposes, it is important to distinguish habits both from goals and from incentive salience desires. The distinction between habits and goals is essential to understanding the difference between the volitional and implemental stages of self-control. And as we emphasized earlier (§2), the habits that are produced by addiction are an importantly different phenomenon from the incentive salience desires that produce addiction. Habits and incentive salience desires may each exert their influence in the absence of the other, though they often go hand in hand.

The primary feature that distinguishes habits from goals is their *motivation-independence*. As habits are associative states that produce a behavior directly when a certain context is encountered, they do not depend for their influence on any motivation to engage in the relevant behavior. This is in contrast with goals, which are almost always activated by and dependent upon a desire to achieve some end.[23] When one ceases to desire the end of a certain goal pursuit, the goal itself is deactivated (Aarts, Custers, & Holland 2007); in contrast, when one ceases to desire the end that is served by a certain habit, the habit remains (Neal, Wood, Wu, & Kurlander 2011). One might, for instance, habitually make a turn that follows the well-worn driving route to one's workplace, when in fact one does not want to go there at all, but rather is going to a restaurant which is actually in the opposite direction. However, one will never set out to pursue the goal of going to one's workplace when in fact one has no desire whatsoever to do so.

The primary feature that distinguishes habits from incentive salience desires is their *motivational neutrality*. In addition to exerting their influence independently from (and even contrary to) one's prior motives, habits also do not *produce* any desire to perform the habitual behavior. In other words, one does not *crave* acting out one's habits. Schroeder and Arpaly (2013) make this point well:

> When one does not do something one wanted to do, there is often a little disappointment or regret. But when one does not make a habitual left turn, there is no disappointment or regret that coincides with not acting out of habit…[one] neither longingly thinks of making the left turn when at other intersections, nor is

[23] A possible exception to this claim is the case of unconscious goal priming by exposure to words semantically associated with a goal (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel 2001).

> behaviorally disposed to get into a position to make the left turn. The habit only has influence upon behavior (231).

This apt observation about the different phenomenologies of habit and desire is confirmed by empirical research. As we have already mentioned (§2), simply learning to notice a habitual behavior seems to be sufficient for ceasing it, implying that once the subject becomes aware of the habitual behavior, it takes little additional self-control to override it (Ladouceur 1979; Bate et al. 2011). Contrast this with incentive salience desires, which are still quite difficult to override even when one is reflectively aware of them.

A third feature of habits distinguishes them from both goals and incentive salience desires: their *behavioral inflexibility*. Neal and Wood (2010) observe that "people rarely substitute habitual behaviors (e.g., a habit of daily jogging) for alternative behaviors that meet the same ostensible goal (e.g., switching from jogging to cycling)" (449). We think this observation reflects an important fact about the structure of habits: they are associations of contexts with *a particular behavior*, not with an end that can be brought about by many different behaviors. Habits rigidly produce a certain behavior, never switching to producing a different behavior that better facilitates some goal. This is illustrated by a study on habitual popcorn eating in the cinema, in which subjects ceased to habitually eat popcorn if they were forced to do so with their nondominant hand (Neal, Wood, & Kurlander 2011). This result shows that these subjects' habit was not really *to eat popcorn*, but rather *to scoop popcorn into their mouths using their dominant hand*. When this behavior was no longer possible, the habit did not cause the subjects to engage in the alternative behavior of eating with their nondominant hands—because *that* is not the particular behavior they associate with the context of the cinema. In contrast, both goals and incentive salience desires are very flexible in the behaviors they produce, dynamically switching between behavioral routines when doing so is adaptive for achieving their end (Bargh et al. 2001; Hassin, Bargh, & Zimmerman 2009).

In summary, habits are best understood as a brute, direct association between a specific context and a rigid behavior, which produces behavior in a way that is unmediated by desire. This distinguishes habits from both goals and incentive salience desires, allowing us to see the task of controlling one's habits as distinct from the task of controlling one's goals. As it arises in the implementation of one's goals, we will call this task the *implemental stage* of self-control.

4.3.2. The role of the incentive salience and self-control systems in creating habits

As Aristotle observed (1984, *Nicomachean Ethics* 1103a-b) and contemporary research has confirmed (Danner, Aarts, & Vries 2010), habits are created by repetition. More precisely, a habit to perform a certain behavior in a certain context is created by an agent's performing that particular behavior in that particular context many times before. This repetition ingrains the automatic association between context and behavior that constitutes the habit.

Both the incentive salience and self-control systems can create and sustain habits by this simple method. If an incentive salience desire is served by regularly performing the same behavior in the same context (say, ordering your usual beer at your favorite bar, or reaching for

the ice cream in your freezer upon arriving at home), then by repeatedly acting on that incentive salience desire, one may create a habit that serves the desire. Insofar as one disapproves of the incentive salience desire, these may be called 'bad habits.' Addicts, who usually spend a good while acting on their addictive desire before seeking help, will thereby acquire many habits that facilitate their addictive behavior. These 'bad habits' will remain even when the addict has overcome her desire for the addictive substance, and may make it more difficult for the addict to remain in control, as Schroeder and Arpaly point out (2013, 228).

On the other hand, one may also inculcate 'good habits' by repeatedly performing a behavior in a context that facilitates one of the cognitive desires or values on the basis of which one exerts self-control. For instance, one might create a habit of walking to the gym immediately after leaving work by simply exerting the self-control required to do so deliberately every day, until it becomes automatic and effortless. Many other examples of the self-controlled creation of habits come from athletics, music, and other skilled behaviors, where one exerts a great deal of self-control to repeat a certain behavior in a precise way during practice (whether a scale on the violin or a free-throw in basketball) and then, as one becomes skilled, is able to do the same behavior automatically and habitually. This self-controlled formation of 'good habits' works just the same way as the formation of 'bad habits' by the incentive salience system: produce the same behavior in the same context over and over again, and *voila!*—a habit is born.

4.3.3. How to improve implemental self-control: implementation intentions

Implemental self-control becomes a challenge when one has a *good goal* that may be thwarted by a *bad habit*. In other words, even once you have succeeded at *volitional* self-control, activating a goal that accords with your cognitive desires, your pursuit of this goal may be hampered by habits that lead to goal-discrepant behaviors. This problem will be especially dire if, as in the case of addicts, one's goal is to change one's behavior from a longstanding pattern produced by the pursuit of a powerful incentive salience desire. As Schroeder and Arpaly observe, bad habits may tip the balance in the addict's self-control conflict, as when an addict finds herself habitually putting herself in situations that make drugs available or tempting.

One strategy for implemental self-control is simply to directly override the habit once it has been triggered. Though this works, it is difficult, causing ego depletion in ordinary subjects (Baumeister et al. 1998; DeWall et al. 2008). Overriding a habit is difficult not necessarily because it is difficult to overcome a habit once it has been detected, but because it requires a great deal of attention regulation to constantly monitor for the cues that trigger the habitual behavior. Given the limitations of our resources for self-controlled attention, this strategy for overcoming bad habits is itself quite limited.

An implemental self-control strategy that may escape these limits is suggested by research on *implementation intentions*, a technique created and investigated by Peter Gollwitzer. Implementation intentions are plans of the form "*if* I encounter X cue, *then* I will perform Y response!" Subjects who form implementation intentions to aid them in a goal pursuit have been shown in a large number of studies to pursue their goals much more effectively than subjects who simply form a goal intention (of the simpler form "I will do X!"). A meta-analysis of 94

studies involving over 8,000 participants found that the improvement of goal pursuit by implementation intentions over mere goal intentions is highly statistically significant, and medium-to-large in effect size (Cohen's $d = 0.65$; Gollwitzer & Sheeran 2006).

The helpful effects of implementation intentions seem to be largely due to the automatic association such intentions create between the 'if' cue and the 'then' response. Subjects who form implementation intentions afterwards show a strong automatic association between the 'if' cue and the 'then' response, reacting far more quickly than controls to words associated with the 'then' response after primed with the 'if' cue (Aarts, Dijksterhuis, & Midden 1999; Webb & Sheeran 2007; 2010; Adriaanse, Gollwitzer, De Ridder, de Wit, & Kroese 2011a). This association leads subjects to quickly and automatically execute the intended 'then' response when they encounter the specified 'if' cue. The automaticity of this process explains why implementation intentions are just as effective (and in some cases *more* effective) when subjects suffer from impairments in executive control caused by cognitive load (Brandstätter, Lengfelder, & Gollwitzer 2001; Cohen, Bayer, Jaudas, & Gollwitzer 2006), ego depletion (Webb & Sheeran 2003), drug withdrawal (Brandstätter et al. 2001), schizophrenia (Brandstätter et al. 2001), ADHD (Gawrilow & Gollwitzer 2007; Gawrilow, Morgenroth, Schultz, Oettingen, & Gollwitzer 2012), or old age (Zimmermann & Meier 2009). The automaticity of implementation intentions is also indicated by studies showing that subjects will execute the 'then' response of their implementation intentions even when the 'if' cue is presented subliminally (Gollwitzer & Schaal 1998; Bayer, Achtziger, Gollwitzer, & Moskowitz 2009).

The attentive reader will have already noticed that the kind of state created by implementation intentions—an automatic association between a cue and a response—is one and the same as the kind of state we have identified with *habit*. This implies that implementation intentions can enable an agent to deliberately create *new* cue-response associations that can compete with and override her old cue-response associations, i.e. her habits. If this is correct, then implementation intentions may provide a powerful tool for overriding unwanted habits and thus improving implemental self-control.

The research has borne this hypothesis out: subjects who form implementation intentions are significantly more successful at creating new habits and overriding old habits than control subjects who form mere goal intentions to do so (Aarts & Dijksterhuis 2000; Sheeran & Orbell 2000; Holland, Aarts, & Langendam 2006; Orbell & Verplanken 2010; Webb, Sheeran, & Luszczynska 2010; Adriaanse et al. 2011a). As we would expect, reaction-time tasks indicate that implementation intentions break habits by creating a new association between the cue and the intended 'then' response, which competes with the old association between the cue and the habitual response. After forming an implementation intention to break a habit, subjects react equally quickly to words associated with the intended 'then' response as they do to words associated with the habitual response, indicating that the implementation intention levels the associative playing field (Adriaanse et al. 2011a). As the experimenters themselves put it: "implementation intentions eliminated the cognitive advantage of the habitual means in the 'horse race' with the alternative response" (Adriaanse et al. 2011a, 503). This gives the agent's self-control system a much better chance of winning the larger 'horse race' with the incentive salience system for the control of behavior.

We should thus predict that forming implementation intentions should help agents to

overcome incentive salience temptation; and the available data supports this prediction. With regards to non-addicted subjects, many studies have shown implementation intentions to significantly improve success in *dieting*, an activity that requires overcoming incentive salience desires for unhealthy foods (Verplanken & Faes 1999; Achtziger, Gollwitzer, & Sheeran 2008; Adriaanse, Vinkers, De Ridder, Hox, & De Wit 2011b). Regarding the effectiveness of implementation intentions in overcoming addiction, there is an unfortunate dearth of research. However, one study has found that forming implementation intentions helped adolescents to quit smoking, though only for those who had a 'weak or moderate' smoking habit as measured by a standard scale (Webb, Sheeran, & Luszczynska 2010).

It is important to note that since implementation intentions aid specifically with implemental self-control, they will only facilitate self-control success among subjects who have already succeeded in overcoming their incentive salience desires in both the deliberative and volitional stages of self-control. If self-control fails in either of these prior stages, then the deck will be stacked too heavily in favor of the incentive salience system for a purely implemental intervention such as forming implementation intentions to make much of a difference. Perhaps this is why implementation intentions on their own did not affect the most addicted subjects' success at quitting smoking.

More generally, since success at all three stages of self-control is required for an agent to fully overcome incentive salience temptation, the most effective interventions to aid self-control will involve a combination of the stage-selective interventions we have advocated here. One existing intervention that follows this prescription is Gollwitzer and Oettingen's 'Mental Contrasting with Implementation Intentions' (MCII) method, in which subjects first undergo the mental contrasting procedure—thus facilitating volitional self-control—and then form implementation intentions—thus improving implemental self-control. It should be no surprise that the MCII method is highly effective in aiding subjects to achieve their goals (Oettingen, Hönig, & Gollwitzer 2000; Stadler, Oettingen, & Gollwitzer 2009; Duckworth, Grant, Loew, Oettingen, & Gollwitzer 2011; Gawrilow, Morgenroth, Schultz, Oettingen, & Gollwitzer 2012; Houssais, Oettingen, & Mayer 2013). We can speculate that combining mindfulness training with the MCII method would augment self-control even further, comprising a 'triple threat' of interventions that improve self-control in the deliberative, volitional, and implemental stages. Whether or not this 'MMMCII' method (Mindfulness Meditation, Mental Contrasting, and Implementation Intentions) would in fact be effective in overcoming both addictive and non-addictive temptation is a question for further empirical work.

5. Conclusion

In this paper we have proposed and defended a model of self-control that applies to addicts and non-addicts alike (represented in Figure 1 below).

Intentional action is the product of a competition between (at least) two systems: the incentive salience system and the self-control system. As we argued in §2, the incentive salience system is not only the source of addictive desires, but is the source of many of our ordinary, non-addictive desires as well. Due to the associative manner in which they are formed, these

incentive salience desires are stubbornly independent of an agent's reflective judgments about what is valuable. This gives rise to the problem of self-control: the challenge of resisting one's incentive salience desires when they do not align with one's cognitive desires. We argued in §3 that the capacity to exert self-control plays an independent role in determining behavior over and above the relative strengths of an agent's desires. This fact is illustrated most vividly by cases where the capacity to exert self-control is impaired (as in ego depletion) or lost altogether (as in vmPFC lesioning). The empirical evidence thus lends significant credence to the Platonic idea that there are two parts of the soul, one rational and the other appetitive, that compete for control over action.

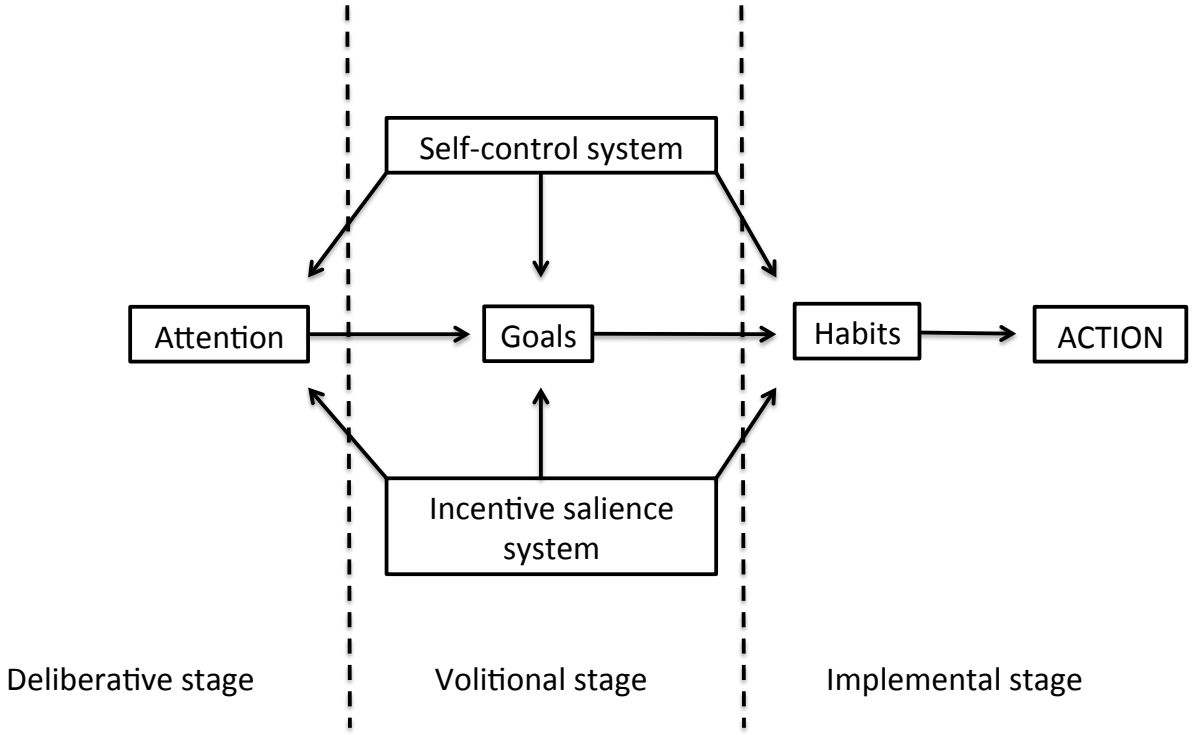

Figure 1: the three-stage model of self-control

As we argued in §4, this competition proceeds in stages. We distinguished three of these stages: deliberative, volitional, and implemental. In the deliberative stage, an agent forms a judgment as to what course of action would be best. Since the judgment the agent reaches depends upon the considerations she attends to when deliberating about what to do, deliberative self-control is a matter of directing *attention* in order to resist the biasing pull of craving. In the volitional stage, an agent forms an intention to act in accordance with her deliberative judgment. What this amounts to is the activation of a *goal*, a mental state that guides behavior toward the achievement of a certain end. Since incentive salience desires automatically activate goals regardless of whether the agent judges them good, an agent must exert self-control in order to make her goals accord with her evaluative judgments. Finally, in the implemental stage, an agent must guide her behavior in pursuit of her chosen goal. Whether she succeeds in doing so depends upon her habits—the automatic associations between contexts and behaviors she has formed in the past. Since habits guide behavior independently from goals, the regulation of *habits*—both by overcoming bad habits and by forming good ones—is a third task of self-control, separate from the two preceding. An agent must succeed in all three of these stages of self-control in order to conform her actions to her cognitive desires.

This single model captures the predicaments of the addict and non-addict alike. The incentive salience desires that render the addict's actions so wildly out of sync with her values are present in non-addicts as well, though in less extreme form. And thus the non-addict will also sometimes act in ways she does not endorse, driven by desires that motivate independently of her conception of the good. The non-addict can resist these desires by exerting self-control; but the addict can do this too. The task of self-control is far more difficult for the addict—which is why

it is often unreasonable to blame addicts for giving in to temptation even when we might blame a non-addict for doing so. But self-control is possible for addicts, especially with strong incentives and assistance from others. Indeed, this is just what recovery from addiction is: the addictive desire does not go away, but the recovering addict learns to control her behavior in spite of it.

Thus addicts are not so different from the rest of us as we may have thought. But that may be because we underestimated our own similarity to addicts, rather than the other way around. There is a tendency to think of human agency as an entirely rational affair: we simply do whatever we think is most likely to get us what we want. The heuristics and biases literature has undermined this picture somewhat over the past few decades, but only by showing us how we are not always rational in selecting the *means* to our ends (Ariely 2008; Kahneman 2011). The model we have defended here shows that the irrationality—or arationality—of human agency goes a step deeper: our ends themselves can be set by desires that are utterly divorced from what we take to be rationally desirable. The activity of controlling our actions is thus not merely a matter of figuring out what we ought to do; it is a matter of fighting to control our minds and actions in accordance with our reasons. To borrow Plato's metaphor, being a human agent is more like struggling with stubborn horses for control over a chariot than it is like calculating a utility function. Those of us who are lucky enough not to suffer from addiction might come to understand ourselves better by acknowledging that there is an addict in us all.[24]

[24] We thank Dylan Bianchi, Matthias Jenny, Bernhard Salow, Ian Wells, and the participants at the Mechanisms of Self-Control Workshop at King's College London, and at The Affective Face of Desire Symposium at Rennes, for helpful comments and discussion.

References

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology*, *78*(1), 53–63.

Aarts, H., Custers, R., & Holland, R. W. (2007). The nonconscious cessation of goal pursuit: When goals and negative affect are coactivated. *Journal of Personality and Social Psychology*, *92*(2), 165.

Aarts, H., Custers, R., & Veltkamp, M. (2008). Goal priming and the affective-motivational route to nonconscious goal pursuit. *Social Cognition*, *26*(5), 555–577.

Aarts, H., Dijksterhuis, A. P., & Midden, C. (1999). To plan or not to plan? Goal achievement or interrupting the performance of mundane behaviors. *European Journal of Social Psychology*, *29*(8), 971–979.

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, *87*(1), 23–37.

Achtziger, A., Gollwitzer, P. M., & Sheeran, P. (2008). Implementation intentions and shielding goal striving from unwanted thoughts and feelings. *Personality and Social Psychology Bulletin*, *34*(3), 381–393.

Adriaanse, M. A., Gollwitzer, P. M., De Ridder, D. T. D., de Wit, J. B. F., & Kroese, F. M. (2011a). Breaking habits with implementation intentions: A test of underlying processes. *Personality and Social Psychology Bulletin*, *37*(4), 502–513.

Adriaanse, M. A., Vinkers, C. D. W., De Ridder, D. T. D., Hox, J. J., & De Wit, J. B. F. (2011b). Do implementation intentions help to eat a healthy diet? A systematic review and meta-analysis of the empirical evidence. *Appetite*, *56*(1), 183–193.

Ahmed, S. H. (2010). Validation crisis in animal models of drug addiction: Beyond non-disordered drug use toward drug addiction. *Neuroscience & Biobehavioral Reviews, 35*(2), 172-184.

Ahmed, S. H., Guillem, K. and Vandaele, Y. (2013). Sugar addiction: Pushing the drug-sugar analogy to the limit. *Current Opinion in Clinical Nutrition and Metabolic Care, 16*(4), 434-439.

Alberts, H. J. E. M., Mulkens, S., Smeets, M., & Thewissen, R. (2010). Coping with food cravings. Investigating the potential of a mindfulness-based intervention. *Appetite*, *55*(1), 160-163.

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: the effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, *19*(2), 87–98.

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces that Shape Our Decisions*. New York: HarperCollins Publishers.

Aristotle (1984). Nicomachean ethics. In J. Barnes (ed.), *The Complete Works of Aristotle, Vol. II* (pp. 1729-1867). Princeton: Princeton University Press.

Badger, G. J., Bickel, W. K., Giordano, L. A., Jacobs, E. A., Loewenstein, G., & Marsch, L. (2007). Altered states: The impact of immediate craving on the valuation of current and future opioids. *Journal of Health Economics*, *26*(5), 865–876.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*(6), 1014.

Baron-Cohen, S., Leslie, A., & Frith, U. (2003). Does the autistic child have a 'theory of mind'? *Cognition*, *21*(1), 37-46.

Bate, K. S., Malouff, J. M., Thorsteinsson, E. T., & Bhullar, N. (2011). The efficacy of habit reversal therapy for tics, habit disorders, and stuttering: A meta-analytic review. *Clinical Psychology Review*, *31*(5), 865–871.

Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry*, *2*(2), 107–122.

Baumeister, R., Bratslavsky, E., Muraven, M., & Tice, D. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252-1265.

Bayer, U. C., Achtziger, A., Gollwitzer, P. M., & Moskowitz, G. B. (2009). Responding to subliminal cues: do if-then plans facilitate action preparation and initiation without conscious intent? *Social Cognition*, *27*(2), 183–201.

Bechara, A., Damasio, A., & Damasio, H. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7-15.

Berridge, K. (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*, 391-431.

Bijleveld, E., Custers, R., & Aarts, H. (2010). Unconscious reward cues increase invested effort, but do not change speed–accuracy tradeoffs. *Cognition*, *115*(2), 330–335.

Bijleveld, E., Custers, R., & Aarts, H. (2011). Once the money is in sight: Distinctive effects of conscious and unconscious rewards on task performance. *Journal of Experimental Social Psychology*, *47*(4), 865–869.

Bowen, S., & Marlatt, A. (2009). Surfing the urge: Brief mindfulness-based intervention for college student smokers. *Psychology of Addictive Behaviors*, *23*(4), 666–671.

Bowen, S., Chawla, N., Collins, S. E., Witkiewitz, K., Hsu, S., Grow, J., et al. (2009). Mindfulness-based relapse prevetion for substance use disorders: A pilot efficacy trial. *Substance Abuse*, *30*(4), 295–305.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology*, *81*(5), 946–960.

Brewer, J. A., Elwafi, H. M., & Davis, J. H. (2012). Craving to quit: Psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. *Psychology of Addictive Behaviors*, *27*(2), 366-379.

Brewer, J. A., Mallik, S., Babuscio, T. A., Nich, C., Johnson, H. E., Deleone, C. M., et al. (2011). Mindfulness training for smoking cessation: Results from a randomized controlled trial. *Drug and Alcohol Dependence*, *119*(1-2), 72–80.

Cesario, J., Plaks, J. E., & Higgins, E. T. (2006). Automatic social behavior as motivated preparation to interact. *Journal of Personality and Social Psychology, 90*, 893-910.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92.

Cohen, A.-L., Bayer, U. C., Jaudas, A., & Gollwitzer, P. M. (2006). Self-regulatory strategy and executive control: implementation intentions modulate task switching and Simon task performance. *Psychological Research*, *72*(1), 12–26.

Custers, R., & Aarts, H. (2005a). Beyond priming effects: The role of positive affect and discrepancies in implicit processes of motivation and goal pursuit. *European Review of Social Psychology*, *16*(1), 257–300.

Custers, R., & Aarts, H. (2005b). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology*, *89*(2), 129–142.

Custers, R., & Aarts, H. (2007a). In search of the nonconscious sources of goal pursuit: Accessibility and positive affective valence of the goal state. *Journal of Experimental Social Psychology*, *43*(2), 312–318.

Custers, R., & Aarts, H. (2007b). Goal-discrepant situations prime goal-directed actions if goals are temporarily or chronically accessible. *Personality and Social Psychology Bulletin, 33*(5), 623-633.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Penguin Putnam.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral Brain Research*, *41*(2), 81–94.

Danner, U. N., Aarts, H., & Vries, N. K. (2010). Habit vs. intention in the prediction of future behavior: The role of frequency, context stability and mental accessibility of past behavior. *British Journal of Social Psychology*, *47*(2), 245–265.

Denzler, M., Förster, J., & Liberman, N. (2009). How goal-fulfillment decreases aggression. *Journal of Experimental Social Psychology*, *45*(1), 90–100.

DeWall, C., Baumeister, R., Gailliot, M., & Maner, J. (2008). Depletion makes the heart grow less helpful: Helping as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin, 34*, 1653-1662.

DeWall, C., Baumeister, R., Stillman, T., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology*, *43*(1), 62-76.

DiLeone, R. J., Taylor, J. R., and Picciotto, M. R. (2012). The drive to eat: Comparisons and distinctions between mechanisms of food reward and drug addiction. *Nature Neuroscience 15*(10), 1330-1335.

Duchaine, B. C., Yovel, G., Butterworth, E. J., & Nakayama, K. (2006). Prosopagnosia as an impairment to face-specific mechanisms: Elimination of the alternative hypotheses in a developmental case. *Cognitive Neuropsychology*, *23*(5), 714–747.

Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., & Gollwitzer, P. M. (2011). Self-regulation strategies improve self-discipline in adolescents: benefits of mental contrasting and implementation intentions. *Educational Psychology*, *31*(1), 17–26.

Finkel, E.J. and Campbell, W.K. (2001). Self-control and accommodation in close relationship: an interdependence analysis. *Journal of Personality and Social Psychology, 81*(2), 263-277.

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification*, *31*(6), 772–799.

Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review*, *11*(3), 211–233.

Förster, J., Liberman, N., & Higgins, E. T. (2005). Accessibility from active and fulfilled goals. *Journal of Experimental Social Psychology*, *41*(3), 220–239.

Gailliot, M., & Baumeister, R. (2007). Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior. *Personality and Social Psychology Bulletin, 33*(2), 173-186.

Gawrilow, C., & Gollwitzer, P. M. (2007). Implementation intentions facilitate response inhibition in children with ADHD. *Cognitive Therapy and Research*, *32*(2), 261–280.

Gawrilow, C., Morgenroth, K., Schultz, R., Oettingen, G., & Gollwitzer, P. M. (2012). Mental contrasting with implementation intentions enhances self-regulation of goal pursuit in schoolchildren at risk for ADHD. *Motivation and Emotion*, *37*(1), 134-145.

Gilbert, D. T., Gill, M. J., & Wilson, T. D. (2002). The future is now: Temporal correction in affective forecasting. *Organizational Behavior and Human Decision Processes*, *88*(1), 430–444.

Gollwitzer, P. M. (1990). Action phases and mind-sets. In E. T. Higgins and R. M. Sorrentino (eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior*, *Volume II* (pp. 53–92). New York: The Guilford Press.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: the importance of implementation intentions. *Personality and Social Psychology Review*, *2*(2), 124–136.

Gollwitzer, P. M., & sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, *38*, 69–119.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*(4), 495–525.

Hassin, R. R., Bargh, J. A., & Zimerman, S. (2009). Automatic and flexible: The case of non-conscious goal pursuit. *Social Cognition*, *27*(1), 20-36.

Hendrickson, K. L., & Rasmussen, E. B. (2013). Effects of mindful eating training on delay and probability discounting for food and money in obese and healthy-weight individuals. *Behavior Research and Therapy*, *51*(7), 399–409.

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology, 43*(3), 497–504.

Holland, R. W., Aarts, H., & Langendam, D. (2006). Breaking and creating habits on the working floor: A field-experiment on the power of implementation intentions. *Journal of Experimental Social Psychology*, *42*(6), 776–783.

Holton, R. (2009). *Willing, Wanting, Waiting*. New York: Oxford University Press.

Holton, R. & Berridge, K. C. (2013). Addiction between compulsion and choice. In N. Levy (ed.), *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience* (pp. 239-268). New York: Oxford University Press.

Houssais, S., Oettingen, G., & Mayer, D. (2013). Using mental contrasting with implementation intentions to self-regulate insecurity-based behaviors in relationships. *Motivation and Emotion*, *37*(2), 224-233.

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*(5), 450–463.

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., & Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective

experience. *Emotion*, *10*(1), 54–64.

Johannessen, K. B., Oettingen, G., & Mayer, D. (2012). Mental contrasting of a dieting wish improves self-reported health behavior. *Psychology & Health*, *27*(sup2), 43–58.

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

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128.

Kappes, H. B., & Oettingen, G. (2011). Positive fantasies about idealized futures sap energy. *Journal of Experimental Social Psychology*, *47*(4), 719–729.

Kawada, C. L. K., Oettingen, G., Gollwitzer, P. M., & Bargh, J. A. (2004). The projection of implicit and explicit goals. *Journal of Personality and Social Psychology*, *86*(4), 545–559.

Ladouceur, R. (1979). Habit reversal treatment: learning an incompatible response or increasing the subject's awareness? *Behavior Research and Therapy, 17*(4), 313-316.

Leland, J., & Grafman, J. (2005). Experimental tests of the Somatic Marker hypothesis. *Games and Economic Behavior*, *52*(2), 386–409.

Lillis, J., Hayes, S. C., Bunting, K., & Masuda, A. (2009). Teaching acceptance and mindfulness to improve the lives of the obese: A preliminary test of a theoretical model. *Annals of Behavioral Medicine*, *37*(1), 58-69.

Marchiori, D., & Papies, E. K. (2014). A brief mindfulness intervention reduces unhealthy eating when hungry, but not the portion size effect. *Appetite*, *75*, 40–45.

Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science*, *19*(3), 255–260.

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45(3), 594-597.

Moore, A., & Malinowski, P. (2009). Meditation, mindfulness and cognitive flexibility. *Consciousness and Cognition*, *18*(1), 176–186. doi:10.1016/j.concog.2008.12.008

Moretti, L., Dragone, D., & di Pellegrino, G. (2009). Reward and social valuation deficits following ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *21*(1), 128–140.

Muraven, M., Collins, R. L., & Neinhaus, K. (2002). Self-control and alcohol restraint: An initial application of the Self-Control Strength Model. *Psychology of Addictive Behaviors*, *16*(2), 113–120.

Neal, D. T. & Wood, W. (2010). Automaticity in situ and in the lab: the nature of habit in daily life. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (eds.), *Oxford Handbook of Human Action* (pp. 442-457). New York: Oxford University Press.

Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits—A repeat performance. *Current Directions in Psychological Science*, *15*(4), 198–202.

Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin*, *37*(11), 1428–1437.

O'Reilly, G. A., Cook, L., Spruijt-Metz, D., & Black, D. S. (2014). Mindfulness-based interventions for obesity-related eating behaviors: a literature review. *Obesity Reviews*, *15*(6), 453–461.

Oettingen, G. (2012). Future thought and behavior change. *European Review of Social Psychology*, *23*(1), 1–63.

Oettingen, G., Hönig, G., & Gollwitzer, P. M. (2000). Effective self-regulation of goal attainment. *International Journal of Educational Research*, *33*(7), 705–732.

Oettingen, G., Mayer, D., & Thorpe, J. (2010a). Self-regulation of commitment to reduce cigarette consumption: Mental contrasting of future with reality. *Psychology & Health*, *25*(8), 961–977.

Oettingen, G., Mayer, D., Timur Sevincer, A., Stephens, E. J., Pak, H. J., & Hagenah, M. (2009). Mental contrasting and goal commitment: The mediating role of energization. *Personality and Social Psychology Bulletin*, *35*(5), 608–622.

Oettingen, G., Pak, H., & Schnetter, K. (2001). Self-regulation of goal setting: turning free fantasies about the future into binding goals. *Journal of Personality and Social Psychology*, *80*(5), 736–753.

Oettingen, G., Stephens, E. J., Mayer, D., & Brinkmann, B. (2010b). Mental contrasting and the self-regulation of helping relations. *Social Cognition*, *28*(4), 490–508.

Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology*, *29*(4), 374–383.

Plato. (1997). Republic. In J. M. Cooper & D. S. Hutchinson (Eds.), *Plato: Complete Works* (pp. 971-1224). Indianapolis: Hackett Publishing Company.

Quinn, J. M., Pascoe, A., Wood, W., & Neal, D. T. (2010). Can't control yourself? Monitor those bad habits. *Personality and Social Psychology Bulletin*, *36*(4), 499–511.

Railton, P. (2012) That obscure object, desire. *Proceedings and Addresses of the American Philosophical Association, 86*(2), 22-46.

Robinson, S., Sandstrom, S. M., Denenberg, V. H., & Palmiter, R. D. (2005). Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. *Behavioral Neuroscience*, *119*(1), 5–15.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews*, *18*(3), 247–291.

Robinson, T. E., & Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction*, *96*(1), 103–114.

Robinson, T., & Berridge, K. (2008). The incentive sensitization theory of addiction: some current issues. *Philosophical Transactions of the Royal Society, 363*(1507), 3137-3146.

Rorty, A. O. (1980). Where does the akratic break take place? *Australasian Journal of Philosophy, 58*(4), 333-346.

Rothermund, K. (2003). Automatic vigilance for task-related information: Perseverance after failure and inhibition after success. *Memory & Cognition*, *31*(3), 343–352.

Saver, J.L. and Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia, 29*(12), 1241-1249.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge: Belknap Press.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*(1), 33–46.

Schoenbaum, G. and Roesch, M. (2005) Orbitofrontal cortex, associative learning, and expectancies. *Neuron, 47*, 633–636.

Schroeder, T. & Arpaly, N. (2013). Addiction and blameworthiness. In N. Levy (ed.), *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience* (pp. 214-238). New York: Oxford University Press.

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology*, *83*(6), 1261–1280.

Shamay-Tsoory, S.G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia, 45*, 3054–3067.

Sheeran, P., & Orbell, S. (2000). Using implementation intentions to increase attendance for cervical cancer screening. *Health Psychology*, *19*(3), 283–289.

Sripada, C. S. (2014) How is willpower possible? The puzzle of synchronic self-control and the divided mind. *Nous, 48*(1), 41-74.

Stadler, G., Oettingen, G., & Gollwitzer, P. M. (2009). Physical activity in women: Effects of a self-regulation intervention. *American Journal of Preventive Medicine*, *36*(1), 29–34.

Tapper, K., Shaw, C., Ilsley, J., Hill, A. J., Bond, F. W., & Moore, L. (2009). Exploratory randomised controlled trial of a mindfulness-based weight loss intervention for women. *Appetite*, *52*(2), 396–404.

Tranel, D., Damasio, H., Denburg, N.L., & Bechara, A. (2005). Does gender play a role in functional asymmetry of ventromedial prefrontal cortex? *Brain, 128*, 2872-2881.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Theoretical Criminology*, *29*(9), 1159–1168.

Verplanken, B., & Faes, S. (1999). Good intentions, bad habits, and effects of forming implementation intentions on healthy eating. *European Journal of Social Psychology*, *29*(56), 591–604.

Vidrine, J. I., Businelle, M. S., Cinciripini, P., Li, Y., Marcus, M. T., Waters, A. J., et al. (2009). Associations of mindfulness with nicotine dependence, withdrawal, and agency. *Substance Abuse*, *30*(4), 318–327.

Vohs, K. D., & Faber, R. J. (2007). Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of Consumer Research*, *33*(4), 537–547.

Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources. *Journal of Personality and Social Psychology*, *88*(4), 632–657.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, *94*(5), 883-898.

Wallace, R. J. (1999). Addiction as defect of the will: Some philosophical reflections. *Law and Philosophy*, *18*(6), 621–654.

Watson, G. (1975). Free agency. *The Journal of Philosophy*, *72*(8), 205–220.

Watson, G. (1999). Disordered appetites: addiction, compulsion, and dependence. In Jon Elster (ed.), *Addiction: Entries and Exits* (pp. 3-28). New York: Russell Sage Foundation.

Webb, T. L., & Sheeran, P. (2003). Can implementation intentions help to overcome ego-depletion? *Journal of Experimental Social Psychology*, *39*(3), 279–286.

Webb, T. L., & Sheeran, P. (2007). How do implementation intentions promote goal attainment? A test of component processes. *Journal of Experimental Social Psychology*, *43*(2), 295–302.

Webb, T. L., & Sheeran, P. (2010). Mechanisms of implementation intention effects: The role of

goal intentions, self-efficacy, and accessibility of plan components. *British Journal of Social Psychology*, *47*(3), 373–395.

Webb, T. L., Sheeran, P., & Luszczynska, A. (2010). Planning to break unwanted habits: Habit strength moderates implementation intention effects on behavior change. *British Journal of Social Psychology*, *48*(3), 507–523.

Westbrook, C., Creswell, J. D., Tabibnia, G., Julson, E., Kober, H., & Tindle, H. A. (2011). Mindful attention reduces neural and self-reported cue-induced craving in smokers. *Social Cognitive and Affective Neuroscience*, *8*(1), 73-84.

Wheeler, S. C., Briñol, P., & Hermann, A. D. (2007). Resistance to persuasion as self-regulation: Ego-depletion and its effects on attitude change processes. *Journal of Experimental Social Psychology*, *43*(1), 150–156.

Witkiewitz, K., Marlatt, G. A., & Walker, D. (2005). Mindfulness-based relapse prevention for alcohol and substance use disorders. *Journal of Cognitive Psychotherapy*, *19*(3), 211–228.

Wyvell, C. L., & Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward "wanting" without enhanced 'liking' or response reinforcement. *Journal of Neuroscience*, *20*(21), 8122–8130.

Wyvell, C. L., & Berridge, K. C. (2001). Incentive sensitization by previous amphetamine exposure: Increased cue-triggered "wanting" for sucrose reward. *Journal of Neuroscience*, *21*(19), 7831–7840.

Zeidan, F., Johnson, S. K., Diamond, B. J., David, Z., & Goolkasian, P. (2010). Mindfulness meditation improves cognition: Evidence of brief mental training. *Consciousness and Cognition*, *19*(2), 597–605.

Zgierska, A., Rabago, D., Zuelsdorff, M., Coe, C., Miller, M., & Fleming, M. (2008). Mindfulness meditation for alcohol relapse prevention: a feasibility pilot study. *Journal of Addiction Medicine*, *2*(3), 165–173.

Zimmermann, T. D., & Meier, B. (2009). The effect of implementation intentions on prospective memory performance across the lifespan. *Applied Cognitive Psychology*, *24*(5), 645–658.