This work was supported by the Wellcome Trust [WT104848/Z/14/Z].

Levy, N. (2018) ‘Addiction: The Belief Oscillation Hypothesis’. In Hanna Pickard and Serge Ahmed (eds). *The Routledge Handbook of Philosophy and Science of Addiction*, Routledge.

5
ADDICTION
The belief oscillation hypothesis

Neil Levy

In popular, philosophical and many scientific accounts of addiction, strong desires and other affective states carry a great deal of the explanatory burden. Much less of a role is given to cognitive states than to affective. But as Pickard and Ahmed (2016; see also Pickard 2016) note, addiction may be as much or more a disorder of cognition as of compulsion or desire. Pickard's focus is on denial. In this chapter my focus will be different. I will argue that in many cases at least, we can explain the lapses of abstinent addicts by way of processes that do not involve motivated reasoning (as denial or self-deception plausibly do). Mechanisms that have the role of updating beliefs in response to evidence may alter addicts' judgments concerning what they have most reason to do (in the precise circumstances in which they find themselves), and thereby cause them to act accordingly.[1]

My focus is on abstinent addicts, understood here as addicts who sincerely resolve to refrain from consuming the drug to which they are addicted and who are committed to this resolution. For many of these abstinent addicts, the resolution is not an idle wish. They may act on it in myriad ways. They may make changes, large and small, in their lives to try to make it more likely that they will abide by it. They may move homes, jobs, withdraw from their social circle, and so on, because they believe (on good grounds) that so doing will help them to remain abstinent. They may enter treatment programs, and they may commit resources to being abstinent (they may, for instance, pay for expensive treatments). We should not underestimate the efficacy of these kinds of steps. Most addicts do succeed in giving up, and many do so without any external help at all (Heyman 2009). Nevertheless, and notoriously, many abstinent addicts relapse, some after long periods of abstinence.

I am concerned with explaining these relapses and, more proximally, with explaining lapses. A *lapse* is an episode of consumption by a hitherto abstinent addict. A *relapse* is a return to regular substance abuse. Contrary to a popular myth, lapses do not always result in relapses. Addicts may lapse and then return, with more or less difficulty, to durable abstinence. However, relapses begin with lapses. Explaining lapses is not sufficient for explaining relapses, but it will contribute very significantly to an explanation of relapse.

The kinds of explanations of lapses that circulate in popular thinking about addiction do not seem to come close to explaining all of them. Commonly, these explanations cite irresistible urges, or the duress that withdrawal involves. On the first account, addicts lapse because they are gripped by a desire to consume that overcomes their powers of resistance. On the second,

withdrawal is so awful that they are powerfully motivated to do almost anything to alleviate it. No doubt the strength of desires plays a role in explaining some lapses, and no doubt addicts are motivated to alleviate the pains of withdrawal. But there is plenty of evidence, direct and indirect, that neither explanation, alone or combined, is adequate.

First, there is extensive evidence that addicted people possess a great deal of control over their behaviour, even when they are in the grip of strong desires. Consumption of drugs, even among addicts, is price sensitive (Elster 1999; Neale 2002), which suggests a responsiveness to perfectly ordinary incentives. Consumption remains price sensitive among alcoholics even after a priming drink (Fingarette 1988). Second, the epidemiological evidence – the fact that the majority of addicts 'mature out' of their addiction – is hard to square with an irresistible desire account. But the withdrawal account does even worse. Not all drugs produce highly aversive withdrawal symptoms when an addicted person abstains from them (West 2006). Further, lapses may occur months, even years, after the last episode of consumption – long after the person has put withdrawal and its symptoms behind them. In fact, there is some evidence that addicts may deliberately go through withdrawal in order to lower their tolerance to the drug, and thereby return to drug taking requiring a smaller dose for the same effect (Ainslie 2000).

While there is extensive experimental evidence for reduced control by addicts (see, for instance, Hester and Garavan 2004), they retain a sufficient degree of control, measured by their responsiveness to ordinary incentives, for it to be very mysterious how this reduction of control could explain lapses except in a small minority of cases. Problems with executive function can explain impulsive behaviour: they can explain lapses that resemble the reactivation of a habit (for instance, they might be able to explain why an abstinent person with an addiction might find herself impulsively accepting the offer of a drink). But many lapses involve some degree of planning and preparation: rather than simply accepting an offer, the person resolves to seek out an opportunity for consumption and then constructs a plan to procure drugs (a plan that may be quite complex: it may be necessary to concoct a cover story for work or family, to secure money to score, and so on). Simple executive dysfunction can't explain this kind of activity. Or rather, it can't explain it *directly*: executive dysfunction may play a role in explaining lapses in these cases by playing a role in explaining how it can come about that addicts come to judge that (all things considered) they ought to procure and consume a drug.

Towards a judgment-centred account

Most of the drug procuring and consuming behaviour of addicts is entirely normal, in the following sense: it gives every appearance of being what psychologists call controlled behaviour, or of being explicable by what philosophers call belief/desire or folk psychology. Given the reasons-responsiveness of the behaviour – the way in which it is sensitive to perturbations in the environment, which give the agent reasons to inhibit or modulate their activity – the best explanation of why it unfolds as it does is that the agent *desires* to procure and consume drugs and that this desire has led them to form the *intention* to do so. Their behaviour is then controlled by this intention and other beliefs, particularly instrumental beliefs about how this is best brought about (most efficiently, at lowest cost in terms of resources and risk, and so on). Sometimes when abstinent addicts lapse they do so by consuming an immediately available drug. Episodes like that can often be explained by impulsivity, absence of control or reactivation of habits. But sometimes lapses involve a long series of behaviours aimed at procuring the drug first (or even first acquiring the money to procure the drug), and only subsequently consuming the drug hours or even days later. This long series of instrumental actions seems best explained by attributing to them the all-things-considered *judgment* that they ought to consume the drug.

Richard Holton (2009) introduced the helpful term 'judgment-shift' to refer to the overhasty revision of a resolution, in response to evidence or facts against which the resolution was meant to be proof. I suggest that judgment-shift underlies and explains many lapses. Addicted persons are especially vulnerable to drug-related judgment shifts, I claim. They have formed the resolution to abstain from drugs, and intend to abide by the resolution in circumstances like the ones in which they now find themselves. But in these circumstances they find themselves on what I have elsewhere called the 'garden path' to consumption (Levy 2016). Due to features of the neurobiology of addiction, finding themselves on the garden path disposes them to undergoing a cascade of judgment-shifts: first shifting toward judging that they ought to step down the garden path and subsequently to judging that they ought to consume.

Addiction causes a number of changes in the brain, such as those that cause the executive dysfunction already mentioned and those that cause anhedonia in response to withdrawal. Central to the addiction phenotype, however, is the response of the midbrain (mesolimbic) dopamine system to drugs of addiction (Hyman 2005; Kalivas and Volkow 2005). To understand the significance of this response, we need first to understand the system's role in ordinary cognition.

The midbrain dopamine system is widely held to be a reward prediction system. It has the role of signalling *unexpected* reward. In classic experiments performed on monkeys, activity in this system was measured when the animals were given a reward. In the initial condition, the reward (a squirt of fruit juice) was unexpected. There was a spike in phasic dopamine upon reward delivery. Subsequently, a signal (a light or a tone) was given prior to delivery of the juice. Once the monkey learned the signal that predicted the juice, there was no longer a spike in phasic dopamine in response to delivery of the reward (indicating that the spike was not tracking actual reward). Instead, the spike occurred in response to the (unexpected) signal. The signal was a sign that the world was better than the animal expected (Schultz *et al.* 1992; Schultz *et al.* 1997). Conversely, failure to deliver the reward subsequent to the signal led to dopamine falling below the resting baseline. Phasic dopamine thus seems to signal the state of the world relative to the organism's expectations. It thereby orients the animal toward the reward, making it more likely that they reap its benefits (dopamine is known to play a role in regulating attention: Corlett *et al.* 2007; Fletcher and Frith 2009).

In one way or another (by driving up dopamine directly or indirectly, or preventing reabsorption), however, addictive drugs 'hijack' this system (Carter and Hall 2012). Because addictive drugs drive up dopamine via their chemical action, they prevent it from accurately tracking reward value. When the addicted person is exposed to a cue that signals drug availability (the sight of a person with whom they have used in the past, say, or a sound, sight or smell they associate with using – anything that has been paired with consumption in the past often enough), they experience the spike in phasic dopamine that signals the availability of unexpected reward. But they also experience a spike in dopamine on consumption of the drug. For the relevant system, this spike is a signal of unexpected reward, despite the fact that, for the experienced user, the reward should be expected.

A predicted consequence of the fact that drugs of addiction produce a spike in phasic dopamine, thereby mimicking the signal of unexpected reward, is that the relevant system will not be able to adapt to the actual reward value of the drug. Instead, there will be a dissociation between the actual reward value of the drug for the person and the value placed on it by the midbrain dopamine system, helping to explain how addicts can be powerfully motivated to pursue drugs they no longer enjoy; the famous dissociation between 'liking' and 'wanting' (Berridge and Robinson 1995; Robinson and Berridge 2003). In fact, there is recent evidence that there is at least some accommodation of drug rewards with repeated exposure, both in animal models

(e.g. Willuhn *et al.* 2014) and humans (e.g. Martinez *et al.* 2007). Further, recent work has shown that the phasic dopamine response to drug-related cues is of a similar magnitude to that of non-drug cues (Cameron *et al.* 2014). However, the mesolimbic system is composed of populations of functionally discrete neurons and this evidence is compatible with the existence of systems that fail to adapt to the value of drug rewards and therefore generate dysfunctionally large – or dysfunctionally orienting – prediction errors.

I think it is a mistake to think of the midbrain dopamine system as simply a reward prediction system. At least some components of the prediction errors it generates have a more general role: to signal the unexpected generally, not just unexpected reward. Corlett *et al.* (2004) showed that activation of rostral prefrontal cortex (rPFC) – a primary target of dopamine projections – was correlated with violations of expectations on a task unrelated to reward. If I am right in thinking that the dopamine system has the role of updating expectations generally, then it should be thought of as a representation-update system. It has the function of changing the organism's model of the world. There is, in fact, evidence directly linking dysregulation of this system to delusional belief formation. Corlett *et al.* (2007) measured rPFC activity in delusional patients and a control group in response to violations of expectations. They found attenuated response to violations of expectation and abnormally high responses to expected events in the patient group. Corlett *et al.* (2006) found that magnitude of prediction error response in controls predicted the likelihood of delusions following administration of the hallucinogen ketamine. All of this is evidence the midbrain dopamine system plays a role in updating the organism's causal model of the world, in response to violations of expectations.

The fact that the behaviour of lapsing addicts often appears to be controlled behaviour – the fact, that is, that it is incentive-sensitive in a very ordinary way – provides us with good reason to think that when these addicts engage in behaviour aimed at procuring and consuming drugs, their behaviour is controlled by the all-things-considered judgment that they ought, on this occasion, to consume the drug (in broadly the same kind of way in which my bread buying and toasting behaviour is controlled by the judgment that toast would be nice). I suggest that the dysfunction that addictive drugs provoke in the midbrain dopamine system provides a mechanism for judgment-shift.

The mechanism is best presented in the framework of prediction error minimisation models of brain function (Hohwy 2013; Clark 2016). According to these models, the brain constantly predicts the input it 'expects' to receive. It then updates its expectations on the basis of actual input, to minimise the prediction error: the gap between expected and actual input. Updating expectations is updating a representation: it involves an alteration in the model of the world. It is therefore updating a doxastic state (the state may, but need not, have all the features associated with a belief, such as serving as a basis for domain-general inference; nevertheless, it is a representational state that may be correct or incorrect, and that is sufficient to qualify it as doxastic). I suggest that the midbrain dopamine system is, or is part of, a prediction minimisation mechanism.

When the addicted person who has resolved to be abstinent is presented with cues predictive of drug availability (again, anything reliably enough associated with drug-taking to serve as a conditioned stimulus), she experiences a spike in phasic dopamine that signals a gap between the expected input and the actual input. As assessed by the relevant subpersonal mechanism, the world is (much) better than expected: it contains the promise of a large reward. The person must now update her model of the world, to minimise the prediction error. This process occurs at the subpersonal level: the person is unaware of the content of the expectation or of the prediction error. Whether she will undergo a change in personal-level attitudes or representations depends on whether the error can be minimised at lower levels of the processing hierarchy. According

to the predictive coding framework, errors propagate up when there is no causal model available at the level at which the error occurs that can minimise it sufficiently (conversely, models propagate down). If the error is large enough and no such model is available at subpersonal levels of processing, model update may occur at the personal level.

This may be experienced by the person as a sense that some options are more rewarding than others. This is a sense that may be understood as a representational state. Suppose, for instance, the cue is a person with whom she has consumed in the past. She may experience a sense that being with the person will be rewarding, but she will be unaware that the reason she experiences this sense is because that person is associated with the availability of drugs, such that their presence is sufficient to cause a prediction error. She may instead just think that it would be fun to relive some old times with her friend. A judgment that going back to an old haunt with her friend would be rewarding may be the personal-level consequence of subpersonal representation update, caused by mechanisms that minimise prediction error.

The decision to spend some time with an old friend is innocuous enough, but it may place the person on the garden path all the way to consumption. Her friend may take her to somewhere that is more strongly associated with consumption, triggering another prediction error and another round of representation (and belief) update. She may thereby proceed step by step down the garden path, at each step unaware that she is heading toward consumption. It is only at the last step that she forms the judgment that (on this occasion, at least), procuring and consuming the drug is best. At the end of the garden path may be a relatively impulsive decision to consume. But the decision is the formation of a judgment: a controlling state that represents the world as being a certain way. Once this judgment is in place it is available to control behaviour generally, in just the same kind of way as her earlier resolution to be abstinent controlled her behaviour before she set out on this path. Note, too, that under the right conditions – when a prediction error is large enough and especially when competing rewards are not salient (lack of competition from competing rewards may be necessary, given the evidence that non-drug cues may trigger dopamine spikes of the same magnitude as drug cues: Cameron *et al.* 2014), the path from cue to personal level judgment may be much shorter.

Addicted people remain rational agents. They are not simply at the mercy of subpersonal mechanisms that cause representational states to change without apparent reason (any more than those of us who are not addicts are). We all make decisions and form judgments in ways that are guided by the outputs of subpersonal mechanisms, which dispose us to take some options seriously and dismiss others and may strongly incline us toward a particular judgment. Making decisions in this way is not a limitation on our rationality, but partially constitutive of it. We could not make decisions at all if we did not have mechanisms that narrowed the search space for us, ensuring that the set of options between which we deliberate are tractably few. This kind of mechanism often helps to constitute our rational processing. In this case, however, it is dysfunctional and produces a maladaptive response.[2] But because the person remains a rational agent, concerned to make sense of her own behaviour, she will justify her behaviour to herself in a way that is congruent with her former resolution to remain abstinent.

As Holton (2009) has emphasised, resolutions are never absolute. Rather, they come hedged with implicit escape clauses: conditions under which they are appropriately revised. I do not act irrationally or akratically if I revise or abandon my resolution to jog first thing in the morning should I realise that my partner is unwell, or that the road is flooded, or my neighbour needs urgent help. We do not, and cannot, specify the full range of such escape clauses: they are indefinitely many. Because this is true, we can quite easily deceive ourselves into judging that an escape clause has been triggered. I may, for instance, revise my resolution to skip dessert in the face of the thought that it's a special occasion. Of course, sometimes it *is* a special occasion,

and I am right to do so, but I may multiply special occasions until I am consuming dessert more often than I am not. The addicted person, too, is likely to explain her behaviour by seeing an implicit escape clause as triggered ("today has been especially stressful"; "just this once won't hurt"; "it would be hurtful to my friend to refuse", or whatever). This may not be mere post-facto rationalisation, in the sense that the unavailability of such a confabulated explanation of her own behaviour might be sufficient to prevent her abandoning her resolution. She has expectations about her own behaviour too, and the inability to update these expectations in a way that maintains a sufficient degree of coherence among her personal-level beliefs may constitute a stronger pressure to maintain the resolution than to abandon it.

Judgment-shift and control

Most people – laypeople and specialists alike – regard addiction as causing or constituting a pathology of control over behaviour. On the picture I have sketched in this chapter, the behaviour of the addicted person is controlled behaviour: at each moment, she acts rationally, given her own all-things-considered judgments. That does not entail, however, that she possesses rational control over her own behaviour. Or rather, the rational control she possesses may not extend across time sufficiently for us to say that she is genuinely in control.

The agent possesses control over her behaviour, including her drug-seeking and consuming behaviour, because it is appropriately sensitive to her reasons. Such reasons-sensitivity is what control consists in. As Fischer and Ravizza (1998) have argued, an agent or a mechanism possesses control over an action or a state of affairs to the extent to which she or it is capable of tracking reasons (which they call reasons-receptivity) and responding to some of those reasons (which they call reasons-reactivity). The addicted person's behaviour exhibits a great deal of such sensitivity, and therefore of control. This is true for most of the time when she is abstinent and abides by her resolution, and it is also true when her behaviour is controlled by her judgment that, all-things-considered, she ought to procure and consume the drug. But, while her behaviour may exhibit a great deal of control whichever judgment controls it, the shift from abstinence to seeking and using drugs is caused by mechanisms that exhibit an impairment in control (as measured by the strength and range of the reasons to which they are responsive). Prediction error minimisation mechanisms may be said to realise exquisite control in agents, when they are appropriately set up to track genuinely reason-giving states. But, in this case, the mechanism is dysfunctional and its capacity to track reasons is impaired.[3]

We began by noting that accounts of addiction very often give central place to loss of control as an explanatory notion. If the account offered here is on the right track, loss – or at least significant impairment – of control is indeed central to explaining lapses, and not only in those cases in which executive dysfunction or habit is at its heart. Extant control-based theories misplace the locus of the diminution of control when it comes to explaining many lapses. They explain lapse by a loss of control of behaviour, whereas on my account the loss of control is over judgment. Because the mechanisms that cause judgement-shift are significantly impaired in the range of reasons to which they are capable of responding, the person's control over her judgments is itself significantly impaired. Sometimes, at least, this impairment is significant enough that we may regard the person as losing control.

The fact that loss of control is central to explaining lapses is important, because many of the normative implications, and indeed implications for treatment, that are supposed to follow from extant control-based accounts may follow from mine too. If, for instance, a significant impairment of control excuses, then we have good grounds for reducing or eliminating blame for addicts in many circumstances. The fact that a loss of control is central to lapses may also

have implications for the incentives to which we can expect addicted people to respond, and therefore to the question of how we can help them moderate or eliminate drug use. Caution is needed in drawing normative and non-normative inferences, however: the fact that an agent lost control because a mechanism failed to be sufficiently sensitive to reasons, or was sensitive to cues that were not reasons, need not be an indication that the mechanism is generally reasons-insensitive. The same mechanism may realise control under some conditions while failing to do so under others.

Conclusion

People lapse for a variety of reasons, and there may be a multiplicity of mechanisms that explain lapses. Plausibly, they sometimes experience a loss of control over behaviour. But often they continue to engage in controlled behaviour even while lapsing. Judgment-shift explains how agents may lose control in one sense, while retaining it in another: they may experience an uncontrolled shift from one controlled state to another. It is possible that judgment-shift explains some cases of ordinary akrasia (thereby explaining how we may lose control "with calm and even with finesse" (Austin 1979: 198). More dramatic cases, like shifting from a judgment to which the agent is strongly committed and behind which she resolutely stands, may require dysfunction in the machinery controlling judgment.

In this chapter I have suggested that the failure of the midbrain dopamine system to adapt to the reward value of addictive drugs (and, possibly, of gambling too; Ross *et al.* 2008) provides a mechanism for such dramatic judgment-shifts. The way in which this system responds to cues predictive of the availability of the drug explains how the abstinent person may find herself on the garden path to consumption. If this picture is correct, it gives us targets for intervention: besides targeting the neuropsychological mechanisms that are dysfunctional, we can target the environment in which the person finds herself, to enable her to avoid the cues that trigger relapse.

Acknowledgements

I am grateful to Serge Ahmed and Hanna Pickard for detailed and challenging comments. Their comments led to significant improvements in almost every aspect of this chapter. This work was supported by a generous grant from the Wellcome Trust; WT104848/Z/14/Z, *Responsibility and Healthcare*.

Notes

1 In this chapter I use 'addiction' to encompass a problematic pattern of alcohol consumption as well as (other) drugs, and I use 'drugs' from now on to encompass alcohol. Making a distinction between 'alcohol' and 'drugs' is not only neurochemically unfounded, it is also politically invidious. It furthers what might be called the othering of those who consume illegal drugs and contributes to the perception that it is these alone that constitute the major social problem. Alcohol is responsible for more drug-related harm than any other drug (Nutt *et al.* 2010). Of course, many more doses of alcohol than of illegal drugs are consumed annually, so the incidence of harm is not a direct reflection of the harmfulness of the drug. Taylor *et al.* (2012) produced a consensus ranking of the harmfulness of drugs. Alcohol was ranked the fourth most harmful drug, behind heroin, crack cocaine and crystal meth, but ahead of many illegal drugs including cocaine, amphetamines and ecstasy.

2 At least, there is a strong case for saying the mechanism is dysfunctional *if* a subsystem in the nucleus accumbens fails to adapt to the reward value of drugs due to their production of a spike in phasic dopamine. Right now, we cannot be confident that this is in fact the case. If the response of the relevant

mechanisms is of a similar magnitude to non-drug rewards, then we may conclude that the system is working as designed, to orient the organism to rewards in its environment. In that case, the judgment-shift that occurs in addiction would be of the same kind we see in ordinary weakness of the will, and we would need to look elsewhere for what (if anything) makes addiction pathological. Addiction is, of course, complex, socially and neurobiologically, and we shouldn't expect the hypothesis offered here to constitute anything like a complete explanation of its power.

3 Fischer and Ravizza are, in my view, insufficiently sensitive to the factors that may diminish, without entirely removing, control. Elsewhere, I suggest ways of extending their account to accommodate the different dimensions along which control comes in degrees (Levy 2017). It is important to recognise the ways in which control is degreed, because its loss is almost never total: in the majority of those cases we (rightly) regard as involving a loss of control, the relevant mechanisms remain capable of responding to *some* reasons.

References

Ainslie, G. (2000) "A research-based theory of addictive motivation", *Law and Philosophy* 19(1): 77–115.

Austin, J. L. (1979) "A plea for excuses", in *Philosophical Papers*, *3rd edition*, Oxford, UK: Oxford University Press, pp. 175–204.

Berridge, K. C. and Robinson, T. E. (1995) "The mind of an addicted brain: neural sensitization of 'wanting' versus 'liking'", *Current Directions in Psychological Science* 4(3): 71–76.

Cameron, C. M., Wightman, R. M. and Carelli, R. M. (2014) "Dynamics of rapid dopamine release in the nucleus accumbens during goal-directed behaviors for cocaine versus natural rewards", *Neuropharmacology* 86: 319–328.

Carter, A. and Hall, W. (2012) *Addiction Neuroethics: The Promises and Perils of Neuroscience Research on Addiction*, Cambridge, UK: Cambridge University Press.

Clark, A. (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford, UK: Oxford University Press.

Corlett, P. R., Aitken, M. R., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., Robbins, T. W., Bullmore, E. T. and Fletcher, P. C. (2004) "Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning", *Neuron* 44(5): 877–888.

Corlett, P. R., Honey, G. D., Aitken, M. R., Dickinson, A., Shanks, D. R., Absalom, A. R., Lee, M., Pomarol-Clotet, E., Murray, G. K., McKenna, P. J., Robbins, T. W., Bullmore, E. T. and Fletcher, P. C. (2006) "Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis", *Archives of General Psychiatry* 63(6): 611–621.

Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R., Shanks, D. R., Robbins, T. W., Bullmore, E. T., Dickinson, A. and Fletcher, P. C. (2007) "Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions", *Brain* 130(9): 2387–2400.

Elster, J. (1999) *Strong Feelings: Emotion, Addiction and Human Behavior*, Cambridge, MA: The MIT Press.

Fingarette, H. (1988) *Heavy Drinking: The Myth of Alcoholism as a Disease*, Berkeley, CA: University of California Press.

Fischer, J. M. and Ravizza, M. (1998) *Responsibility and Control: An Essay on Moral Responsibility*, Cambridge, UK: Cambridge University Press.

Fletcher, P. C. and Frith, C. D. (2009) "Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia", *Nature Reviews Neuroscience* 10(1): 48–58.

Hester, R. and Garavan, H. (2004) "Executive dysfunction in cocaine addiction: evidence for discordant frontal, cingulate, and cerebellar activity", *Journal of Neuroscience* 24(49): 11017–11022.

Heyman, G. (2009) *Addiction: A Disorder of Choice*, Cambridge MA: Harvard University Press.

Hohwy, J. (2013) *The Predictive Mind*, Oxford, UK: Oxford University Press.

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

Hyman, S. E. (2005) "Addiction: a disease of learning and memory", *American Journal of Psychiatry* 162(8): 1414–1422.

Kalivas, P. W. and Volkow, N. (2005) "The neural basis of addiction: a pathology of motivation and choice", *American Journal of Psychiatry* 162(8): 1403–1413.

Levy, N. (2016) "Addiction, autonomy and informed consent: on and off the garden path", *Journal of Medicine and Philosophy* 41(1): 56–73.

Levy, N. (2017) "Implicit bias and moral responsibility: probing the data", *Philosophy and Phenomenological Research* 94(3): 3–26.

Martinez, D., Narendran, R., Foltin, R.W., *et al.* (2007) "Amphetamine-induced dopamine release: markedly blunted in cocaine dependence and predictive of the choice to self-administer cocaine", *The American Journal of Psychiatry* 164(4): 622–629.

Neale, J. (2002) *Drug Users in Society*, New York, NY: Palgrave.

Nutt, D. J., King, L. A. and Phillips, L. D. (2010) "Drug harms in the UK: a multicriteria decision analysis", *The Lancet* 376(9752): 1558–1565.

Pickard, H. (2016) "Denial in addiction", *Mind & Language* 31(3): 277–299.

Pickard, H. and Ahmed, S. H. (2016) "How do you know you have a drug problem? The role of knowledge of negative consequences in explaining drug choice in humans and rats", in N. Heather and G. Segal (eds), *Addiction and Choice*, Oxford, UK: Oxford University Press, pp. 29–48.

Robinson, T. E. and Berridge, K. C. (2003) "Addiction", *Annual Review of Psychology* 54: 25–53.

Ross, D., Sharp, C., Vuchinich, R. E. and Spurrett, D. (2008) *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling*, Cambridge, MA: MIT Press.

Schultz, W., Apicella, P., Scarnati, E. and Ljungberg, T. (1992) "Neuronal activity in monkey ventral striatum related to the expectation of reward", *Journal of Neuroscience* 12(12): 4595–4610.

Schultz, W., Dayan, P. and Montague, P. R. (1997) "A neural substrate of prediction and reward", *Science* 275: 1593–1599.

Taylor, M., Mackay, K., Murphy, J., McIntosh, A., McIntosh, C., Anderson, S. and Welch, K. (2012) "Quantifying the RR of harm to self and others from substance misuse: results from a survey of clinical experts across Scotland", *BMJ Open* 2(4): 10.1136/bmjopen-2011-000774.

West, R. (2006) *Theory of Addiction*, Oxford, UK: Blackwell Publishing.

Willuhn, I., Burgeno, L. M., Groblewski, P. A. and Phillips, P. E. M. (2014) "Excessive cocaine use results from decreased phasic dopamine signaling in the striatum", *Nature Neuroscience* 17(5): 704–709.