

COMPULSION AND CHOICE IN ADDICTION ¹

RICHARD HOLTON AND KENT BERRIDGE

Despite a wealth of recent empirical findings, the debate on addiction remains polarized along traditional lines. In one camp stand those who see the characteristic actions of the addict as driven by something very much like a disease: a pathologically intense compulsion that they can do nothing to resist. In a book published in 1812, Benjamin Rush quoted an alcoholic giving powerful expression to this approach:

Were a keg of rum in one corner of a room, and were a cannon constantly discharging balls between me and it, I could not refrain from passing before that cannon, in order to get at the rum. (Rush 1812, p. 266)

The example impressed Williams James, who quoted it and continued with the story of an alcoholic who supposedly chopped his hand off in order to be given brandy. (James 1890, Vol. II, p. 543). The understanding found its way into literature. Saying that he was ‘following what psychologists tell us’, Oscar Wilde described the lure of opium on Dorian Gray:

Men and women at such moments lose the freedom of their will. They move to their terrible end as automatons move. Choice is taken from them, and conscience is either killed, or, if it lives at all, lives but to give rebellion its fascination and disobedience its charm. (Wilde 1891, Ch. 16)

Modern expressions tend to be less dramatic, but the basic conception remains much the same. Many contemporary theorists insist that the addict is in the grip of a brain disease that removes control over their actions and so requires treatment rather than condemnation.

In the opposing camp stand those who see addictive behavior as involving ordinary choices, and so as something that takes place within the domain of ordinary intentional action. This approach sees an addict’s decision to take a drug as motivated by a standard structure of beliefs and desires. It harks back to an earlier understanding that saw alcoholics as simply over-fond of their drink, but in recent years it has received new impetus in the hands of certain economists and behavioral psychologists.²

Members of the disease camp point to the extraordinarily self-destructive behavior that addicts exhibit, and to the burgeoning literature that suggests that their brains are functioning in abnormal ways. Members of the ordinary choice camp point to findings that show that addicts often respond to incentives in normal ways. For example, most succeed in getting over

¹ This is a shortened and simplified version of (Holton and Berridge 2013). In addition to the people we thank there, we would like to thank the editors of the current volume for many helpful suggestions.

² The idea that addiction involved some kind of compulsion doesn’t really take hold until the 18th century. For discussion see (Sournia 1990).

their addictions by their mid-30s, often with minimal help.³ Further, many addicts beyond that age stop taking drugs if the incentives are great enough and clear enough. Anesthesiologists and airline pilots who, having been once detected in their addiction, are required to pass random and frequent drug tests on pain of dismissal, are remarkably good at giving up.

The two approaches are typically seen as quite incompatible. If addiction is a brain disease, then there is no role for willpower or self-control. To take a representative example, the book from a recent television series lists as one of the ‘seven myths of addiction’ the idea that ‘addiction is a willpower problem’, and goes on to say:

This is an old belief, probably based upon wanting to blame addicts for using drugs to excess. This myth is reinforced by the observation that most treatments for alcoholism and addiction are behavioral (talk) therapies, which are perceived to build self-control. But addiction occurs in an area of the brain called the mesolimbic dopamine system that is not under conscious control. (Hoffman and Froeke 2007, p. 37)

We agree that the mesolimbic dopamine system is centrally involved in addiction, and that the workings of that system do not appear to be under direct conscious control (in the sense that there doesn’t seem to be much that one can deliberately do to directly affect it). But it is one thing to say that people cannot control their mesolimbic dopamine system; quite another to say that they cannot control how it influences their actions. In a parallel way, there isn’t much that people can deliberately do to influence their perceptual systems, but that doesn’t mean that there is nothing they can do to control the effects of what they perceive on their actions.

Our contention here is that, once they are shorn of their claims to the whole truth, the disease approach and the choice approach are not so incompatible as has been made out. Our aim is to present the outlines of a middle path. The findings from brain science are solid enough. There is good evidence that the brain of an addict is importantly different from that of a normal non-addicted individual—indeed there is even some reason to think that the addict’s brain might have started out with a vulnerability to addiction. Certainly once addiction is under way, the desire for the addictive drug (amongst which we include alcohol and nicotine) takes on a life of its own, with an intensity that is particularly, perhaps uniquely, high. The desire becomes insulated from factors that, in normal intentional behavior, would undermine it, and so persists even when the addict knows that acting on it would be highly damaging. The addict may recognize that taking the drug again will incur the loss of family, friends, job, and most that makes life worth living, and yet still continue to take it. More surprisingly, addicts need not even like the thing that they are addicted to: they need gain no pleasure from it, nor anticipate that they will. Nor need they be motivated by a desire to avoid the horrors of withdrawal. Alcohol or heroin addicts often relapse long after withdrawal is over, and cocaine addiction is no less potent for having a relatively mild withdrawal syndrome. Addicts may relapse when they see nothing good in their drug whatsoever. They may see it as

³ This point is made very forcefully in (Heyman 2009, Ch. 4). He draws his conclusion from examination of national population surveys—not just surveys of addicts. He argues that most of those who remain addicted do so because they suffer from other psychiatric illnesses.

nasty, damaging and worthless in every respect. Yet they may still want it, and want it, moreover, in a particularly immediate and intense way—perhaps more immediately and more intensely than most other people ever experience.

Nevertheless, the intensity and power of an addictive desire does not mean that addicts are automata, powerless spectators moved by their desires. For whilst addictive desires are very strong, the human capacity for self-control is also highly developed; much more developed, it seems, than in rats. Addicts do not actually cross into the paths of cannonballs or their equivalents, despite William James' colorful assertion. They go around or wait for a lull. Smokers on airplanes postpone their urge to smoke until the flight is over.

As these cases show, addictive urges can be controlled, at least for a short while, and sometimes for longer if the stakes are high enough and clear enough. The experience of self-control that everyone has at certain moments is a veridical one: self-control is a real phenomenon, something that can be used to control acting on addictive desires, even if at a considerable cost and, for most addicts, subject to occasional failure. We should thus not be thinking of addictive desires as things that are impossible to resist, but as things that are very difficult to resist.⁴ Our moral evaluations should reflect this fact, and our scientific account of addiction should explain why resistance is difficult and why failure happens on the occasions it does.

Our aim here is to articulate such a model, one that explains why addictive desires have the distinctive features they have, but that also explains how they can be controlled. The account has two parts. The first explains what goes wrong in addiction, making the case for the incentive salience approach, which sees addiction as primarily a disorder of the wanting system. The second explains why, nevertheless, addicts remain agents. We start by outlining what we think is wrong with the pure choice model.

I. PURE CHOICE MODELS

We cannot hope to survey all of the different pure choice models here, but some brief comments will serve to show why we think that they cannot provide a complete explanation of addiction. A choice model can, of course, easily explain the behavior of those who willingly and knowingly take addictive drugs. But addicts frequently say that they have been somehow captured by the addiction—that they wish that they could escape it but that something is making it very hard for them to do so. Some listeners might dismiss these comments as disingenuous or self-deceived, but we think there is something in what they say.

How can choice models make sense of this capture? They have two approaches. One is to ascribe to addicts abnormal desires; the other is to ascribe to them mistaken beliefs. Advocates of the first approach typically see addicts as having steep temporal discount curves—they see them as having much stronger desires for the present and immediate future than for the more

⁴ Could they sometimes be truly irresistible? It seems rash to rule that out, although it is hard to be sure quite what the claim means: That no incentive *would* overcome it? That no incentive *could*?

distant future. Since addictive drugs normally involve a nasty period of withdrawal, an already addicted agent whose focus is on the immediate future will want to avoid embarking on the suffering that such a process will involve, even if they know that the long-term effects will be beneficial. Of course they might well prefer not to have started consuming the drug in the first place—in this way advocates of this approach can make sense of the idea that they are really addicted and are not simply willing consumers—but given the state that they are in now, continuing to take the drugs is preferable to withdrawal.⁵

Advocates of the second approach typically see addicts as mistaken, at least initially, about the effects of their drugs (they believe that they will not become addicted, or that addiction will not be so bad); or they see them as failing to take into account the consequences of current consumption for their future state: by focusing only on their current options, addicts fail to see that consuming addictive drugs now will lower their overall well-being in the future.⁶

The two approaches may be combined: mistaken beliefs might explain why addicts fall into addiction, and then the steep discount curves might explain why they stay there; and elements from these approaches might be used to supplement other accounts. Indeed we ourselves are inclined to think that there are important insights to be had here. In particular, there is good evidence that ignorance has an important role in the process of acquiring an addiction. But we do not think that an ordinary choice account can provide the fundamental explanation of what is distinctive about addiction. For if they were right, then a former addict who had been through the pains of withdrawal should be the least likely to consume again. They would no longer have the cost of withdrawal to endure; and they, of all people, would be well informed of their own vulnerability to addiction, of how nasty it is, and of the cost of not looking to the future. We are not talking here of the person who really prefers to be addicted; they will just start consuming again, although such a person would be unlikely to have put themselves through the process of withdrawal in the first place. But the person who genuinely wanted to be free of the drugs should be uniquely well qualified to ensure that they remain so.

Yet that is not what we find. People who have come through withdrawal, and gained much self-knowledge in the process, are much more likely to take up drugs again than those who never started, a process that is typically triggered by cues that are associated with the previous addiction. Indeed, withdrawal seems largely irrelevant in the process of maintaining addiction. Not only do people consume again after having gone through it, cravings are experienced long before it comes in, and some highly addictive drugs—most notably, cocaine—have minimal withdrawal symptoms. A pure choice model struggles to explain these features. So let us turn to the disease models that do better with them.

⁵ The most influential presentation of this line is from Becker and Murphy; for a simplified presentation see (Skog 1999). Becker and Murphy give no explanation of how addicts get into the state of addiction; that is left to be explained by exogenous factors.

⁶ See Loewenstein (1999) and Herrnstein and Prelec (1992) for versions of the first approach; and Heyman (2009, Ch. 6) for a detailed development of the second. An addict, understood on Heyman's lines as one who fails to think about their future, will be behaviorally equivalent to the steep discounter who doesn't care about it; but this will derive from features of their beliefs rather than of their desires.

II. DISEASE MODELS

There are many disease models of addiction. To get some traction on the debate, we divide these into four, at the cost of some simplification. The first sees addiction as habit: drug-taking actions are triggered automatically in particular situations, independently of the subject's beliefs and desires. The second sees it as involving distorted pleasure: addictive drugs 'hijack' the subject's pleasure circuits, or conversely, as a desperate attempt to escape the intense suffering of withdrawal states, distressing life circumstances, and/or continuing existential anguish. The third, using reinforcement learning theory, sees the distortion as affecting not the pleasure itself, but the subject's beliefs about what will give them pleasure. The final account, which we endorse (whilst denying that this provides the whole story about addiction), involves desire: consumption of addictive drugs gives rise to pathologically intense desires or cravings, states that are largely insulated from the subject's beliefs and other desires. We start by briefly outlining those with which we disagree.

Habit accounts

In its simplest form the habit model follows the classic stimulus-response account laid down in the early 20th century by Thorndike. In its early behaviorist guise, this approach was linked with skepticism about mental states altogether; but such an approach has few supporters now, and we shall say nothing about it. More interesting is the idea that habits stand alongside, but independent of, the agent's beliefs and desires (Wise 2004; Everitt et al. 2008). Contemporary versions of the theory hold that drugs induce brain systems of action to perform particular behaviors when cued, behaviors that have been established by previous drug taking episodes—much like a shoe-tying habit but even more strongly automatic.

If addictive states were understood this way, that would provide some explanation of why they are insensitive to addicts' desires to stop. But the account implies that drug taking is unmotivated, and most likely to surface when the addict's attention is distracted. That belies the intensely motivated nature of addictive urges, and the observation that attentively thinking about drugs, rather than being distracted, is the most dangerous situation for an addict. Whilst some aspects of habitual behavior might be important—reaching unthinkingly for a cigarette—the account cannot easily explain why an agent will take drugs in full awareness of what they are doing, but quite contrary to their views of what is best.⁷

Pleasure accounts

So let us move the second class of accounts, those premised on excessive pleasure.⁸ Clearly many addicts do get great pleasure from the drugs they take. If drugs can 'hijack' the pleasure circuit, giving a disproportionate amount of pleasure to those who consume them, then this

⁷ For work on the areas in which pure habit accounts do provide good explanations see Wood and Neal (2007).

⁸ Thorndike's original account of learning was in terms of pleasure, though he later came to talk purely in terms of stimulus and response. Historically then, pleasure based accounts represent something of a reversion to an earlier idea.

would give rise to a very strong learned desire for them.⁹ And if the pleasure per unit decreased over time, as tolerance developed, the agent would want more and more of the drugs to compensate (Wise 1980, 1985).

This account was once thought to be bolstered by the finding that the addictive drugs boost the mesolimbic dopamine system: either by stimulating the production of dopamine (in the case of amphetamine, nicotine, caffeine); by reducing the production of substances like GABA that themselves reduce the amount of dopamine released (opiates, perhaps THC); by reducing the level of substances that break down dopamine (alcohol); or by reducing the activity of the system that reabsorbs dopamine (cocaine, and perhaps amphetamine). Add the premise that the mesolimbic dopamine system is the pleasure system, and we have what looks like a compelling picture.¹⁰

A converse ‘withdrawal’ version of the pleasure account focuses on addictive escape from the displeasure of withdrawal, or of other distressing conditions of life, which, it is posited, lead the addict to pursue drugs to regain an acceptable level of pleasure. This view is supported by findings that the brains of addicts often have lower levels of dopamine receptors available (specifically the D2 type of dopamine receptors that can be measured by current neuroimaging techniques). Repeated taking of drugs induces ‘downregulation’, a neuronal compensation in which neurons reduce their production of the D2 dopamine receptors in response to receptor overstimulation by those drugs (Volkow et al. 2004). Further, it has been suggested that some addiction-prone individuals have lower levels of D2 dopamine receptors to begin with—possibly associated with a innate tendency to have higher levels of dopamine responses to reward-related events, which might play the same role as drugs in providing overstimulation, and consequent downregulation, of their D2 receptors.

But the essential premise that brain dopamine is a mechanism of pleasure is false, in our view. A host of findings have now shown conclusively that the primary role of the dopamine system is not to do with pleasure. In rats, suppressing the dopamine system does not result in a lack of pleasure responses to sweet substances. Likewise, human subjects whose dopamine systems are suppressed artificially, or as a result of Parkinson’s disease, give normal pleasure ratings to sugar. Conversely, elevated dopamine levels in either rats or human beings do not result in greater pleasure (Berridge 2012, p. 1132).

Dopamine thus does not seem to be directly concerned with liking. We will suggest that it is concerned with the creation of wanting. This might not matter if there were nonetheless a very tight correlation between liking and wanting: if liking invariably resulted in wanting, and if wanting were invariably the result of prior liking. But the very results that show that they are

⁹ We speak of ‘hijacking’ and ‘disproportionate pleasure’ here, but accounts that think that there is no rational constraint on what gives one pleasure will find it hard to make sense of this. To that extent, this approach will lapse back into a choice account, in which the agent acts on desires for their strongest pleasure.

¹⁰ For a recent popular presentation of such an approach by a neuroscientist, see Linden (2011, Ch. 2). Linden writes ‘Addictive drugs, by co-opting the pleasure circuitry and activating it more strongly than any natural reward, create deeply ingrained memories that are bound up in a network of associations’ (p. 53).

distinct states also show that, whilst they might typically be linked by causal connections, sometimes those connections will fail. We will argue that this is crucial for understanding addiction.

Learning accounts

So let us move to those models that see addiction as resulting from learning. Admittedly in a simple behaviorist model learning is not a very substantial notion: there isn't much more to it than the idea that a subject's behavior changes as a result of what happens to them, and hardly anyone could disagree that that is true of addiction. But in more cognitivist models, the idea of learning is much more specific: it is the idea of forming predictive associations, that is, beliefs.

These accounts see addiction as stemming, not from heightened pleasure itself, but from mistaken beliefs about pleasure. Addictive drugs hijack the circuits that learn about pleasure, and so they distort the memories that are used to guide future desires. One popular theory of reward learning holds that dopamine spikes indicate 'reward prediction errors': dopamine is released whenever an outcome is better than expected (Sutton and Barto 1998). Applied to addiction, the idea is that dopamine-stimulating drugs cause an exaggerated prediction error (Schultz et al. 1997; Redish 2004). Consumption of the drug itself doesn't have to be especially pleasurable, since the effect on the dopamine system is to trigger a large prediction error as if it were pleasurable, with the result that the 'memory' of the pleasure greatly exceeds any actual pleasure. On this approach, addicts' fundamental desires are desires for pleasure. Since they mistakenly believe that consumption of the drug will give them pleasure, this results in a strong instrumental desire to consume.

We think that this is mistaken. We will present instead a model—the incentive salience model—that sees addiction as driven by desires that have no essential connection with beliefs about what will be liked, or about what will be beneficial in other ways. The key idea here is that the dopamine signals are not learning signals, in the sense that they do not give rise to beliefs, predictions or memories (real or apparent) at all. Instead, they give rise to desires directly—or, more accurately, to a sensitivity to experience desires when cued with appropriate stimuli. The desire felt is not an instrumental desire, driven by an intrinsic desire for pleasure; instead, it is an intrinsic desire for the drug, a desire that may lead to action even in the face of contrary desires, and in the face of beliefs that consumption will have terrible consequences.

Before we explain the evidence for such an account in any detail, let us get clearer on the distinctions we have just outlined: that between wanting and liking; and that between the formation of beliefs and the acquisition of desires.

III. DISTINGUISHING WANTING AND LIKING

In one sense it is obvious that wanting and liking are distinct, at least if we think of liking in terms of pleasure: wanting typically comes before one gets the thing wanted, whereas the pleasure typically (though not invariably) comes once one has got it. And liking and wanting can also come apart as a result of false beliefs. We can want something that we believe we will

like, even though we won't in fact like it: perhaps we haven't tried it before, or have forgotten that we didn't like it, or we believe for some reason that our reaction will be different to last time. For parallel reasons we can like something and not want it: we might not realize that we like it, or we might have other reasons for foregoing it.

So the real issues do not concern the identity of wanting and liking. Instead we think that they are two-fold. One concerns the causal relations between wanting and liking, and their embodiment in particular brain mechanisms. The second concerns the relation of wanting to expected liking. We take these in turn.

Causal relations between wanting and liking

Does liking invariably cause wanting? Conversely, is wanting always caused by liking? We answer 'no' to both questions. While we accept that brain activations that cause increases in liking typically cause increases in wanting too, we think that these mechanisms are separable, so that under the right conditions liking can be generated without wanting, and wanting without liking.

The evidence here came originally from studies of rats. Since rats can't talk we need to have some non-verbal behavioral indicators of wanting and of liking. Wanting is straightforward: rats want something if they try to get it. (This is where we assume that issues of self-control will not intrude; things are more complicated with human beings as we shall see later.) Liking is harder to identify. But a set of results indicate that a range of evolved facial expressions—including tongue protrusions and lip sucking—are correlated with liking for the sensory pleasure of tastes across a wide range of species including rats, monkeys and human infants (Berridge 2000; Berridge and Kringelbach 2008).

Once we have distinct criteria for wanting and liking, we find that one can be induced without the other. If rats' dopamine levels are suppressed, they are no longer prepared to work to gain food rewards that they would previously have worked for. At the extreme, they will not eat pleasant foods that are freely available, even though they still display strong liking for them once the foods are placed in their mouths. Indeed, rats who have 98% of the dopamine neurons in their nucleus accumbens and neostriatum chemically destroyed, and who would starve to death were they not intragastrically fed, nevertheless maintain their normal liking reactions, indicating that pleasure in food is unchanged. So liking is not sufficient for wanting. Conversely, by boosting rats' dopamine levels we find that their wanting can be increased without their liking being increased—we will discuss an example of this shortly (Berridge 2007; Smith et al. 2011). Indeed wanting can be artificially engendered in rats without any signs of liking at all (Peciña et al. 2003; Wyvell and Berridge 2000; Faure et al. 2010; Smith et al. 2011; Tindell et al. 2005; Berridge and Valenstein 1991). In the past decade, the distinction between liking and wanting has also been confirmed in a number of human studies based on ratings of their own experience of sensory pleasures, such as cocaine and other addictive drugs (Leyton 2010; Lawrence et al. 2003).

Relation of wanting to expected liking

The second issue concerns the relation of wanting to expected liking. Can subjects want something whilst believing they will not like it? And conversely, can they believe that they will

like something and not want it? This is where the talk of learning fits in: can subjects come to learn that they like something, and yet not go on to form a desire for it?

Here again the empirical evidence suggests that wanting without expected liking is indeed possible. Consider first a set of experiments done by Cindy Wyvell, which showed that levels of wanting could be manipulated even though the associated levels of liking or expected liking remained constant. The experiment worked in two stages. The first got rats to associate a random stimulus (a noise) and an activity (lever pressing) by pairing each with a sugar reward. As a result of this pairing, the rats acquired a conditioned response: they came to press the lever when they heard the noise, even if no sugar was present.¹¹ Their facial responses showed that they liked the sugar when they did get it.

The second stage involved probing the effect of changes in dopamine level on this behavior. To achieve this, cannulas were inserted into the rats' brains, enabling their mesolimbic dopamine systems to be affected directly by microinjection. A control group received an inert substance through this cannula, whilst the other group received amphetamines, which greatly increase dopamine release. The effects of the dopamine could then be determined by observing the differences between the two groups.

Both groups continued to like the sugar. Indeed, they liked it to the same degree, evidence that dopamine does not produce pleasure. Importantly though, the amphetamines did not seem to increase anticipated pleasure from the sugar either: when given the lever to press, the amphetamine group did not press it any more frequently than the control group.

The difference came when the rats heard the noise that they had been conditioned to associate with sugar. Now both groups increased their lever pressing. But rats in the amphetamine group pressed it dramatically more: more than four times as frequently as before, and more than 50% more than the rats in the no-amphetamine group. And this effect was switched on and off as the noise went on and off.¹²

It appears that the increased dopamine levels resulted in a massive amplification of the conditioned response that was already present. It is very hard to explain this result in terms of changes in expectation, for we have no reason to think that hearing the noise caused change in the rats' beliefs about how pleasurable the sugar would be. The rats were not learning anything new; and the effects fell off as soon as the noise ceased.

There are many other experiments that confirm that learning, which is what would be involved in expected liking, is different from wanting. For example, the rats discussed above who have lost nearly all of their mesolimbic dopamine due to neurochemical 6-OHDA lesions are still quite capable of learning new values about food rewards. When a previously liked food is made unpalatable by inducing nausea, the dopamine depleted rats will learn to react to it with signs of disgust, in just the same way as normal rats. Similarly, mice who have been

¹¹ See Robinson and Berridge (2003, pp. 41–3) for further discussion of why this feature is important.

¹² It's an interesting question why the sight of the lever didn't itself work as a cue. Clearly not all cues are created equal.

genetically engineered to lack dopamine are still able to learn basic Pavlovian reward associations (Berridge 2012, pp. 1139–40). Learning, in the sense of the formation of new beliefs or of the formation of new behavior, does not seem to be essentially dependent on the dopamine system.

So what exactly is dopamine doing? As we saw from the Wyvell experiments, it is involved in the generation of desires in response to certain stimuli. To see why such a system might have evolved, let us start by employing some relatively a priori considerations about creatures like us and about the kind of wanting system that would benefit us.

IV. MODELING THE WANTING SYSTEM

Some creatures are tightly locked into a specific pattern of consumption: an insect that eats the leaves from a single plant species, or a koala that eats the leaves from four. Such creatures can have their tastes hard-wired. Other creatures are more opportunistic, adapting their consumption patterns to what is available. Human beings, like rats, are at the far end of this continuum. Although some of our desires are perhaps hard-wired, most are highly plastic.

Let us think in the abstract about how a creature with plastic desires will structure its consumption. We assume that it has some way of telling, when it samples a given food, how good that food is in providing it with what it needs. The goal of the creature is to maximize its consumption of things that are good for it. How could it go about that?

One way would be for the creature simply to try each thing that it comes across to see how good it is and then consume it if it is; but obviously that would be highly inefficient, since it would involve constantly retrying things which had already been shown to be bad. A second would be to learn what is good for it, in the full sense of that term: the creature would develop beliefs about which foods are good, and then, given its desire to consume what is good, it would form instrumental desires for just those foods (Dickinson and Balleine 2010). A third possibility would be to avoid forming the beliefs at all. Instead the creature could directly form its desires on the basis of what it had discovered to be good. That is, it could form intrinsic desires for the good foods, without recourse to any beliefs or predictions about them.

This third possibility would have some advantages. It could be simpler and easier to implement than a belief based system, and in some ways more robust. So let us consider how it might work. It would need to do two things. First it would need to form long-term intrinsic desires for the kinds of food it has found good; second it would need to act on those desires in the presence of the foods. For this it would need two systems, a desire formation system which would create intrinsic desires, and a consumption system which would regulate the creature's consumption in accord with those desires.

We suggest that both rats and human beings exhibit both of these systems, working together in just this way. To see this for rats, return to the findings of Cindy Wyvell. As mentioned, Wyvell found that boosting dopamine caused huge increases in short-term wanting in the presence of the relevant cue. But this was not all. She also ran a parallel set of experiments on rats who had received earlier amphetamine injections during the initial

conditioning, rather than at the time of the later stimulus. She found that this sensitized their brains in a long-term way. Despite being free of the drug for ten days, the later stimulus of the sound elicited twice the rate of lever pressing from these rats as it did from a control group who had not received the sensitizing injections. The dopamine was creating long-term intrinsic desires (Wyvell and Berridge 2001; see also Tindell et al. 2005; Smith et al. 2011).

So here we have evidence of exactly the kind of model mentioned. Dopamine is involved in two processes: the laying down of long run intrinsic desires; and then the triggering of those desires in the presence of the relevant cues.

V. ADDICTION AS MALFUNCTION OF THE WANTING SYSTEM

Now that we have the model in place, our account of addiction can be quick. What would happen if a subject consumed a substance that caused an artificial boost in dopamine levels? The effect would be two-fold. First, it would experience a boost in its immediate desire for the substance. Second, it would experience a boost in its long-run desire for it. That desire would be cued by the substance itself, or by other cues that were around at the time that the substance was initially consumed. If the dopamine signal was strong enough, the ongoing sensitization could be very great, potentially persisting indefinitely.

Our claim is that this is just what happens in cases of addiction. Since the addictive drugs artificially stimulate the dopamine system so powerfully, they give rise to long lasting dispositional desires. The dispositional desires are triggered by cues surrounding the consumption of the drugs: the drugs themselves, but also, given the associative nature of the process, the places in which they are consumed, the paraphernalia surrounding their consumption, and so on. Since these are intrinsic and not instrumental desires, they are not undermined by the belief that consumption of the drugs will not be pleasurable, or that it will be harmful in some other way. These dispositional desires may persist long after the subject has stopped taking the drugs, and has gone through any associated withdrawal. A cue provided by seeing the drug, or the environment in which it was once taken, or even by imagining it, may provoke a powerful occurrent desire for it; and if this results in further consumption, the whole pattern will be repeated.

This seems to fit the facts very well. Or at least, it fits some of the facts very well, the pathological facts, those concerning the way that addiction differs from ordinary behavior. But it might seem that this has taken us too far. For what are we to make of those aspects of addiction that make it seem very much like ordinary behavior? Can we preserve the idea that addicts are nonetheless sensitive to standard incentives?

The crucial point here is that, in human beings, the incentive salience process that we have sketched does not necessarily lead directly to behavior.¹³ It typically leads instead to cravings:

¹³ This is not to deny that incentive salience effects can work unconsciously in a way that takes them fairly directly to behavior. See Winkielman and Berridge (2004). But such behavior is still susceptible to self-control; it is just that the subject doesn't see the need to exert it.

to powerful desires that tend to crowd out other considerations. (See Loewenstein (1999) for a good discussion of how cravings tend to narrow one's focus.) Many philosophers make a sharp contrast between desires and intentions. Desires are the inputs to deliberation; it is quite rational to have many that conflict. Intentions are the outputs of deliberation; they are insulated from reconsideration and lead directly to action, and so they generally need to be consistent (Holton 2009). Cravings seem to come somewhere between the two. Whilst they have many of the features of standard desires, they are not easily thought of as inputs to deliberation. Rather, they lead directly to action unless something stops them. Stopping them requires self-control; to this we now turn.

VI. SELF-CONTROL

Both philosophers and psychologists tend to view desires as a fundamentally uniform class. Roughly, they are the states that move an agent to action. In contrast we think that they are heterogeneous. So far we have focused just on one kind, the desires, or cravings, that result from the incentive salience process. As we mentioned at the beginning, we also have other, more rationally tractable desires: a desire to take a holiday in St Petersburg, say, or to be healthy, or to treat a particular person well. And many of these are intimately connected with our beliefs. If we come to think that St Petersburg is too Western to reveal the true Russia, and it is the true Russia that we are after, then our desire to visit will be undermined. In contrast, the cravings that result from the incentive salience process are not typically undermined by the belief that they are harmful. (For an excellent discussion of such desires see Railton (2012).) Many actual desires may combine an element of both types; indeed the very case that Railton uses as illustrative of the more cognitive desire—a desire for an espresso—is very plausibly a case in point.)

But if we have at least two different sorts of desires—together perhaps with other factors that also influence our behavior, like our habits—then the question arises of what it is that determines what we will do. This is a difficult and complex question that we cannot hope to fully answer here. But one thing that we think has become clear in recent years is that it is not fully determined by the relative strength of the different sorts of desires. We also need to factor in a more active control on the part of the agent.

Although the details remain controversial, a wealth of psychological research supports the idea that self-control should be taken seriously. Self-control develops in children after the development of desires; it is effortful; it is depleted by various factors including stress, fatigue, and its prior exercise; and it can be developed and deployed more or less successfully (for general discussion see Holton (2009).) A failure to behave a certain way might indicate a lack of desire to behave that way. Alternatively it might indicate that a desire, even the kind of craving that results from addiction, is being held in check by self-control.

To say that self-control is real is not to deny that its exercise is sensitive to the agent's beliefs and desires. Agents can be well motivated to employ it, if they think that there is something to be gained from it, and that its employment will be successful. Alternatively, if they think that it will bring little benefit, or that the benefits can be gained more easily another

way, or that it is unlikely to succeed, they will be far less likely to employ it, and even if they do initially employ it, given that it is effortful, they will be far more likely to give up.

As we have seen, the pathology of addiction means that addicts will experience strong cue-driven cravings long after withdrawal is over, especially at particular moments such as when a drug cue is encountered in a moment of stress or emotional excitement. But this is not the end of the story. Whilst there is some evidence that addictive drugs can diminish self-control by damaging the prefrontal cortex (Volkow et al. 2004), there is no reason to think that addicts lose it altogether. Indeed, the fact that addicts can get themselves off their addictions is strong evidence that they do not. Controlling cravings may be tremendously hard work, but that it is not to say that it is impossible. Understanding when it is that addicts will continue to consume and when they will not thus requires an understanding of how their cravings interact with their self-control. Whilst we do not have even the beginnings of a real account here, we identify the following factors as very likely to be relevant to the pattern of activity that we remarked on at the outset, in particular the responsiveness of addicts to incentives, and their tendency to escape their addictions in their late twenties or early thirties.

(a) the strength of the self-control system

There is evidence that self-control, regulated primarily by the pre-frontal cortex, continues to develop in strength into the mid-twenties, typically maturing rather earlier in women than in men (see, for instance, Luna and Sweezy 2004; Goldstein et al. 2009; and, for a popular review, Sabbagh 2006).

(b) the efficiency with which the self-control system is employed

A great deal of research indicates that there are techniques that enable agents to better deploy their self-control. Forming prior intentions and then acting on them without reopening the question of what to do seems important, as does avoiding tempting situations. Similarly, mindfulness techniques can enable agents to stand back from their desires in ways that make their self-control more effective. It is still an open question how effective such techniques can be against the kinds of cravings engendered by addiction, but initial research indicates that they can make a difference (Prestwich et al. 2006; Kober et al. 2010). Again, skill in using the self-control system is something that we might expect to increase with age.

(c) the role of desires

Addicts who have strong motivations for giving up rather than continuing are more likely to employ their self-control to overcome cravings. And it does seem likely that concerns about partners, families, and careers will become more pressing as people reach their late twenties and early thirties. Conversely, since dopamine levels start to fall from the teenage years onwards, the power of the cravings may themselves diminish.

(d) the role of belief

If addicts think that there is little reason to give up today, since giving up tomorrow will be just as good, there will be little motivation to employ self-control. Vague concerns about health and well-being are often of that form; there can be a sense that, whilst giving up is something that needs to be done at some point, one more dose won't hurt. In contrast, the incentives that

have been shown to work well—for instance the knowledge that certain dismissal from a much valued job will follow a single positive drug test—guarantee an immediate cost or benefit. We suspect that much the same is true of a price rise; whilst it is true that paying the higher price just one more time is probably within the addict’s reach, there is no escaping the fact that a higher price is being paid. The other set of relevant beliefs concern the efficacy of exerting self-control. If the addict is convinced that they will succumb despite their best efforts—if not today, then surely soon—the motivation to try will be much reduced. And here, presumably, the addict’s own theory of addiction will have a part to play. If they think of the addiction as resulting in behavior that is quite outside their control, they will be far less motivated to try to control it (a point that has been noted many times by Albert Bandura; see for instance Bandura (1999).)

VII. THE EXTENT OF ADDICTION, AND ITS RATIONALITY

We have talked about addictions that are caused by drugs—by substances that interfere directly with the dopamine system, and gain their incentive salience effect from that interference. But what of the many other kinds of behavioral addictions—addictions to gambling, shopping, sex or the internet—that feature so prominently in current discussion. Can we give an account of them? Or is the theory we have given bound to say that they these are not really addictions?¹⁴

Clearly our account is bound to say that there is an important difference between substance and behavioral addictions. The latter do not, so far as we know, involve mechanisms that short-circuit the dopamine system in the way the former do. Nevertheless, there is good reason to think that they too work through the incentive salience system, and so that they too can result in cue driven cravings that are relatively insulated from other desires and from beliefs about what is good. Of course, if the dopamine system has not been short-circuited, then these behavioral addictions must have originated from behavior that was pleasurable, or was in some other way recognized by the agent’s dopamine system as being beneficial. But the assessment of the dopamine system might be at odds with the agent’s more cognitive beliefs about the value of the activity; and even if it is not, once the intrinsic desires have been established, they will tend to persist through changes in the agent’s assessment at any level. Even if the agent stops liking the thing concerned, a well-established incentive salience desire will degrade very slowly. The result can be behavior that looks very like the addiction engendered by drugs (Dill and Holton 2014). (Further evidence that drug and behavioral addictions have much in common comes from the cases of Parkinson’s patients who respond to their dopamine supplement by developing addictive behavior (O’Sullivan et al. 2009). We leave open the question of whether other behaviors that look rather like chemical addictions—

¹⁴ We have made the traditional division between substance addiction and behavioral addiction, but it could be that some substances give rise to addiction-like behavior without hijacking the dopamine system in the way we have discussed, and so should be grouped with the behavioral addictions. Sugar might be like that, and perhaps, though here the findings are controversial, cannabis. So a more careful distinction would be between the dopamine-hijacking addictions, and those that are not. But we will stick with the more traditional terminology.

those resulting from obsessive compulsive disorder, for instance—should also be understood in the same way.)

This brings us finally to an issue that we have largely skirted up till now, that of the rationality of addicts. Most ordinary choice models see addicts as quite rational, though working with unusual desires or false beliefs (perhaps there is some irrationality in how they arrived at those beliefs, but that doesn't affect the rationality of how they act upon them). Most disease models see the addict as largely arational: addictive actions hardly count as intentional actions at all, and so fall outside the scope of rationality. In contrast, the account that we have developed here sees the addict as potentially irrational in two ways. One is familiar: if considered views about what would be best diverge from action, then both substance addicts and behavioral addicts will frequently be akratic, in ways that have at least a prima facie claim to irrationality. The second is rather less familiar. If what we have said is right, then something goes badly wrong with the process by which substance addicts (but not behavioral addicts) form their desires: substances come to be desired independently of any pleasure or other benefits that they bring. There has been much discussion in philosophy of whether intrinsic desires can be irrational. What we are suggesting is that substance addiction results from the malfunctioning of a normally rational system for creating intrinsic desires. This seems to us as clear a case of an irrational intrinsic desire as one is ever likely to find.

VIII. CONCLUSION

We started by stressing the need to find a middle path. Our attempt to find one has involved exploring the interaction between two different systems: one that regulates our desires, and one that controls which desires we act on. Addiction results from the malfunction of the first; insofar as it does not result in a complete loss of agency, that is thanks to the second. In a sense then, both the disease model and the choice model are describing something real; but each gives a picture that is partial. We hope that we have gone some way to putting them together.

References

- Bandura, A. (1999). A sociocognitive analysis of substance abuse. *Psychological Science*, 10, 214–7.
- Berridge, K. (2000). Taste reactivity: measuring hedonic impact in infants and animals. *Neuroscience and Biobehavioral Reviews*, 24, 173–98.
- (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, 191, 391–431.

- (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. *European Journal of Neuroscience*, 35, 1124–43.
 - and Kringelbach, M. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology*, 199, 457–80.
 - and Valenstein, E. (1991). What psychological process mediates feeding evoked by electrical stimulation of the lateral hypothalamus? *Behavior Neuroscience*, 105, 3–14.
- Dickinson A. and Balleine, B. (2010). Hedonics: the cognitive-motivational interface. In Kringelbach, M. and Berridge, K. (eds.). *Pleasures of the Brain*. Oxford: Oxford University Press, pp. 74–84.
- Dill, B. and Holton, R. (2014). The addict in us all. *Frontiers in Psychiatry*, 5:139. doi: 10.3389/fpsyt.2014.00139.
- Everitt, B., Belin, D., Economidou, D., Pelloux, Y., Dalley, J., and Robbins, T. (2008) Neural mechanisms underlying the vulnerability to develop compulsive drug-seeking habits and addiction. *Philosophical Transactions of the Royal Society of London B Biological Science*, 363, 3125–35.
- Faure, A., Richard, J., and Berridge, K. (2010). Desire and dread from the nucleus accumbens: cortical glutamate and subcortical GABA differentially generate motivation and hedonic impact in the rat. *PLoS One*, 5, e11223.
- Goldstein, R., Craig, A., Bechara, A. et al. (2009). The neurocircuitry of impaired insight in drug addiction. *Trends in Cognitive Science*, 13, 372–80.
- Herrnstein, R. and Prelec, D. (1992). A theory of addiction. In Loewenstein, G. and Elster J. (eds.). *Choice Over Time*. New York: Russell Sage Press. pp. 331–60.
- Heyman, G. (2009). *Addiction: A Disorder of Choice*. Cambridge MA: Harvard University Press.
- Hoffman, J. and Froeke S. (eds.) (2007). *Addiction: Why Can't They Just Stop?* Emmaus, PA: Rodale Press.
- Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford: Clarendon Press.
- and Berridge, K. (2013). Addiction between compulsion and choice. In Neil Levy (ed.). *Addiction and Self-Control*. Oxford: Oxford University Press pp. 239–68.
- James, W. (1890). *Principles of Psychology*. New York: Henry Holt.
- Kober, H., Kross E., Mischel M., Hart C., and Ochsner K. (2010). Regulation of craving by cognitive strategies in cigarette smokers. *Drug and Alcohol Dependence*, 106, 52–5.
- Lawrence, A., Evans A. and Lees A. (2003). Compulsive use of dopamine replacement therapy in Parkinson's disease: reward systems gone awry? *Lancet Neurology*, 2, 595–604.
- Leyton, M. (2010). The neurobiology of desire: dopamine and the regulation of mood and motivational states in humans. In Kringelbach M. and Berridge K.. *Pleasures of the Brain*. New York: Oxford University Press, pp. 222–43.
- Linden, D. (2011). *The Compass of Pleasure*. New York: Viking.

- Loewenstein, G. (1999). A visceral account of addiction. In Elster, J. and Skog O. (eds.) *Getting Hooked*. Cambridge: Cambridge University Press, pp. 235–64.
- Luna B. and Sweezy J. (2004). The emergence of collaborative brain function: fMRI studies of the development of response inhibition. *Annals of the New York Academy of Science*, 1021, 296–309.
- O'Sullivan, S, Evans, A. and Lees, A. (2009). 'Dopamine dysregulation syndrome: an overview of its epidemiology, mechanisms and management. *CNS Drugs*, 23, 157–70.
- Peciña, S., Cagniard, B., Berridge, K., Aldridge, J. and Zhuang, X. (2003). Hyperdopaminergic mutant mice have higher "wanting" but not "liking" for sweet rewards. *Journal of Neuroscience*, 23, 9395–9402.
- Prestwich, A., Conner, M. and Lawton, R., (2006). Implementation intentions: can they be used to prevent and treat addiction? In Wiers R. and Stacy A. (eds.). *Handbook of Implicit Cognition and Addiction*. Thousand Oaks, CA: Sage, pp. 455–71.
- Railton, P. (2012). That obscure object, desire. *Proceedings and Addresses of the American Philosophical Association*, 86, 22–46.
- Redish, A. D. (2004). Addiction as a computational process gone awry. *Science*, 306, 1944–7.
- Robinson, T., and Berridge, K. (2003). Addiction. *Annual Review of Psychology*, 54, 25–53.
- Rush, B. (1812) *Medical Inquiries and Observations, Upon the Diseases of the Mind*, 1st Edition. Philadelphia: Kimber and Richardson.
- Sabbagh, L. (2006). The teen brain, hard at work. *Scientific American Mind*, August/September, 20–25.
- Schultz, W., Dayan, P. and Montague, R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–9.
- Skog, O. (1999). Rationality, irrationality and addiction—notes on Becker's and Murphy's theory of addiction. In Elster, J. and Skog O. (eds.) *Getting Hooked*. Cambridge: Cambridge University Press, pp. 173–207.
- Smith, K., Berridge, K. and Aldridge, J. (2011). Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proceedings of the National Academy of Science*, 108, E255–E264.
- Sournia, J. (1990). *A History of Alcoholism*. Oxford: Blackwell.
- Sutton, R. and Barto, A. (1998). *Reinforcement Learning*. Cambridge MA: MIT Press.
- Tindell, A., Smith, K., Berridge, K., and Aldridge, J. (2009). Dynamic computation of incentive salience: "wanting" what was never "liked". *Journal of Neuroscience*, 29, 12220–8.
- Berridge, K., Zhang, J., Peciña S. and Aldridge J. (2005). Ventral pallidal neurons code incentive motivation: amplification by mesolimbic sensitization and amphetamine. *European Journal of Neuroscience*, 22, 2617–34.

- Volkow, N., Fowler, J. and Wang G. (2004). The addicted human brain viewed in the light of imaging studies. *Neuropharmacology*, 47, 3–13.
- Wilde, O. (1891). *The Picture of Dorian Gray*. London: Ward Locke and Co.
- Winkielman P. and Berridge, K (2004). Unconscious emotion. *Current Directions in Psychology*, 13, 120–3.
- Wise, R. (1980). The dopamine synapse and the notion of 'pleasure centers' in the brain. *Trends in Neuroscience*, 3, 91–5.
- (1985). The anhedonia hypothesis: Mark III. *Behaviour and Brain Science*, 8, 178–86.
- (2004). Dopamine, learning and motivation. *National Review of Neuroscience*, 5, 483–94.
- de Wit, H. and Chutuape, M. (1993). Increased ethanol choice in social drinkers following ethanol preload. *Behavioral Pharmacology*, 4, 29–36.
- Wood, W. and Neal, D. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114, 843–63.
- Wyvell, C. and Berridge K. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward. *Journal of Neuroscience*, 20, 8122–30.
- (2001). Incentive sensitization by previous amphetamine exposure: increased cue-triggered “wanting” for sucrose reward. *Journal of Neuroscience*, 21, 7831–40