

To appear in N. Heather and G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship*, Oxford: Oxford University Press.

Addiction, Compulsion, and Weakness of the Will: A Dual-Process Perspective

Edmund Henden

Abstract

How should addictive behavior be explained? In terms of neurobiological illness and compulsion, or as a choice made freely, even rationally, in the face of harmful social or psychological circumstances? Some of the disagreement between proponents of the prevailing medical models and choice models in the science of addiction centres on the notion of “loss of control” as a normative characterization of addiction. In this article I examine two of the standard interpretations of loss of control in addiction, one according to which addicts have lost free will, the other according to which their will is weak. I argue that both interpretations are mistaken and propose therefore an alternative based on a dual-process approach. This alternative neither rules out a capacity in addicts to rationally choose to engage in drug-oriented behavior, nor the possibility that addictive behavior can be compulsive and depend upon harmful changes in their brains caused by the regular use of drugs.

1 Introduction

Addiction is a paradoxical phenomenon. On the one hand, addicts seem knowingly to expose themselves to harm, including emotional distress, legal and financial problems, health problems and so on. Many report feeling miserable and wanting to quit. They are commonly thought of as having lost control over their drug-taking behavior. On the other hand, many addicts seem to be susceptible to a wide range of ordinary incentives, including money, and counter-incentives such as the risk of harm, suggesting that they do in fact exercise a substantial degree of control over this same behavior. These apparently conflicting observations are frequently used by adherents of what are held to be the two main opposing models of addiction. The “medical model” sees addictive behavior as compulsive and symptomatic of neurobiological illness. The “choice model” sees it as the manifestation of harmful social or psychological circumstances that have led addicts to freely and, in a certain sense, rationally choose to engage in this behavior.

What exactly is it that the proponents of these two models disagree about? Some of their disagreement concerns the correct normative characterization of addiction where the bone of contention is the notion of “loss of control.” Loss of control refers to a highly heterogeneous phenomenon for which a precise definition is difficult to give, the trouble being that the empirical evidence does not seem to favour one normative characterization more than any other. In this article I want to suggest a way of thinking about addiction that resolves the apparent paradox. If correct, there will be no need for any normative conflict between medical models and choice models of addiction.

The structure of the article is as follows: in the following two sections I examine two alternative interpretations of addicts’ control problem. According to the first, they have lost free will; according to the other, their will is weak. Contrary to what many have claimed, I argue that addicts typically neither have lost free will nor display ordinary weakness of will. In the subsequent section I propose an alternative interpretation of addicts’ control problem that is based on a dual-process approach. According to this proposal, addicts’ lack of control stems from a certain sort of malfunctioning of their will.¹ This sort of malfunctioning does not rule out the capacity in addicts to act freely and, in a certain sense, rationally choose to engage in drug-oriented behavior, nor does it rule out the possibility that this behavior can be compulsive and depend upon harmful changes in their brains caused by the regular use of drugs.

2 Addiction and Freedom of the Will

What does it mean to “lose control of one’s actions”? It certainly means not doing what one, in some way, “wants to do”, i.e., where what one wants to do expresses the content of one’s will. Saying of someone that she has “lost control of her actions” is, therefore, to say that she is acting against her own will. Very generally, the term “will” refers to a cognitive state that has some action as part of its content, as well as an executive capacity to bring the world into conformity with that content.² It is common to distinguish, broadly, between two different senses in which persons may be said to “act against their own will.” They can do so in the sense that their will is too weak to sustain the performance of what they want to do. Such persons are

¹ The sort of view I am proposing is, of course, not new. For different versions, see e.g., Redish, Jensen, and Johnson (2008), Holton and Berridge (2013), Schroeder and Arpaly (2013).

² There are different views of exactly what sort of capacity this involves. According to one common view, it can be characterized, roughly, as the capacity to choose and to act in the light of reasons (see e.g., Raz 1997). This will be the view I shall assume in what follows.

said to suffer from “weakness of the will.” But they can also act against their will in the sense of being compelled by external forces to do other than they want. Such persons are said to lack “freedom of the will.” In both cases, although the person fails to express their will in action, the normative implications are very different. In the former, the person is held to be blameworthy (or at least rationally criticizable), in the latter she is not. That is because it is only in the latter case that persons are held to have been prevented from translating the content of their will into action. Much of the philosophical discussion about addiction has consequently revolved around determining to which of these normative categories addicts’ control problems belong. That is, does their loss of control typify weakness of will or unfreedom of the will? My own view, which I shall argue for in a moment, is that it typifies neither – at least not as they are ordinarily understood. Let me start with the claim that addicts’ control problem consists in a loss of freedom of the will.

2.1 Addiction as Loss of Free Will: Irresistible Desires and Rational Incapacity

According to a plausible view, one (at least) necessary condition of a person having free will with respect to a particular action at a given time is that she has the capacity to refrain from that action at that time. If addicts’ control problem is lack of free will, it follows that addicts typically lack the capacity to refrain from drug-oriented behavior. Those who accept this claim tend to assume that what prevents addicts from exercising their will is that their addiction creates such an overwhelming desire for drugs as to be literally irresistible.³ To say that a person’s desire is “irresistible” is to say that she is *unable* to resist acting on it. This is commonly interpreted as meaning that she has lost a rational capacity – broadly construed, the capacity to respond to reasons of a certain type (Fisher and Ravizza 1998). Given the plausible view that capacities are general or multi-track in nature – they don’t simply manifest themselves as single possibilities, but rather as whole rafts of possibilities – a person can be said to have lost such a capacity if a whole host of relevantly similar counterfactuals are true of her (Smith 2003). Consider, for example, a cocaine addict. What makes it the case that she has lost the capacity to resist cocaine? According to the reasons-responsiveness view, it would be true to say of her that she would not respond to reasons to refrain from cocaine – cocaine of ever so slightly different kinds, in ever so slightly different circumstances – in a suitable context of relevantly similar possible worlds in which she was given what she took to be good and sufficient reasons

³ For a classic statement of this view, see Frankfurt (1971).

to refrain from cocaine. In short, addicts lack free will with respect to their drug-oriented choices and actions because they have lost the capacity to respond to reasons to abstain.

Is there any evidence that addicts suffer from this sort of rational incapacity? No. On the contrary, there appears to be plenty of evidence that they do not. The evidence demonstrates that a wide range of counterfactuals of the sort just mentioned is actually false of many addicts. Several studies suggest, for example, that addictive behavior varies as a function of costs, benefits and cultural values, and many addicts appear to quit successfully unaided when they reach their early thirties (Heyman 2009; Pickard 2013). One particularly interesting illustration of addicts' capacity to control their drug-oriented motivation is provided by a form of behavioral therapy known as contingency management (CM) treatment. In CM, every time a desired behavior occurs, tangible and immediate reinforcers are applied. When it does not occur, they are withdrawn (see Petry et al. 2011). When the method is applied to substance disorders, they commonly use an exchange system. Patients receive vouchers for metabolic evidence of drug abstinence. Every time an addict submits a scheduled urine sample (typically three times a week) that tests negative for drugs, she earns a voucher that can then be traded in for desirable but inexpensive goods, such as restaurant gift certificates, clothing, or electronics. The value of the vouchers increases with each consecutive instance of proven drug abstinence. Conversely, if the patient uses drugs or fails to submit a scheduled sample, the value of the voucher is reset to the starting point. In combination with counseling, CM has proved a surprisingly effective method (Heyman 2009). Several studies have shown that, compared with control groups receiving traditional psychological counseling, a significantly larger number of subjects in the voucher groups remain abstinent during and after CM treatment.

Now, it is reasonable to infer that CM treatment must impact on addicts' capacity for reasons-responsiveness, either by strengthening their motivation to exercise this capacity or by strengthening the capacity itself – or perhaps by some combination of the two. For present purposes, what is interesting about it is that it suggests that some capacity must be present also *before* treatment with CM. In fact, voluntarily submitting to CM treatment and complying with the therapist's instructions seems evidence of such capacity. The effectiveness of CM, in other words, appears difficult to explain unless one assumes that many addicts do in fact possess a rational capacity, at least to some extent. If this capacity is diminished as a result of drug abuse, it would presumably explain why they find it difficult to abstain. But if many addicts typically possess rational capacity, it is not plausible that addiction typically causes a loss of free will with respect to drug-related choices and actions. The view that addicts necessarily lack free will is therefore likely to be false.

Should the control problem be conceptualized then in terms of ordinary weakness of will? Some have argued that it should (e.g., Benn 2007). In section 2.2 I present two different versions of this view (corresponding to two different accounts of weakness of will) and argue that both are mistaken.

2.2 Addiction and Weakness of the Will

The claim that addiction is a species of weakness of the will can be interpreted in two ways, corresponding to two different views of the nature of weakness of will. According to what, perhaps, has become the dominant view, weakness of will refers to the violation of a “resolution,” where the resolution is a kind of decision, intention, plan, or policy that is formed precisely in order to remain firm in the face of contrary desires one expects to arise when the time to act comes (Holton 1999). According to this view, weakness of will is manifested in an over-readiness to abandon one’s resolutions for exactly the type of reasons they were meant to overcome. If addicts typically are weak-willed in this sense, their control problem consists in a persistent lack of resolve with respect to abstinent behavior. So, for illustration, imagine an addict who now forms a resolution to abstain from taking drugs in the future. As the opportunity for consumption draws near, she succumbs to the temptation to rationalize her reluctance to abstain by changing her mind about what would be best, for example, by giving too much weight to certain considerations that appear to provide reasons for consumption, such as the immediate pleasures of the drug, and then revises her resolution to abstain accordingly, usually regretting it afterwards. This sort of weakness involves diachronic conflict, the irrationality of which is displayed in a form of incoherence between her motivational state at the time of action and her long-term attitudes toward these states.

However, not all agree that weakness of will must involve violation of a resolution. The traditional view, dating back to Aristotle, is that weakness of will is the same as “akrasia,” i.e., intentional action freely performed against the person’s consciously held judgement that, on the whole, it would be best not to perform the action (Mele 2009). According to this view, weakness of will is expressed in a failure to comply with one’s own best judgments while still holding them. It involves, in other words, a synchronic conflict, the irrationality of which is displayed in a form of incoherence in the person’s attitudes at the time of action. If addicts are weak-willed in this sense, they typically might not be forming resolutions to give up their addictions. Rather, their control problem might consist, precisely, in their failure to form such resolutions. Typically, what they find hard might be to decide to do or commit themselves to doing what their own judgment is telling them would be the best option all things considered. There need

be no sense, therefore, of them constantly changing their minds about what *is* best. Rather, their problem is that they cannot make up their minds *to actually do it*. Their defect consists in a failure to comply with their best judgment.

Before proceeding, let me add a brief remark about these two views, as there has been some controversy about which of them truly captures the nature of weakness of will (Holton 1999; Mele 2009; May and Holton 2012). I shall assume, for present purposes, that there can be both akratic and non-akratic forms of weakness of will. I do so because it seems to me that insofar as we care about this phenomenon at all, we care about it because of its normative significance, because weak-willed people are held to be rationally criticizable (or blameworthy) for succumbing to temptation.⁴ Thus, it is widely agreed to be reasonable to expect or demand that persons displaying weakness of will resist acting on their contrary desires – something that is commonly taken to imply the possession (unlike persons lacking in free will) of capacities of resistance.⁵ In other words, weakness of will is displayed by persons insofar as they are criticizable (or blameworthy) for making insufficient effort to exercise capacities they are believed to possess in order to do what they in some sense want to do. If the importance of weakness of will – at least for most practical purposes – is associated with its normative characterization, I think it would be a reasonable hypothesis to assume that the ordinary notion of weak-willed action corresponds to a more general notion of action lacking in self-control due to insufficient effort rather than to any of the more technical definitions discussed in the philosophical literature.⁶ This means, of course, that weakness of will is displayed both by persons who exert insufficient effort to make up their mind about what to do, i.e. to make decisions, form resolutions or commitments, as well as by persons who exert insufficient effort to stick to such decisions, resolutions or commitments in the face of temptation. One might wonder, though, why it matters whether addicts typically display weakness of will in the akratic or non-akratic sense. There is, however, a practical reason why it might matter: helping addicts overcome a tendency to unreasonably revise their resolutions to give up drugs may seem to require a different kind of approach than helping them make such resolutions in the first place.

⁴ By “we” here I mean us ordinary folk. I take weakness of will to be, at its core, a folk-psychological notion. There may still, of course, be other reasons why someone might care about it (e.g., philosophical or scientific reasons).

⁵ For a different view, however, see Watson (2004).

⁶ For empirical evidence that the folk notion of weakness of will may not be identical with any one of the two standard definitions in the philosophical literature, see May and Holton (2012).

While the former requires helping them learn ways of avoiding drug-associated thoughts, cues, or situations that might lead to a reconsideration of their resolutions, the latter seems more a matter of strengthening their motivation to actually do what they themselves judge is best, that is, to support and encourage them to translate their best judgments into practical plans or commitments to maintain abstinence.

Now, deciding whether addiction typifies akratic or non-akratic weakness of will is obviously problematic because it is very difficult to assess whether addicts in general retain the judgment that it would have been better to abstain or, by the time they have embarked on their addictive behavior, whether they have revised it. Nevertheless, both alternatives seem to have some evidence in their favor. In sections 2.2.1 and 2.2.2 I shall present some of this evidence. However, I shall argue that even if addiction is frequently associated with both diachronic and synchronic forms of irrationality, it is still misleading to treat it as a case of ordinary weakness of will.

2.2.1 Addiction as Non-Akratic Weakness of the Will: Hyperbolic Discounting

The best evidence that addicts tend to be weak-willed in the (non-akratic) resolution-violating sense is that their discount rates tend not to remain constant over time. “Discount rate” is the rate at which they discount the utility of future rewards, such as, for example, the benefits of a drug-free life. There is plenty of research suggesting that addicts have a tendency to discount the utility of future rewards, not by a fixed proportion per period of time (exponentially) – which would have led to temporally stable preferences – but by a proportion that declines as the length of the delay increases (hyperbolically), leading to regular and systematic preference reversals (Ainslie 2001; Bickel and Marsch 2001). Consider an alcoholic who prefers before breakfast not to drink at tonight’s dinner party but still has a strong craving for alcohol. As the opportunity to drink draws closer in time and the prospect of drink begins to weigh more heavily with her, the rate at which she discounts abstinence rises rapidly relative to the rate at which she discounts consumption. When the opportunity to imbibe finally presents itself, her estimate of its utility outweighs her estimate of the utility of abstinence with the result that her preference reverses. It may seem plausible that such “hyperbolic discounting” leads to weakness of the will, that is, an unreasonable over-readiness to abandon one’s resolutions. This view depends

on the assumption that hyperbolic discounting causes addicts to change their judgement of what is best at the time of action, i.e. when the alcohol or drugs become available for consumption.⁷

Now, although addicts find it extremely hard to remain abstinent over the longer term (many of them probably due to hyperbolic discounting), the view that loss of control in addiction should be conceptualized in terms of non-akratic weakness of will faces a difficulty: in many cases addictive behavior does not seem to involve judgment shifts at the time of action. Three key observations support this: first, addictive behavior sometimes displays a strongly habitual element – a fact that seems important for understanding relapses among users who have quit and are attempting to remain abstinent (Schroeder and Arpaly 2013). Such relapses seem frequently triggered by environmental cues and often appear to occur without much conscious awareness of the behavior undertaken. Explaining them in terms of a shift in considered judgments about what it would be best to do, therefore seems implausible. Second, addicts sometimes seem to retain a strong sense of *the disvalue* of their drug-oriented behavior even as they are carrying it out. For instance, there is evidence suggesting that they sometimes make conscious and strenuous efforts to resist, even indeed while seeking or taking drugs. As Robert West (2006) puts it, “when the restraint fails, there is often (but not always) no sense of the addict having changed his mind and deciding to engage in the behavior as a positive step; rather the sense is of a failure to exert control followed by regret and a feeling of having let oneself down” (p. 133). But this seems difficult to explain on the assumption that addicts typically abandon their judgement that abstaining is best while satisfying their addiction. Third, addicts may continue to seek and take drugs even when they derive no pleasure from their consumption, even in the absence of withdrawal – even, in fact, when they are convinced that taking drugs is a disastrous course of action for them (Robinson and Berridge 1993). Once again, addicts would seem on this observation to retain at times their judgment of abstinence as the most valuable alternative while carrying out their drug-oriented behavior.

2.2.2 Addiction as Akratic Weakness of the Will: “Wanting” and “Liking”

⁷Levy (2006; 2014) defends a version this view. It is worth noting that if motivating preferences are revealed in actual behavior and a gap may obtain between such preferences and a person’s evaluative ranking of her alternatives, preferences can reverse in the absence of judgment shifts. The view that hyperbolic discounting causes weakness of the will depends, therefore, on the assumption that motivating preferences always track considered judgments. For a different view, see, e.g., Lowenstein (1999).

The above observations may be taken as supporting the view that addicts' control problem must be one of akratic rather than of non-akratic weakness of the will. Akratic weakness of will, remember, is action carried out contrary to the person's consciously held judgment of what it would be best to do all things considered. An initial objection against this view could be that there simply is no evidence showing that it is even possible intentionally to perform actions while at the same time consciously judging that it would be best to refrain. All we can rely on to support this possibility are introspective reports of the persons themselves and such reports are notoriously unreliable (Levy 2011a). But there is other evidence. Of particular interest for our present purposes is the series of neuroscientific experiments conducted by Terry E. Robinson and Kent C. Berridge (see e.g., Robinson and Berridge 1998). Based on these experiments, they distinguish two components of motivation that are mediated by different psychological processes and neural substrates, what they call "liking," which is associated with an affective or cognitive value, e.g., subjective pleasure, goodness or predicted utility; and "wanting," which is associated with incentive salience – the degree to which a stimulus is action-driving. While "liking" and "wanting" normally go together so that we "want" the things we "like" (e.g. the value associated with some environmental cue serves as a trigger to activate and direct "wanting"), Robinson and Berridge provide evidence that they actually come apart in addiction, often making addicts "want" things they do not "like." The possibility of a decoupling between "wanting" and "liking" supports the possibility of a dissociation between intentional action and best judgment; if "wanting" can make addicts perform actions contrary to what they "like," and if judging some action as the best option entails "liking" it, then this is evidence that it is possible to perform actions contrary to one's best judgment (see also Holton and Berridge 2013).

So, granting that intentional action contrary to one's best judgement is possible, should we, on the basis of the observations mentioned in the last section, conclude that addiction typically involves akratic rather than non-akratic weakness of will? I think there are reasons to resist that conclusion. But let me begin by addressing a worry some readers may have at this point. Why do we need to choose between the view of addiction as either a species of akratic or non-akratic weakness of will? Why can addicts not exhibit both forms of weakness of will? I think both questions are reasonable. Addicts are not all alike. They differ in circumstances, the drugs they use, their social and personal resources (e.g. abilities and motivation), and their beliefs about the value of their options. Moreover, as we have seen, there is credible evidence suggesting that they often violate their resolutions as well as that they sometimes act contrary to their best judgments. Addicts' control problem often appears, then, to involve diachronic as

well as synchronic forms of irrationality. Still, I think there is an important sense in which addiction and ordinary weakness of the will differ. To see this we need to go back to what is arguably the commonsense notion of weakness of will according to which persons who try too little to exercise capacities they are assumed to possess in order to do what they in some sense want to do are criticizable (or blameworthy). Now, what determines an effort's "insufficiency," and thereby criticizability, is a normative matter. Presumably, there are certain shared expectations and norms guiding what counts as sufficient effort in various contexts. Thus, when we observe people failing to do what they want to do, we seem to have ways of correctly answering questions such as: *did she try hard enough?* I have elsewhere argued that the notion of sufficient effort is normative in the sense of being relative to the level of effort which, other things being equal, would have been sufficient for a normal person successfully to perform an action of the same type if she was as strongly motivated to perform it (Henden 2013). By "normal person" I mean someone whose capacity for self-control equals that of the majority of adults, and whose motivational system is congruent. In other words, if a person, according to this view, fails to do what she wants to do because her effort to do so is insufficient relative to this ordinary standard, she can be said to display weakness of will. In such cases, she is not, as we may say, sincerely trying. Hence she is criticizable (or blameworthy) for her behavior.

This gives us, I think, a way of distinguishing addiction from ordinary weakness of the will. The reason is that many addicts who try to abstain appear to try very hard to give up drugs, and many of them *are* wholly sincere in their effort to quit. Nevertheless, they fail again and again. If this failure cannot be explained by lack of capacity, it must be explained by insufficient effort. However, even though the effort they put into it is insufficient relative to what is actually required of them to abstain, it might still count as sufficient relative to *ordinary standards*. That is, if a normal person had made the same effort in similar circumstances, it would seem reasonable to expect her to succeed. Assuming, then, that many addicts display a strong will to give up drugs, it would seem unfair to criticize them for weakness of will. Rather, they seem instead to face some extraordinary obstacle peculiar to them that evades most normal persons. One observation supporting this is that we seem inclined to consider addicts *much less* criticizable (or blameworthy) for their drug-oriented behavior than weak-willed persons for their weak-willed behavior. Consider, for example, a cocaine addict and a weak-willed non-addict both of whom take cocaine.⁸ Suppose both act contrary to their best judgements (or fail

⁸ On what grounds, it could be asked, can one distinguish a non-addicted user from an addicted one when their actual mental states and behavioral patterns are the same? The difference, I think

to keep to their resolutions). Their drug-oriented motivations, we will assume, were resistible at the time of action. Now, we are clearly more inclined to criticize the non-addict than the cocaine addict for their wrongful behavior. The reason seems obvious: while we assume it would be relatively easy for the non-addicted user to refrain from cocaine in this situation, it is, we assume, comparatively harder for the cocaine addict to do the same. This normative difference affects our attitudes towards the two cases, suggesting that our attitudes to addiction and ordinary weakness of will differ significantly. Admittedly, this observation alone is not conclusive evidence that addiction is not a species of weakness of will. Still, it gives us, I think, good reason to explore a different approach to the loss of control in addiction. It is to this approach I now turn.

3 Addiction and Compulsion

If we cannot conceive of addicts' loss of control as loss of free will nor as ordinary weakness of will, how *should* we conceive of it? To answer this question, we might begin by noting that an important feature of addictive behavior is the regularity with which it occurs. The property of "being addicted" refers to a certain kind of relation a person has, not to some isolated act of consumption, but to a pattern of behavior, enacted on a regular basis in characteristic circumstances, which the person finds extremely difficult to override by intentional effort. Explaining an action in terms of addiction usually involves seeing it as part of such a behavioral pattern (in fact, it is not clear it even makes sense to speak of one-off addictive actions).

Lack of control over behavioral patterns is, of course, also what tends to be emphasized in clinical descriptions of compulsive behavior. Such behavior is characterized as strongly cue-dependent in the sense that it is regularly triggered by certain situations, places or people associated with the type of behavior in question; there is a feeling of being driven again and again to behave in precisely that particular way (often in spite of oneself), and it is a common

it is plausible to say, resides in certain counterfactuals being true when the person is an addicted and false when he is a non-addicted user. For example, it would be true of the addict but false of the non-addict that, as his supply of drugs wanes, he would begin obsessing over them, perhaps to the point of making an extraordinary effort to obtain them (often at great cost to himself). And were his drug use to become associated with displeasure, emotional distress, or health problems, it would be true of the addict but false of the non-addict that he would continue to consume the drug, often experiencing a physical compulsion to do so. See Skog (2003) for defense of a counterfactual definition of addiction.

experience that resistance, however sincere, becomes increasingly difficult over time. Since all these features are typically also present in addictive behavior, it seems reasonable to infer that addictive behavior is compulsive in the clinical sense. Indeed this is part of the standard medical definition of addiction (American Psychiatric Association 1994).⁹ If this is correct, it serves to further distinguish addiction from ordinary weakness of will since the latter is not definitionally tied to patterns of behavior exhibiting features of compulsivity. The question then becomes in what sense individuals who are exhibiting compulsive patterns of behavior are “acting against their own will.”

3.1 Attention, Bias, and the Will

One important aspect of the will concerns its relation to attention. That is because “executive control,” an umbrella term used in psychology for top-down cognitive processes that regulate, coordinate, and control other cognitive processes that are necessary for the initiation and monitoring of goal-directed actions (such as reasoning, planning or inhibition), involves directed attention, the capacity to voluntarily focus or shift attention. In fact, in some theories of executive control, they are not even clearly separated – executive control is treated as more or less identical to the mechanism controlling the deployment of attention (Miyake et al. 2000). Why this is relevant to our understanding of the loss of control in addiction is because there appears to be plenty of evidence that addicts’ attention is biased toward drug-associated stimuli (for a review, see e.g. Field and Cox 2008). An attentional bias is a certain sort of disruption of a person’s attentional selection process. This process is “biased” toward a particular kind of stimuli if the stimuli intrudes on her experience by capturing her attention and this capturing is part of a pattern and occurs in a systematic rather than random fashion. Attentional bias is, in other words, a statistical tendency or inclination to direct attention at a particular kind of stimuli (McKay & Efferson 2010).

Now, evidence for drug-related attentional bias in addiction is provided by a variety of experiments measuring implicit cognition. In the addiction Stroop task, for example, addicts are asked to name the color of drug-related words, and the time they take to name the colour of

⁹ It is worth noting that while, in the philosophical literature, the term “compulsive behavior” tends to be used to characterize behavior caused by “irresistible desires,” there is no assumption in the clinical literature that behavior must be caused by “irresistible desires” in order to count as compulsive. When I speak of “compulsion” in what follows, it is the clinical notion I shall have in mind. For a discussion, see Henden, Melberg and Røgeberg (2013).

these words is compared to those for drug-unrelated words (e.g., words related to musical instruments). The challenge is to focus attention on color while blocking out the words' meaning. Attentional bias is indexed as the difference between participants' mean color-naming reaction time in trials with drug-related and those with drug-unrelated words (Cox et al. 2006). Using this approach, researchers have found that while addicts exhibit significantly slower reaction times and are more prone to error when naming the color of drug-related words, control participants do not exhibit this pattern. The result has been demonstrated in users of a variety of different drugs, including alcohol, cannabis, cocaine, heroin, and tobacco. The standard interpretation of this "Stroop interference" is that, compared to non-addicts, drug-related words capture addicts' attention, causing excessive processing of the semantic content of these words, thereby disrupting their color naming.

Based on the addiction Stroop and other paradigms developed to measure attentional bias, many researchers believe attentional bias toward drug-associated stimuli plays an important part in explaining the maintenance or escalation of drug-oriented behavior, including relapses among users who have quit and are attempting to remain abstinent.¹⁰ Robinson and Berridge's influential theory, briefly discussed in section 2.2.2. explains why repeated drug use may lead to drug-related attentional bias. According to this theory, it is caused by a process they call "incentive-sensitization" in which repeated drug use produces a dopaminergic response that becomes sensitized (i.e. progressively larger) by making certain regions in the brain involved in the motivation of behavior more easily activated by drugs or drug-associated stimuli. As a consequence, these drug-associated stimuli acquire powerful "incentive properties" by drastically enhancing their capacity to grab the person's attention and be

¹⁰ One caveat is in order: there has been some debate in the psychological literature concerning the interpretation of the evidence from the addiction Stroop, and not all agree that it demonstrates a bias in addicts' attentional selection processes. Some have suggested, for example, that the delayed color naming could result from attempts to suppress the processing of drug-related words, or from a generic slowdown in cognitive processing as a consequence of experienced craving induced by the drug-related words rather than from a selective attention to those words (see e.g., Algom et al. 2004). The standard interpretation does, however, appear to be corroborated even when more direct measures of attentional bias are used, such as the monitoring of eye movements while the subject completes a visual probe task in which drug-related and control pictures are presented. For discussion, see Field and Cox (2008).

perceived as particularly salient. According to Robinson and Berridge, it is this process of incentive-sensitization that causes a decoupling of “wanting” (incentive salience) from “liking.”

In human addicts, drug-related attentional bias caused by incentive-sensitization presumably works through the human decision-making system, as evidenced in the ability of addicts, in general, to delay, alter and in some cases, substitute, their drug-oriented behavior based on deliberation. Typically, addicts plan when and how to obtain the drugs, taking all sorts of considerations into account. In order fully to explain how drug-related attentional bias produces drug-oriented behavior in humans, therefore, we need to know how the attentional bias interacts with the addicts’ decision-making system.¹¹ In section 3.1.1, I present in broad outline what I think might be a plausible view of this interaction.

3.1.1 Addiction as a Malfunctioning of the Will: A Dual-Process Perspective

It has become common in cognitive psychology to distinguish between two modes of decision-making; one fast, intuitive and effortless that is shaped by biology and implicit learning, the other slow, analytical and effortful, and shaped by culture and formal tuition.¹² While the former mode – often referred to as type-1 processes – depends on environmental cues, is associative, automatic, and can control behavior directly without the need for controlled attention, the latter – often referred to as type-2 processes – depends on de-contextualization, is rule-based, and requires controlled attention and effort.¹³ To achieve rational decision-making and reliably contribute to the person’s goal achievement, the two modes have to work well together. This requires, first, that the person’s type-2 processes can exert an executive function and override the impulsive output of her type-1 processes. For this to happen, her type-2 process must be able to generate a more considered response that is in line with her normative reasons, as well as involve inhibitory mechanisms to suppress the response tendencies of her type-1 processes.

¹¹ Incentive sensitization theory does not deny a role for top-down cognitive control. Robinson and Berridge are clear that whether or not a sensitized response is actually expressed may depend on contextual factors. See e.g., Robinson, Robinson and Berridge (2013).

¹² Parts of what follows are based on Henden, Melberg and Røgeberg (2013). A wide variety of evidence has converged on the conclusion that some sort of dual-process notion is needed to explain how the overall process of decision-making works. For a brief review, see Frankish and Evans (2009).

¹³ There is some disagreement on precisely how these processes should be characterized and distinguished. I cannot enter into this debate here. For discussion, see Stanovich (2009).

Second, the person's type-1 processes must be able to select adequate and relevant information about the practical situation as an input to her type-2 processes (Saunders and Over 2009).

The sort of top-down processes that are associated with executive control exemplify type-2 processes. They are inferentially integrated with the person's propositional attitude system and draw on all her background knowledge and beliefs. In contrast, the processes underpinning attentional bias are examples of type-1 processes. They consist of associative relations in memory between environmental cues and behavioral propensities – relations that can be activated during critical decision points without the person's conscious intention, deliberation, or even awareness. The latter processes are modular in the sense of involving highly specified mechanisms that are inferentially isolated from the person's propositional attitude system.

Now, assuming this dual-process model of decision-making, drug-related attentional biases in type-1 processes might be hypothesized to affect addicts' behavior in a variety of ways depending, most likely, on individual differences between addicts (e.g. differences in personal and social resources, abilities and motivation, type of drug used and so on). For example, one way might be by entering their type-2 processes and shaping their beliefs, desires, and reasoning about what to do. Thus, by persistently directing the person's attention to drug-associated features of their immediate physical and social environment, the processes might cause an over-appreciation of these features as well as blindness or indifference to longer-range goals (Dill and Holton 2014). The result could be that the rate at which these addicts discount the utility of future rewards, such as the benefits of abstinence, is increased drastically relative to the rate at which they discount consumption, leading to regular and systematic preference reversals (and perhaps judgement shifts) of the sort associated with hyperbolic discounting.

However, drug-associated attentional bias might also shape addicts' beliefs, desires, and reasoning *without* causing hyperbolic discounting. Not all addicts have unstable preferences. Some never make any effort to abstain from drugs because they have no desire to quit. Does that mean they are in full control of their drug-oriented behavior? I think most people would be inclined to say no. There is something odd about the idea of a person addicted to a drug being in "full control" of her behavior with respect to that drug. Simple reflection on the meaning of the notion of "being addicted" seems to speak against this possibility. Plausibly, addicts with stable preferences (who do not fight their addiction) will still show symptoms of compulsivity and obsession with respect to drug-seeking and drug-taking (e.g. drug-oriented considerations will always have precedence in their practical reasoning), or *would* show such symptoms if their drugs were to become unavailable or associated with an increase in negative costs (see footnote

8). Compulsivity and obsession – despite their superficial appearance of “too much control” – seem on a deeper level to indicate the opposite of control. The dual-process approach can explain in what sense control is lacking: even if some addicts have stable preferences, all their beliefs and desires will still be infused by drug-associated attentional bias; hence, by taking these beliefs and desires as inputs, their *practical reasoning* itself will in a sense be “out of control.”

But drug-related attentional bias in a type-1 process might also affect an addict’s behavior in a much more direct way. As we have seen, there is evidence that addicts sometimes retain their judgment that abstinence is the most valuable alternative even as they are carrying out their drug-oriented behavior. From a dual-process perspective, the problem in these kinds of cases is not that the type-2 processes are internally biased, but rather that they regularly fail to suppress responses generated by their biased type-1 processes despite their conflicting with considered normative responses generated by their type-2 processes. The Stroop effect serves as a good example of such a dual-process conflict: an automatic type-1 process directs the subject’s attention to one feature of the stimulus (word meaning), which disrupts a type-2 process (deciding the color of the word). Berridge and Robinson’s decoupling of “wanting” and “liking” is another example. As they describe “wantings” they are clearly modular. Not only are the processes that underpin them associative – and presumably operating at high speed and requiring low effort – they are implicit in the sense of being able to drive addicts’ behavior independently of their propositional attitude system (although in some cases they might operate by increasing their experienced cravings). It is reasonable to infer that “wantings” must be a kind of type-1 process. “Likings,” by contrast, seem underpinned by processes that work *through* addicts’ propositional attitude system. That is because they typically manifest themselves in cognitive-affective states such as desires, beliefs, and value judgments. Addicts who “want” to engage in drug-seeking and drug-taking without “liking” it can therefore plausibly be viewed as experiencing a dual-process conflict between a type-1 and a type-2 process.

Now the question was how drug-related attentional bias in a type-1 process might affect addicts’ behavior more directly. The answer, presumably, is that it upsets their *will* in some way. According to a theory that has gained wide acceptance in psychology, executive control functions all draw on the same limited resource. The more this resource is consumed, the more depleted it becomes and the poorer the person performs on subsequent tasks requiring executive control. This so-called “limited-resource model” of executive control is supported by numerous experiments showing that when people engage in executive control, later attempts at executive

control will be less successful (for a review, see Muraven and Baumeister 2000). Now the Stroop task is commonly used as a measure of executive control and cognitive depletion. The faster the participants name the colors of the words (which requires suppressing the initial tendency to name the meaning of these words), the higher their level of current executive control; the slower they are, the more depleted they are assumed to be. Since addicts perform poorly on the addiction Stroop compared to non-addicts, it is reasonable to infer that their executive functions are impaired by cognitive depletion. Plausibly, a drug-oriented type-1 process directing attention toward drug-associated features of the environment has become fixed – most likely due to incentive-sensitization – in entrenched dispositions and patterns of perception and response, thereby increasing the cognitive load on executive functions associated with type-2 processes requiring directed attention. A dramatic increase of cognitive load on directed attention means a corresponding increase in the consumption of executive control resources, which, given the limited supply of these resources, suggests they must detract from other executive functions. This impairs the performance of tasks involving inhibition, reasoning, or planning. Cognitive depletion, then, might explain why addicts regularly fail to override their type-1 processes in cases in which they experience a conflict between type-1 and type-2 processes (see also Levy 2013). In combination, attentional bias and cognitive depletion increase cognitive inflexibility, that is, they reduce addicts' capacity to switch their thinking and attention among different tasks, operations or practical perspectives in response to changing goals or circumstances. Their drug-related decision-making becomes, as a result, more stimulus-bound and less responsive to reasons to abstain from drugs. Since cognitive flexibility is plausibly part of a well-functioning will, drug-associated attentional bias can thus be said to disrupt addicts' will by drastically reducing cognitive flexibility and, in this sense, causing their will to malfunction.

4 Addiction between Illness and Choice

This account no doubt needs more development and defence than I can provide here, but assuming that it is on the right tracks, what are the implications for the supposed opposition between medical models and choice models in addiction science, and for the disagreement between those who claim addicts have lost free will and those who claim they suffer from weakness of will?

First, a malfunctioning of the will of the sort proposed neither implies that addicts' will must be unfree nor that it must be weak. Although cognitive inflexibility makes it much harder for addicts to revise or abandon their drug-oriented decision-making pattern, there is no reason

to assume it makes them unable, in general, to resist their desires for drugs. Unlike a loss of capacities of resistance (and hence free will), cognitive inflexibility might be offset, for example, by a sharp increase of intentional effort, i.e. by forcefully and actively redirecting attention towards drug-unrelated aspects of the situation, or to the normative reasons. Further, it seems hardly plausible that such malfunctioning constitutes ordinary weakness of will. By drastically reducing cognitive flexibility, drug-related attentional bias creates persistent obstacles to addicts' decision-making – obstacles requiring a sustained and extraordinary effort to overcome. Even addicts who exhibit a strong will to give up drugs frequently fail due to the difficulties of maintaining abstinence. Insofar as they do not seem criticizable for weakness of the will, it seems unfair to blame them for their lack of success.

Second, regarding the supposed opposition between medical models and choice models, a malfunctioning of the will of the sort proposed here neither rules out intention and choice in addictive behavior nor compulsion in the clinical sense. That addictive behavior patterns are sustained by cognitive inflexibility does not mean that addicts cannot *choose* to perform drug-oriented actions on the basis of their beliefs and desires. What causes their lack of control is the *shaping* of these beliefs and desires by processes that are completely dissociated from their propositional attitude system. Even in cases in which they perform drug-oriented actions contrary to what they desire, believe, or even judge is best, their actions can still be guided by intentions and they can retain the capacity to refrain from performing them. There is no reason, therefore, to assume that addicts do not intentionally engage in addictive behavior.

Finally, drug-associated attentional bias is underpinned by type-1 processes it is easy to imagine could create behavior patterns that exhibit features of compulsivity. The “compulsion” often reported by addicts might simply be the subjective experience of cue-triggered decision-making, that is, decision-making underpinned by associative relations in memory activated without their conscious intention, deliberation, or even awareness. If this is correct, what distinguishes compulsive from non-compulsive actions would not simply be the causal strength of their motivational antecedents (as is standardly assumed) – strong desires, for example, are felt by most normal persons from time to time – but rather their frequency, computational speed, cue-dependence and dissociated nature, features that together disrupt the normal functioning of their will and therefore make it extremely hard to maintain resistance over time.

5. Conclusion

In this article I have argued that a dual-process approach can explain the seemingly paradoxical features of addictive behavior, that is, its appearance of control and of non-control. It shows

how addiction can involve intentional actions that are freely performed, why addicts can make choices, as well as why addictive behavior still counts as, in an important sense, “compulsive” and “out of control.” It therefore resolves the apparent conflict between medical and choice models of addiction. Addiction, on this view, is a varied and multi-determined behavioral phenomenon. Many different factors – ranging from social and psychological factors shaping reasons and choice to biological changes disrupting the capacity for attentional control – are likely to affect its manifestations in individual cases, the difficulties facing the addict wanting to quit, and what are the best ways of helping her achieve that goal. But if the argument of this chapter is correct, then every addict can be said to suffer from a malfunctioning of the will brought about by the regular use of drugs.

References

- Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
- Algom, D and Chatjout, E and Lev, S. 2004. A rational look at the emotional Stroop phenomenon: a generic slowdown, not a Stroop effect. *J. Exp. Psychol. Gen* 133, 323-338.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental Disorders*. 4th ed. Washington, DC.
- Benn, P. (2007). Disease, addiction and the freedom to resist. *Philosophical Papers* 3, 465-481.
- Bickel, W. K., and Marsch, L.M. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction* 96, 73-86.
- Cox, W. M., Fadardi, J.S., and Pothos, E. M. (2006). The addiction-Stroop test: Theoretical considerations and procedural recommendations. *Psychological Bulletin* 132, no. 3, 443-476.
- Field, M. and Cox, W. M. (2008). Attentional bias in addictive behaviors: A review of its development, causes, and consequences. *Drug and Alcohol Dependence* 97, 1-20.
- Fisher, J.M., and Ravizza, M. (1998). *Responsibility and control: An essay on moral responsibility*. Cambridge: Cambridge University Press.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68, 5-20.
- Frankish, K., and Evans, J. St. B. T. The duality of mind: An historical perspective. In: J. St. B. T. Evans and K. Frankish, (eds.). (2009). *In Two Minds: Dual Processes and Beyond*.

- Oxford: New York, pp. 1-29.
- Henden, E. (2013) Addictive actions. *Philosophical Psychology* 26, no. 3, 362-382.
- Henden, E., Melberg, H.O., and Røgeberg, O.J. (2013). Addiction: Choice or compulsion? *Frontiers in Psychiatry* 4, no. 77, 1-11.
- Heyman, G.M. (2009). *Addiction, a Disorder of Choice*. Harvard University Press.
- Holton, R. (1999). Intention and weakness of will. *The Journal of Philosophy* 96, 241-262.
- Levy, N. (2006). Autonomy and addiction. *Canadian Journal of Philosophy* 36, 427-448.
- Levy, N. (2011). Resisting ‘weakness of the will’. *Philosophy and Phenomenological Research*, Vol. LXXXII No. 1, 134-155.
- Lowenstein, G. A visceral account of addiction. In: J. Elster and O-J. Skog (eds.) (1999). *Getting Hooked: Rationality and Addiction*. Cambridge: The MIT Press, pp. 235-264.
- May, J. and Holton, R. (2012). What in the world is weakness of will? *Philosophical Studies* 157, no. 3, 341-360.
- Mckay, R and Efferson C. (2010). The subtleties of error management. *Evolution and Human Behavior* 5, no. 31, 309-319.
- Mele, A. (2009). Weakness of will and akrasia. *Philosophical Studies* 150, no. 3, 391-404.
- Miyake, A., Friedman N.P., Emerson, M.J., et al. (2002). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology* 41, 49-100.
- Muraven, M., and Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin* 126, no. 2, 247–259.
- Petry, N. M., Sheila M. Alessi S.M., and Rash C.J.. Contingency management treatment of drug and alcohol use disorders. In: J. Poland and G. Graham (eds.) (2011). *Addiction and Responsibility*. Cambridge, Mass: MIT Press, pp. 225-245.
- Raz, J. Incommensurability and agency. In R. Chang (ed.) (1997). *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press, pp. 110-128.
- Redish, D. A., Jensen, S., and Johnson, A. (2008). A unified framework for addiction: vulnerabilities in the decision process.” *Behavioral and Brain Sciences* 31, 415-487.
- Robinson, T. E. and Berridge, K.C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews* 18, 252-255.

- Robinson, T. E., and Berridge, K.C. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews* 28, 309-369.
- Robinson, M.J.F., Robinson, T. E., and Berridge, K.C. (2013). Incentive salience and the transition to addiction. *Biological Research on Addiction*, Volume 2, 391-399.
- Saunders, C and Over, D.E.. In two minds about rationality. In: J. St. B. T. Evans and K. Frankish, (eds.). (2009). *In Two Minds: Dual Processes and Beyond*. Oxford: New York, pp. 317-334.
- Skog, O-J. Addiction: definitions and mechanisms. In: E. Vuchinich and N. Heather (eds.). (2003). *Choice, Behavioral Economics and Addiction*. Elsevier Ltd., pp.157-182.
- Smith, M. Rational capacities, or: How to distinguish recklessness, weakness, and compulsion. In: S. Stroud and C. Tappolet (eds.) (2003). *Weakness of Will and Practical Irrationality*. New York: Oxford University Press, pp. 17-38.
- Stanovich, K. E. Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In: J. St. B. T. Evans and K. Frankish, (eds.). (2009). *In Two Minds: Dual Processes and Beyond*. Oxford: New York, pp. 55-88.
- Watson, G. (2004). Skepticism about weakness of the will. *Agency and Answerability*. New York: Oxford University Press, pp. 33-58.
- West, Robert. *Theory of Addiction* (Oxford: Blackwell, 2006).