



Craving and Control

Victor Tadros¹

Accepted: 6 March 2023
© The Author(s) 2023

Abstract

Pre-reflectively, many addicts seem either not responsible, or less responsible, for their addictive conduct, at least if they lack responsibility for their addiction. Moore believes roughly the following. Addicts lack responsibility, when they do, because addicts are unable to control their conduct. They are unable when certain modal conditions are satisfied. Moore offers different modal conditions in different places. This view can be contrasted with another – that addicts lack responsibility when they do because they act on desires that are not well integrated into their motivations. Moore doubts the significance of integration for responsibility. I argue that addicts lack responsibility when, and because, first, addictive craving is not integrated into personhood in a way that directly grounds responsibility; and second, addicts lack the ability to control their addictive conduct to the extent needed for (full) responsibility. I explain how and why modal accounts of self-control need to be refined beyond ways Moore suggests to appropriately test the ability to control oneself, whether or not they are fundamental in explaining responsibility.

Keywords Addiction · Craving · Responsibility · Self-control

1 Introduction

Michael Moore's important shelf-crushing book *Mechanical Choices* includes a deep investigation of when addicts are excused for their addictive conduct. What is most impressive about Moore's work is not only the range of philosophical issues he addresses, but that in doing so he draws on, and offers significant insight about, the substantial neuroscientific evidence. I will attempt a much smaller contribution in the same vein, but I cannot possibly do justice to the array of arguments in Moore's work.

✉ Victor Tadros
v.tadros@warwick.ac.uk

¹ School of Law, University of Warwick, Coventry, England, UK

Pre-reflectively, many addicts seem either not responsible, or less responsible, for their addictive conduct, at least if they lack responsibility for their addiction.¹ Moore's overall view of this is complex, and somewhat hard to get a grip of. But he believes roughly the following. Addicts lack responsibility, when they do, because addicts are unable to control their conduct. They are unable when certain modal conditions are satisfied. As we will see, Moore offers different modal conditions in different places. This view can be contrasted with another—that addicts lack responsibility when they do because they act on desires that are not well integrated into their motivations. Moore doubts the significance of integration for responsibility.

In contrast, I argue that addicts lack responsibility when, and because, first, addictive craving is not integrated into personhood in a way that directly grounds responsibility; and second, addicts lack the ability to control their addictive conduct to the extent needed for (full) responsibility. Lack of integration does not always excuse but it blocks one path to responsibility. Moore thinks that the modal conditions needed for certain abilities must always be satisfied for responsibility, either because abilities depend on those conditions, or because they are evidence of them. I argue that this is not so where there is an appropriate causal chain between the relevant features of the person and their conduct. In doing so I cast doubt on Moore's handling of Frankfurt cases and his arguments about the irrelevance of integration. However, I suggest that such a chain does not occur in some cases of addictive wrongdoing because craving does not ground responsibility.

Now focus on a second path to responsibility—that the person should have exercised self-control. I remain neutral about whether this idea should be explained modally (though I have my doubts). However, I explain how and why modal accounts of self-control need to be refined beyond ways Moore suggests to appropriately test the ability to control oneself, whether or not they are fundamental in explaining responsibility. Moore and others adopt tests for self-control that overpredict responsibility for wrongdoing. They need correction to ensure that we control for effects on the intrinsic basis of self-control in possible worlds. My argument reinforces Moore's conclusion that this makes experimental tests for the ability hard to perform. Nevertheless, I suggest that there are reasons to believe that addicts have limited powers of self-control despite clinical evidence that might suggest otherwise.

Before getting to the main line of argument, let me frame the paper in three ways. First, I leave aside difficult questions about the relevance of a person's responsibility for becoming an addict.² Second, addiction can have an immediate and a less

¹ I will interchange excuses and responsibility, as I think the relevant excuses are responsibility diminishing.

² The question whether people can be responsible for their addictive conduct because they are responsible for being addicts is known as the tracing problem. I am broadly sympathetic to Michael Moore's assessment of the limited significance of tracing in *Mechanical Choices: The Responsibility of the Human Machine* (Oxford: OUP, 2020) 514–517. However, at 531–533 he argues that addicts cannot be excused for lack of control if they are responsible for being addicts. He draws a parallel between acquiring vicious character traits, which cannot excuse, and addiction. I am less confident about that parallel.

immediate bearing on conduct. Heroin addiction immediately bears on taking heroin; less immediately on stealing to support the addiction. I focus on the way addiction immediately bears on conduct. Third, I focus on excuses for wrongdoing. Addictive conduct includes taking heroin, gambling, watching pornography and drinking coffee, some of which is not wrong. But some of that conduct is wrong because of its effects on others (smoking harms children of smokers, for example); promises not to act on the addiction; erosion or destruction of duty-involving personal relationships; the way the object of addiction is produced (addiction to child pornography, for example); or perhaps even the violation of self-regarding duties.

2 Craving, Desire and Personhood

Responsibility for wrongdoing is sometimes grounded in the positive causal relationship between a person and their conduct, regardless of whether they have control over it—what I will call *productive responsibility*. But if a person is not responsible in this way, they might be responsible because of a failure of *self-control*—what I will call *regulative responsibility*. This Section focuses on productive responsibility; the next on regulative responsibility.

2.1 The Sufficiency of Production

A person can be responsible because of an appropriate causal chain between the relevant pro attitudes of a person—their beliefs, desires, emotions, values, judgments, reasons, or preferences—and their conduct. Different views about productive responsibility involve different causal relations.

Productive responsibility is related to ‘mesh theories’ of free will and responsibility.³ However, there are two qualifications. First, the fact that conduct meshes with the relevant pro attitudes is insufficient for responsibility. Those properties must cause the conduct in the right way. Second, unlike standard mesh theories productive responsibility is only intended to supply a sufficient condition of responsibility for wrongdoing—where it is absent, the person may be responsible for failing to exercise self-control.⁴

An objection to mesh theories is that they imply that a person is responsible even where their conduct is relatively unresponsive to reason. That is so when the person (or the mechanism that produces the action) fails to satisfy the kinds of modal condition that reason-responsive theories of responsibility endorse, or some version of

³ Familiar examples are H. Frankfurt *The Importance of What We Care About* (Cambridge: CUP, 1988); G. Watson ‘Free Agency’ (1975) 72 *Journal of Philosophy* 205.

⁴ I think that this is the kind of view that Wolf outlines and then criticizes in *Freedom Within Reason* (Oxford: OUP, 1990) ch. 2. Responding to Wolf’s criticisms would take me too far afield. I note only that, first, she is concerned with desert in a way that I am not. And second, many of the cases that she uses to put pressure on this kind of view involve background injustice which leads a person to be the kind of person they are. But the injustice is a distraction, and once we eliminate it, the cases are much less forceful.

the ability to do otherwise.⁵ However, I doubt such modal conditions are relevant to responsibility where the causal relationship between the appropriate mental states and the conduct obtains.

Suppose a person acts according to their values, desires, preferences, and so on. Now consider Moore's suggestion for a modal version of the ability to do otherwise that he thinks makes people responsible: X is morally responsible for v only if X was able not to v and able to choose not to v .⁶ It is hard to see why the satisfaction of these conditionals make a difference to productive responsibility.⁷ If a person is responding to reason, or to what they value, in the appropriate way in the actual world, who cares about nearby worlds? What happens in the actual world seems sufficient for responsibility.

Consider:

Retributivism: I value a person's suffering because I am a retributivist, and I believe the person has acted wrongly. I spontaneously express delight at the person suffering, and that stems in the normal way from my beliefs, judgements, desires and so on.

The intuition that I am responsible does not depend on any judgement about the satisfaction conditionals. This is true for both praiseworthy and blameworthy conduct. For example, if retributivism is true, my reaction might be praiseworthy. If false, my reaction is blameworthy. In neither case does this depend on my reactions in possible worlds.

Frankfurt-style backups confirm this view and cast doubt on Moore's modal view. Moore's analysis of Frankfurt cases is as follows. He starts with the assumption that the ability to control one's conduct is needed for responsibility and cashes that out in modal terms. He recognises that this account runs into trouble in Frankfurt cases, and then retreats to the view that modal facts provide a test for the relevant abilities in non-Finkish worlds.⁸ But Frankfurt cases show that these abilities are irrelevant to responsibility. So, it is not clear why Moore's analysis of abilities, even if it is successful, is relevant to productive responsibility. Modal conditions may need satisfying for responsibility for failing to do something, or failing to prevent ourselves from doing things. That's a tough topic that I can't make progress on here.⁹ But I doubt they are needed for standard cases of productive responsibility.

To confirm this, suppose I have a chip in my brain in *Retributivism* that would have been triggered had I not expressed delight in the normal way that guarantees

⁵ See, for example, M. McKenna and C. van Schoelandt 'Crossing a Mesh Theory with a Reason-Responsive Theory: Unholy Spawn of an Impending Apocalypse or Love Child of a New Dawn' in A. Buckareff, C. Moya and S. Rosell *Agency, Freedom, and Responsibility* (New York: Palgrave, 2015) 47.

⁶ *Mechanical Choices* ch.8.

⁷ I investigate this in depth in 'Responsibility Without Modality' (unpublished ms.). See, also, D Heering 'Reason-Responsiveness, Modality and Rational Blind Spots' *Philosophical Studies*, forthcoming.

⁸ See *Mechanical Choices* 296. Non-Finkish worlds are roughly worlds where the intrinsic basis of a disposition is unaltered by the things that would trigger the disposition.

⁹ An alternative view is that absence causation rather than modality makes people responsible for failures as well as successes. See, especially, C Sartorio *Causation and Free Will* (Oxford: OUP, 2016).

my behaviour. The chip bears on the satisfaction of Moore's counterfactuals, but not on responsibility. The simple lesson to be drawn from Frankfurt cases is not just the failure of standard counterfactual principles such as the principle of alternate possibilities. They powerfully illuminate the idea that the appropriate causal relationship between a person and their conduct grounds responsibility.

Some people also object to the implications of mesh theories for willing addicts.¹⁰ But with a causal condition, such theories have attractive implications. Consider a person who likes, values, or desires acting on their addiction to heroin, but where their attitudes do not bear causally on their conduct. Some mesh theories counterintuitively imply that this person is responsible for their addictive conduct. Early Harry Frankfurt, for example, implausibly thought they are because they have second-order volitions in favour of their addictive conduct—they not only want to want the heroin; they want their wanting it to result in action.¹¹ However, if a person's pro attitudes cause their addictive conduct because they are willing addicts, it is plausible that they are responsible. Complexity arises in cases where the causal path from a person's attitudes to their addictive conduct is unusual, as well as cases where a person's attitudes make a modest causal contribution to their conduct. But whether the person is responsible in such cases just is plausibly assessed by considering the significance of the causal relationship between the person and their conduct.¹²

Against this background, let us explore Moore's view about the irrelevance of the fact that a person acts on a desire that is integrated into their general set of motivations. Moore rightly thinks that the strength of desire does not undermine responsibility or ground an excuse. The strength of a desire might strengthen the causal relationship between a person and their conduct, enhancing responsibility. However, Moore thinks this even when the desire is neither rationally formed, nor governed by a person's values.¹³ I think Moore is driven by the idea that acting on such desires is insufficient for an excuse, which I agree with—the person might be regulatively responsible. This is consistent with lack of integration contributing to an excuse by denying one route to responsibility—productive responsibility.

To support his conclusion, Moore considers Mr Ott, a Virginia school teacher, who was accused of inappropriate sexual contact with his 14-year-old stepdaughter. Neurologists claimed that he acted on new sexual desires caused by a brain tumour. Moore claims that Ott might be responsible for his conduct 'if he chose to act on such desires when he could have chosen not to.'¹⁴ I evaluate this case thus. We might doubt whether Ott's desires are integrated into his personhood in a way that

¹⁰ Frankfurt suggested that willing addicts are responsible for their addictive conduct in 'Freedom of the Will and the Concept of a Person' in *The Importance of What We Care About* (Cambridge: CUP, 1988). McKenna and van Schoelandt offer this objection to mesh theories (especially Frankfurt's) in 'Crossing a Mesh Theory with a Reason Responsive Theory' 47. Their objection to mesh theories that meshing is not necessary for responsibility is well taken. I do not claim that it is necessary.

¹¹ Frankfurt's account of second order volitions is outlined in 'Freedom of the Will and the Concept of a Person' 13–16.

¹² I'm grateful to Nadine Elzein for encouraging me to explore this.

¹³ *Mechanical Choices* 527.

¹⁴ *Mechanical Choices* 527. Moore raises doubts about his capacities at 368–370.

supports productive responsibility.¹⁵ If they are not, this leaves room for a modal test like Moore's—but that is a test for whether Ott is regulatively responsible. We will consider that issue later. If Ott is not regulatively responsible, any lack of integration of his desires into his personhood ensures that he is not responsible at all. This explains the compelling idea that Ott's brain tumour was important in grounding an excuse by making it true that his conduct was 'not really him'. This idea is surely important independently of any effect of his tumour on his self-control.

Moore offers two further cases to raise doubts about mesh theories, but neither is very convincing. One involves 'pathetic males [who] kill their female victims out of anger, disappointment, feelings of rejection, sexual frustration, and the like'.¹⁶ The other is Andre Gide's Lafcadio, who pushes an old man from a moving train, thinking he'd feel satisfaction at surmounting conventional morality, but who in the end was only disappointed. Moore plausibly claims lack of integration does not excuse these people even if they do not endorse their conduct.¹⁷

As Moore briefly suggests, these cases do not count against all versions of the idea that certain desires cannot ground productive responsibility.¹⁸ These people act on judgements, and acting on a judgement might ground productive responsibility. They may not be fully committed to those judgements, and they may be deeply flawed. But their cases are different to Ott's, at least if Ott did not act on his judgements at all. If these are cases of responsibility, and I think they are, they at most lead us to revise mesh theories to focus on a wider range of pro attitudes.

2.2 Craving

Now consider whether addicts are productively responsible. I assume that when X is addicted to *d*, X craves *d*. If so, our question is whether craving is integrated into personhood in a way that can ground productive responsibility. I doubt it.

Craving has two features: stimulation and desire. Focus on the stimulation dimension first. Itches provide one helpful comparison. Itches are not desires. They stimulate desires, such as the desire to scratch. There is something similar about satisfying a craving and scratching an itch. Cravings like itches involve physical and psychological discomfort and responding to them alleviates that discomfort. But responding to them is also pleasurable. These two features together at least partly explain acting on cravings.¹⁹ However, as with cravings our motivation to scratch itches is not fully explained in this way. They motivate action more directly. We are agitated to act in a way that cannot be fully explained teleologically.

¹⁵ Moore may disagree. See *Mechanical Choices* 529–530 where he puts extreme conditions on when a psychological feature of a person cannot ground responsibility due to its not being integrated into the person.

¹⁶ *Mechanical Choices* 339.

¹⁷ *Mechanical Choices* 533.

¹⁸ *Mechanical Choices* 340.

¹⁹ The positive and negative stimuli involved in addiction are discussed in Moore *Mechanical Choices* 547–556, drawing on the neuroscientific literature. The details need not concern us here.

But cravings are also distinct from itches in their desire-like qualities. Craving *d* involves representing *d*, just like desiring *d* involves representing *d*. Our itches are represented in our desire to scratch; they do not represent. Our cravings, in contrast, represent the thing craved. In that way, satisfying a craving is more like satisfying a desire. However, although cravings represent, unlike many other desires they do not involve judgements; not even judgements disconnected from values. Craving coffee involves no judgement about coffee at all.

Now focus on whether craving can ground productive responsibility. I doubt it. Mental features that ground productive responsibility are those connected with personhood. Very roughly, they primarily include features concerned with autonomous agency, rational decision-making, and character traits. That explains why certain desires ground responsibility—the content of these desires involves values or judgements, or are caused by features of the person’s character, that make them reflect on the person as a person. Craving lacks these features. If craving just involves representing discomfort and hedonic pleasure, where the person wants to respond to these things, it is not typically integrated into the features of autonomous agency or character that make us productively responsible. The representational feature of craving that makes it desire-like need not involve character, judgement, value, or anything else deeply involved in autonomous agency. Other than that, the main feature of cravings are the stimulations that motivate conduct. But we are not plausibly productively responsible because of mere stimulation.

We can conclude that addicts can be productively responsible for wrongdoing only if they have pro attitudes that cause their addictive conduct that are independent of craving. This will vary between cases, but it is reasonable to suppose that responding to cravings is the dominant cause of a great deal of addictive conduct in unwilling addicts. And that suggests that many such addicts are not productively responsible for wrongdoing. To this extent, they can appropriately claim that it was the addiction rather than them as a person that gave rise to their wrongful conduct.

3 Does Addiction Undermine Self-Control?

Lack of productive responsibility is not always decisive, because some unwilling addicts ought to have resisted their craving. These addicts are not productively responsible for wrongdoing but are *regulatively* responsible because of their ability to exercise self-control. This Section explores how that can be assessed.

3.1 Costs of Refraining

One approach claims that addicts are not responsible if and because refraining from addictive conduct is costly.²⁰ Just as a person is not responsible for their conduct when they succumb to a grave and credible threat, so a person is not responsible for succumbing to addiction because of the costs involved.

²⁰ See, for example, G Yaffe ‘Compromised Addicts’ in D J Coates and N Tognazzini *Oxford Studies in Agency and Responsibility Vol.5: Themes from the Philosophy of Gary Watson* (Oxford: OUP, 2019).

Like Moore, I think this idea fails to establish that addiction excuses. Suppose that X will bear some cost, h , if they refrain from v ing where v ing is *pro tanto* wrong. If h is sufficiently high, X is permitted to v (though, of course, there may be conduct which is so bad that costs to X never permit it). The costs justify their conduct, but they are responsible for it. If h is insufficiently high for justification, though, it might contribute to an excuse. But the excuse cannot just be that were they not to v , X would bear h . They need some excusing explanation for failing to do their duty. Those who point to the costs of not acting on the addiction, then, need a way of showing how these costs ground excuses. Moore suggests costs make a difference to excuses even when they are insufficient to justify, because they make a person's conduct closer to being justified, and it takes less to excuse conduct that is almost justified than that which is not.²¹ I agree.²² But, strictly speaking, this does not make costs bear on excuses; rather they bear on what needs to be excused.

3.2 Addiction and Self-Control

Although addicts sometimes have less to excuse than non-addicts their excuses rely on the possibility that addiction undermines rational agency, rather than figuring in it. Moore suggests a modal test for the degree of self-control that determines when this is so. However, he is not completely clear about the right modal test.

Here is one test he suggests. The ability to control oneself is a scalar notion. Suppose that X vs. The strength of X's ability to refrain from v ing depends on the range of worlds where X refrains from v ing, where relevant worlds are those where X desires not to v . Increase the strength of X's desire not to v in different worlds. If X refrains from v ing in worlds where their desire not to v is relatively weak, they are strongly able to control themselves. If X does not refrain from v ing in worlds where their desire not to v is very strong, they are weakly able to control themselves. Moore adds the following caveat to this picture: worlds where X fails to control themselves only ground an excuse where their failure is not attributable to moral defect.²³

If this is the right test, we face an empirical challenge to the idea that addicts are excused based on their lack of self-control: most addicts refrain from acting on their addiction when the costs are sufficiently high or even where the costs are relatively low if those costs are vividly presented. Many conclude that addicts can control themselves and seek alternative explanations for the persistence of addictive behaviour.

This, I will argue, is a mistake. It is wrong to conclude from the fact that a person controls themselves in high and vivid cost settings that they are able to do so in low/non-vivid cost settings. Here are some examples of the mistaken inference. After noting that addicts often refrain from acting on their addiction, even in a sustained way, if the stakes are high enough, Richard Holton and Kent Berridge conclude that 'addictive urges are not entirely uncontrollable: they can be controlled, at least

²¹ *Mechanical Choices* 319–322.

²² See V Tadros 'Duress and Duty' in S Bazargan-Forward and S C Rickless *The Ethics of War* (Oxford: OUP, 2017) 96.

²³ This is a brief summary of Moore's view in *Mechanical Choices* 350–355.

for a short while, and sometimes for longer if the stakes are high enough and clear enough.²⁴ The last statement is fine, as it is restricted to the ability of addicts to control acting on their urges *if the stakes are high enough and clear enough*. But they wrongly infer that, ‘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. Our moral evaluations should reflect this fact, and our scientific account of addiction should explain why resistance is difficult and why failure happens when it does.’²⁵

They note a range of factors that make it more likely that a person will exercise self-control, including the significance and immediacy of the costs of failing to do so—a person is much less likely to exercise self-control, they suggest, if doing so is needed to avoid vague long-term goals such as good health, and much more likely to do so if doing so is needed to avoid significant immediate costs, such as arrest. The implicit view is that the difference in incentives makes no difference to an addict’s ability to control themselves but makes a difference only to their tendency to exercise self-control. This view, I will argue, is unsound, and so we cannot infer general claims about the ability of addicts to control themselves from the fact that they do so in high or vivid cost settings.

Similarly, Hannah Pickard presents evidence that addicts refrain from addictive conduct when sufficient incentives are present, and visible. And they sometimes manage to control their addictive conduct at certain points in their lives—they ‘mature out’ in their early thirties as responsibilities increase, so the incentives for avoiding addictive conduct increase. From this, she concludes ‘the evidence converges to indicate strongly that addicts are able to control their consumption and choose not to use drugs in many circumstances when there is sufficient reason for them to do so.’²⁶ This statement is fine, as it restricts assessment of the ability to control themselves to circumstances where there is sufficient reason to do so. But she then goes on: ‘In other words, drug use is not straightforwardly compulsive but rather responsive to incentives.’ This conclusion is unwarranted, for it draws a more general inference about the compulsive nature of addiction from the ability of addicts to control themselves in circumstances where they are given sufficient reasons to refrain.

I will suggest a range of reasons why inferences like these are unsound, including one that draws on one of Moore’s observations. Before that, it is worth noting that these analyses conflict with the experiences of addicts. Addicts frequently report feeling compelled to act on their addiction and experience an inability to control their addiction. And addiction has been found to be more common in people with mental health conditions that pre-exist addiction which reduced powers of self-control.²⁷ Addicts often try and fail to kick their habits. Finally, as Moore notes, there is

²⁴ R. Holton and K. Berridge ‘Addiction Between Compulsion and Choice’ in N Levy *Addiction and Self-Control* (Oxford: OUP, 2013) 242.

²⁵ *Ibid.*

²⁶ H Pickard ‘Denial in Addiction’ (2016) 31 *Mind and Language* 277, 281.

²⁷ See K. D. Ersche, P. S. Jones, G. B. Williams, A. J. Turton, T. W. Robbins and E. T. Bullmore ‘Abnormal Brain Structure Implicated in Stimulant Drug Addiction’ (2012) 335 *Science* 601.

significant neuroscientific support for addiction causing a loss of self-control.²⁸ The experimental data would need to be quite compelling to lead us to doubt the effects of addiction on self-control.

Here is a clearer presentation of the argument against the idea that addicts lack self-control:

Cost-Responsiveness

P1: If the costs of addictive conduct are low, X is unable to control their *ving* only if X would not be disposed to refrain from *ving* when the costs are high, or vivid.

P2: Addicts are disposed to refrain from *ving* when the costs are high or vivid.

C: Therefore, addicts are able to control their *ving* when the costs of addictive conduct are low.

For ease of exposition, I will just talk about ‘low/non-vivid cost worlds’ to refer to worlds where the costs are both low and not vivid and ‘high or vivid cost worlds’ to refer to worlds where the costs are either high or vivid. I will assume that X is averse to the relevant costs, and that X is not productively responsible for *ving*.

P1 fails. The fact that a person would prevent themselves from *ving* in high or vivid cost worlds need not bear on their ability to refrain in low/non-vivid cost worlds. Addicts might lack sufficient self-control in low/non-vivid cost worlds to be responsible, whilst gaining greater ability to control themselves in high or vivid cost worlds.²⁹ Philosophers thus mistakenly infer from the tendency of addicts to control themselves in high or vivid cost worlds that they are able to control themselves in their ordinary daily lives where the costs are often low (at least in the short term) and non-vivid.

I will demonstrate that this argument is consistent with the best modal accounts of self-control (though I have some doubts about those accounts). Philosophers who favour modal accounts have been led astray by using the wrong modal tests. The best modal test for the ability to control oneself is probably just this: X is able to control their *ving* because of *c* in their circumstances if X would control themselves from *ving* because of *c* in these circumstances had they tried. Perhaps we need to add further conditions to ensure that people also are able try, though that has less importance in this context. Introducing incentives to try is worse than a distraction. Thus, relying on the empirical evidence to establish the level of self-control, in the way philosophers and psychologists have, is unsound.³⁰

3.3 The Failure to Test Self-Control

Here is my rough account of the ability to control oneself: a person has this ability when their power of self-control must be sufficient to prevent a potential non-rational

²⁸ *Mechanical Choices* 561, drawing on N. Volkow and M. Boyle ‘Neuroscience of Addiction: Relevance to Prevention and Treatment’ (2018) 175 *American Journal of Psychiatry* 719.

²⁹ Watson identifies something like this possibility in ‘Disordered Appetites’ 67, fn.14.

³⁰ Moore broadly agrees. See *Mechanical Choices* 361–364.

cause of action resulting in them acting. Let me flesh this out. A person, X, is subject to some non-rational potential cause of action c . Let us say that the causal power of c varies with its disposition to cause v . Craving is an example: the stronger the craving, the stronger its causal power, where its causal power is the disposition of the craving to cause actions that satisfy the craving. X's powers of self-control over c are powers to ensure that c does not cause X to v . X's powers of self-control vary with their disposition to succeed in preventing c from causing X to v when X exercises their powers of self-control, given varying causal powers of c .³¹ Our self-control is greater when our attempt to exercise self-control will result in our not acting when non-rational potential causes of action have greater power.

Different accounts of addiction focus on either the extent to which addiction strengthens the causal power of the craving for the object of the addiction, or the extent to which it weakens self-control. We need not pick between these options. In different cases, one or the other, or (probably most commonly) both, occur. To support this, note that treatment of addiction both aims to reduce the causal power of addiction (eye movement desensitization and reprocessing (EMDR) has this aim) and increasing powers of self-control (cognitive behavioural therapy (CBT) primarily focuses on this). Both kinds of treatment have some independent measure of success.³²

This picture helps us understand challenges to P1 in *Cost-Responsiveness*. One challenge is familiar from Gary Watson's work. Watson argues that a person who refrains from v ing when presented with high or vivid cost may refrain because they are compelled by the high or vivid costs. If so, the fact that they refrain does not demonstrate that they are able to exercise self-control.³³ Suppose X v s because of c , where c is a non-rational cause of v . We cannot show that c does not compel X to v by establishing that X would not v where c if $c2$. For $c2$ might just compel X to refrain from v ing. If X is compelled to v because of $c2$, X's responsiveness to $c2$ does not establish that X is not compelled to v because of c .

Now consider this response that Moore might appeal to in defence of his modal view: the power of c just is, or is revealed by, competition with other potential causes such as $c2$. So, the fact that X responds to $c2$ just demonstrates limits of the power of c on X. Whilst true this is not an adequate response. I have suggested that the extent to which X is able to control themselves depends both on X's powers of self-control and the causal strength of c . Hold the causal strength of c constant, but reduce X's powers of self-control, and X is less able to control themselves by resisting c . However, if X has limited powers of self-control in general, X will not be

³¹ This must be subject to certain familiar caveats—for example, the power of self-control might be finkish in that external factors might make it fail when we vary causal powers of c because varying those powers alters the intrinsic basis of self-control.

³² See W. Markus and H. K. Hornsvelt 'EMDR Interventions in Addiction' (2017) 11 *Journal of EMDR Practice and Research* 3.

³³ 'Disordered Appetites: Addiction, Compulsion, and Dependence' in *Agency and Answerability: Selected Essays* (Oxford: OUP, 2004) 67–69. In 'Compromised Addicts' Gideon Yaffe claims similarly that addictive conduct is unfree because addicts would not act autonomously were they to refrain from addictive conduct. However, he explains the problem by appealing to costs, and we have already seen the limits of this idea.

able to control themselves by refraining from acting on competing causes that will lead to the person not *v*ing, such as *c*2. The fact that X is caused not to *v* by *c*2 is thus consistent with their having no power of self-control over anything at all. And if they have no power of self-control, they are not able to control themselves to refrain from *v*ing; certainly, they are not able to do so in worlds without competing causes such as *c*2. In other words, the fact that X does not *v* where *c*2 is present may reveal only one aspect of the ability to control oneself—the strength of *c*—not another—X's powers of self-control.

Here is a clear illustration:

Double Addiction. Harry is addicted to both heroin and cocaine. He takes heroin in the actual world, where cocaine is unavailable. If presented with the option of either heroin or cocaine, where he cannot take both, he would take cocaine.

It might be thought that this case tells us that Harry is able to control taking heroin. However, it does so only if Harry controls his desire for heroin in the counterfactual world. And he need not. He may fail to control his desire for cocaine.

If Harry is not controlling his desire for heroin, and nothing else bears on his conduct, this case might tell us that Harry's desire for cocaine is more powerful than his desire for heroin. And if his desire for cocaine is not very strong, it reveals that his desire for heroin is not very strong either. Even thus specified, this test does nothing to reveal Harry's powers of self-control. Suppose his powers of self-control are too weak to refrain from taking either heroin where that is the only option, or cocaine where that is the only option. His choice of cocaine over heroin does not reveal this fact. Indeed, if Harry has no powers of self-control at all, he picks cocaine over heroin if he desires that more. So, his picking cocaine fails to reveal his ability to refrain from taking heroin where that is the only option, as it is in the actual world.

It follows that Moore is wrong to argue that the ability to control oneself not to perform some action in the actual world is appropriately tested by considering worlds where the person desires not to perform the action. For example, Moore tests whether a person is able to control their eating cake as follows: 'the counterfactual about X not choosing to eat cake is tested in possible worlds *close* to the actual world, save that in those worlds: X desires to remain on his diet a great deal (and perhaps a great deal more than he does in this world); and whatever else has to change to generate such strengthened desires by X has also changed.'³⁴ However, the fact that X refrains from eating cake is consistent with X lacking any power of self-control either over eating cake or dieting. If X has no power of self-control over eating cake in the actual world, they lack the ability to control themselves. So, this test does not reveal whether X was able to control eating cake.

What we need to make the test work is to specify that in the counterfactual world, X exercises self-control. But then, of course, we have no reason to include desires or reasons not to take heroin in the counterfactual. We could just ask: would Harry be able to control his taking heroin were he to fully exercise his powers of self-control?

³⁴ *Mechanical Choices* 353.

If we can assess this, high or vivid cost counterfactuals just fall out of the picture. If not, high or vivid cost worlds will not help. The use of costs might be warranted in experimental settings, for they might stimulate people to try to exercise self-control. And from this, their actual powers of self-control might be revealed. I will cast doubt on this in a moment. But even if it is right, it does nothing to vindicate the philosophical significance of these counterfactuals.

3.4 Varying Levels of Self-Control

A second response to *Cost-Responsiveness* relies on the idea that costs can alter a person's powers of self-control. Focus on an implicit assumption in *Cost-Responsiveness*: the empirical evidence shows that addicts exercise self-control in high or vivid cost worlds, and this reveals their ability to control themselves in the actual low/non-vivid cost world. This assumption is warranted only if making the costs high or vivid does not bear on either the causal power of *c*, or *X*'s powers of self-control.

Here is why. We are aiming to test whether a person is able to control themselves by testing whether their powers of self-control are sufficient to make them able to refrain from acting on their addiction, given the causal power of their addiction. But we cannot test this by considering worlds where either the causal power of their addiction is reduced compared with the actual world, or worlds where their power of self-control is increased compared with the actual world. However, in high or vivid cost circumstances, a person's powers of self-control might be greater than in low/non-vivid cost circumstances. Or, the causal power of the object of addiction might vary between low cost and high-cost worlds. If either thing is true, *Cost-Responsiveness* fails. Thus, empirical findings that are supposed to show that addicts are able to control their conduct in ordinary life fail to do so.

Focus, first, on the possibility that greater or more vivid costs alter an addict's powers of self-control. As Moore suggests, assessing this involves grasping the difference between our ability to control ourselves and our tendency to do so. There is no close relationship between these things. Suppose that *X* does not control themselves in the low/non-vivid cost world, but does so in the high or vivid cost world. We need a way of assessing whether this is due to a difference in *X*'s ability to control themselves, or a difference in *X*'s tendency to control themselves.

Because costs give people reasons for action, it is tempting to think that they bear mainly on people's tendency to exercise self-control rather than on the degree to which they have the power of self-control. But costs may also affect powers of self-control. For example, self-control might require concentration. But our powers of concentration vary with our circumstances. A person who will bear great costs if they act on their addiction may have better powers of concentration than those who will not.

To illustrate, suppose that *X* takes heroin in the absence of police officers, but not in the presence of police officers. *X* may have greater powers of self-control in the presence of police officers. Their powers of self-control depend on their ability to concentrate on exercising self-control. And their ability to concentrate might be

affected by the costs. This might be supported by studies that suggest that the effective exercise of self-control involves rehearsing one's resolutions to ensure that they are stable in the face of temptation.³⁵ The ability to concentrate on exercising self-control may be heightened by vivid perception of costs, especially high costs, which heighten attentiveness to the task of exercising self-control by rehearsing one's resolutions. Without that, other things may more easily distract the person's attention, making it harder for them to rehearse resolutions.

It is hard to identify the empirical evidence that would best test this hypothesis. But it is at least one way to interpret the finding that relatively modest rewards are effective in diverting people away from addictive conduct in clinical settings, but are much less effective in the real world, for many addictive substances, including cigarettes and cannabis.³⁶ This finding might be explained by the fact that in clinical settings, participants have heightened awareness and concentration that enhances the ability to control oneself compared with ordinary daily life. The challenge is untangling this explanation from that offered in the next subsection.

3.5 The Varying Causal Power of Craving

A further possibility is that the causal power of the cravings involved in addiction varies with the costs. Moore briefly recognises this problem, suggesting that experimental psychology accurately investigates the ability to refrain from acting on a desire only if it holds tempting desire constant.³⁷ Though he does not suggest it, this requires revising the counterfactual account he offered earlier to assess a person's ability to refrain from eating cake.

Suppose c causes X to v where the costs are low and non-vivid, c is a non-person-integrated cause of X ving, and X 's powers of self-control are invariant. X 's ability to refrain from ving given c depends on the causal power of c . We have already seen that this analysis requires qualification to avoid Watson's objection. But even thus qualified, these counterfactuals are only relevant where the causal power of c is invariant between low/non-vivid cost circumstances and high or vivid cost circumstances. I now offer evidence that this is not so.

Clinical studies report variation in the strength of cravings that depend on presentation of the costs. For example, vivid self-presentation of the long-term bad effects of cigarette smoking has been reported to reduce cravings for cigarettes. Vivid presentation of the costs, it is suggested, does not only impact on the person's powers of self-control, or their tendency to exercise them; it impacts on the causal power of addictive craving.³⁸

This empirical finding has significant intuitive appeal too. Consider:

³⁵ For discussion, see Holton *Willing, Wanting, Waiting* ch.6.

³⁶ See R. K. McHugh, B. A. Hearon, and M. W. Otto 'Cognitive Behavioural Therapy for Substance Use Disorders' (2010) 33 *Psychiatric Clinics of North America* 511.

³⁷ *Mechanical Choices* 361.

³⁸ 'Prefrontal-Striatal Pathway Underlies Cognitive Regulation of Craving' (2010) 107 *Proc Natl Acad Sci USA* 14,811.

Torture: X is being tortured to get them to reveal the passcode for their computer which will result in X losing money.

Torture 2: X is being tortured to get them to reveal the passcode for a bomb that will immediately kill X's child, who is strapped to the bomb in the corner of the room.

X has a stronger reason to resist the torture in *Torture 2* than in *Torture*, which partly explains X's greater tendency to reveal the passcode in *Torture* than *Torture 2*. But holding the torture equal, if X is psychologically normal, X also seems better able to resist in *Torture 2* than *Torture*.

There are several possible explanations. Here is a plausible one that is important in this context. In *Torture*, because the costs to X of revealing the passcode are much lower and less vivid, these do not dominate X's attention. As a result, the torture dominates X's attention. The more of a person's attention some motivating fact occupies, the greater its causal power. Thus, the causal power of the torture to motivate X to reveal the passcode is great.

In *Torture 2*, in contrast, the predicament of X's child dominates X's attention. That prevents the torture dominating their attention as completely. Thus, the torture has less causal power. The greater the causal power of some non-rational motivating fact, the more difficult it is to resist acting on it. Thus, it is more difficult for X to refrain from revealing the passcode in *Torture* than *Torture 2*. Therefore, it is unsafe to conclude from X's ability to refrain from revealing the passcode in *Torture 2* that they are able to do so in *Torture*.

Before comparing addiction, I reinforce the idea that the causal power of motivation depends on the proportion of attention it occupies. Take some potential causal fact, *c*. As other things occupy a greater proportion of the person's mind, *c* occupies less of a person's mind. As a result, other things that occupy the person's mind reduce the potential causal power of *c*. Introspection supports this idea. Imagine selecting between ice cream flavours, where you are relentlessly presented with images of a person eating one flavour that you like—strawberry—and not another that you like equally—chocolate. This disposes you to select strawberry, even though you know have no more reason to select strawberry, would enjoy strawberry no more than chocolate, and so on. This familiar phenomenon is confirmed in clinical settings concerning addiction, where altering the presentation of addictive conduct in the mind of addicts is a promising way of reducing addictive conduct.³⁹

The extent to which a fact dominates attention depends both on how important that fact is to a person, and on how vividly it is presented. *Torture 2* is designed to reflect this. The costs of revealing the passcode are very significant to X and very vividly presented. That explains why we find it compelling that X's attention is heavily focused on these costs, reducing the motivating power of pain.

Now return to addiction. A common empirical finding that philosophers have noticed is that addiction dominates the addict's attention.⁴⁰ This may be partly

³⁹ See Markus and Hornsveid 'EMDR Interventions in Addiction'.

⁴⁰ See, for example, Watson 'Disordered Appetites' 71–73; Moore *Mechanical Choices* 526–528.

because of persistent representation of addictive behaviour in an addict's mind, which powerfully triggers addictive conduct.⁴¹ Furthermore, attentive thinking about what one is addicted to increases the tendency to addictive behaviour.⁴² I assume the intuitive view that addicts lack sufficient control over these representations to ground responsibility for them. They may be able to do so with significant professional support, but this ability is limited in the ordinary world of many addicts. If the causal power of some psychological facts depends on the proportion of attention that they receive, this explains the causal power of addiction.

The fact that an addict would resist their addiction were the costs high or vivid enough does not demonstrate otherwise. For reasons I have given, the causal power of addiction may well reduce in the face of high or vivid costs, as these occupy part of the addict's attention. It may also reduce in unusual settings, such as clinical settings, where costs are presented in an unfamiliar environment in an especially vivid and direct way. Addictive craving may no longer dominate the attention of the addict as fully in these scenarios, reducing its causal power, and leaving the person able to resist. This is another reason why it is unsafe to draw conclusions about the ability of addicts to control themselves in the real world from how they perform in high or vivid cost settings—their cravings may grab more attention in the low/non-vivid cost world than in the high or vivid cost world, with a resulting variation in their causal power.

4 Conclusion

None of this implies the failure of modal accounts of the causal power of addiction or the ability to control oneself that Moore favours (though as I suggested, modal considerations are irrelevant to productive responsibility). I have doubts about whether the modal facts are more than the upshot of a more fundamental analysis of these things. As evidence, think about how we draw modal conclusions—by independently grasping the fundamental basis of the thing that the modal view is supposed to account for. But I do not rely on that idea here. Even if modal facts aren't fundamental, they can be a helpful tool for understanding the causal power of craving and the power of self-control. But to play this role, we need the right modal test.

Moore suggests two different tests for self-control. In one, we hold constant everything in the actual world, except we alter the strength of a person's desire not to take the addictive substance. We have seen more than one reason why this test is inappropriate. First, it is insensitive to differing levels of self-control—it may just reveal that the person cannot control acting on the new desire. Second, it fails to account for the fact that properties of the actual world might interact with alterations in the strength of desire in a way that leads to alterations in the intrinsic features of self-control and craving.

⁴¹ See C P Müller 'Episodic Memories and their Relevance for Psychoactive Drug Use and Addiction' (2013) 7 *Frontiers in Behavioural Neuroscience* 1.

⁴² See Holton and Berridge 'Addiction Between Compulsion and Choice' 245.

A second test can be inferred from Moore's brief remarks in the context of assessing the empirical data: modal tests must be constructed to ensure that alterations in the strength of a person's desires not to act on their addiction in nearby worlds do not affect the intrinsic properties of their cravings. This is a better test, but it is still inadequate, because it fails to account for the fact that altering the relevant desires might also affect intrinsic properties of the power of self-control.

A better modal test just involves considering how the person would have performed if they had tried to exercise self-control over their addictive conduct. We don't even need to worry about masks, at least to assess the person's ability to control themselves—a person whose powers of self-control are masked lacks the ability to control themselves. For example, suppose that an addict has strong powers of self-control, they don't try to exercise them, but if they had tried to exercise them an evil demon would have made them ineffective. This person lacks the ability to control themselves. We need to control for masks when considering general abilities or dispositions, but not when considering the particular abilities we are concerned with here.

There is a difficult question about whether those who fail to attempt to exercise general abilities are responsible where their exercise would be masked.⁴³ Perhaps the failure to try to exercise the power of self-control undermines a person's excuse even where their attempt would have failed because their powers would have been masked. This needs more careful analysis than I can give it here. I have my doubts. If the person knows that their attempt would have been masked, and that explains why they don't try, they are surely not responsible because of their failure to try. But that seems true just because of the more basic fact that their inability to control themselves ensures their failure to try has no bearing on their responsibility. But considering what the person would do were they to have strong reasons to refrain from addictive conduct is just a distraction in a philosophical account of the ability to control oneself.

There are also difficult questions about whether people need the ability to try to resist to be responsible if they fail to try. But whilst this question is theoretically important and interesting, I doubt that it has much significance for assessing the responsibility of addicts, for it is not clear how addiction impacts on a person's ability to try to resist their addiction, rather than their ability to resist if they try.

Giving people strong reasons to control themselves has some value in clinical settings—to test whether people can control themselves we have to get them to try, and providing incentives is a way to do this. But we have also seen several reasons why clinical experiments aimed at assessing the ability of addicts to refrain from acting on addiction fail to provide compelling evidence of the ability to exercise self-control in the real world. They fail to control for alterations in the intrinsic bases of the two components of that ability. The empirical data—that addicts often do resist acting on their addiction when the costs are high or vivid—is at most some evidence

⁴³ One approach involves working out whether the person's failure to exercise their ability is a cause of their conduct, which is a thorny problem. For discussion, see P. Dowe *Physical Causation* (Cambridge: CUP, 2000) ch. 6, and in the context of responsibility C. Sartorio *Causation and Free Will* 128–132 and 'Replies to Critics' (2018) 175 *Philosophical Studies* 1545, 1547–1548.

that they have the level of self-control needed for full responsibility. It is consistent with the empirical data that either a person's powers of self-control, or the causal power of their craving, varies with the size or vividness of the costs. This reinforces Moore's conclusions about the limits of current experimental data for whether addicts are excused. And it implies serious challenges in designing experiments that test the things relevant to whether addicts are excused for their addictive conduct in the real world.

In the light of that, self-reports by addicts that they experience an inability to control themselves in the real world has significance in vindicating the pre-theoretical intuition that addiction excuses because it compels. We don't know how well addicts are able to control themselves in the actual world. But even if we lack decisive evidence, the experiences of addicts suggest that their abilities are weaker than the empirical data might suggest.

Acknowledgements I am grateful to participants in the workshop on *Mechanical Choices* at Rutgers. I'm especially grateful to Nadine Elzein and Alec Walen, for excellent comments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.