UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Responsibility and Addiction: A Defense of a Control-Based Theory of Moral Responsibility

Permalink

Author

Koepke, Cami

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Responsibility and Addiction:
A Defense of a Control-Based Theory of Moral Responsibility

A dissertation submitted in partial satisfaction of the requirements for
the degree of Doctor of Philosophy

in

Philosophy

by

Cami Esther Koepke

Committee in charge:

       Professor Dana Nelkin, Chair
       Professor David Brink, Co-Chair
       Professor Matthew Fulkerson
       Professor Craig McKenzie
       Professor Manuel Vargas

2022

The dissertation of Cami Esther Koepke is approved, and it is accepted in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

Table of Contents

Acknowledgements

I want to first and foremost thank Dana Nelkin and David Brink for guiding and encouraging me throughout this project. Thank you for always finding the best in my ideas and for supporting me through all the ups and downs of graduate school, dissertation writing, parenthood, and a pandemic. Thank you also to Matthew Fulkerson and Monique Wonderly for taking the time to read papers with me and for answering my endless questions.

Other faculty at UC San Diego offered me constant support throughout my extended stay in the department. I owe a special thank you to Don Rutherford for nurturing my interest in bioethics and for general life advice. Thank you also to Eric Watkins for your willingness to discuss absolutely anything I could possibly worry about with respect to philosophy or life. And thank you to Reuven Brandt for enthusiastically discussing with me every twist and turn of my arguments and offering examples and resources when I felt I had hit a dead end in an idea.

Beside great faculty, studying philosophy at UC San Diego afforded me the opportunity to be surrounded by supportive colleagues. Thank you to my fellow graduate students who took the time to read and comment on my work, including Kathryn Joyce, Kathleen Connelly, Cory Davia, Jonathan Knutzen, Chirag Barai, and Craig Agule. Thank you especially to Rosalind Chaplin for co-working with me throughout the COVID pandemic and for providing daily motivational support during the months prior to my defense. Also, thank you to my dearest friend Elizabeth Oldenkamp for patiently reading and commenting on my work throughout all of my programs. I have greatly appreciated your insightful perspective.

To audiences at UC San Diego's Moral and Political Philosophy Seminar and Agency Reading Group and to the Northwestern University Society for the Theory of Ethics and Politics, thank you for your feedback on chapter 4.

Also thank you to all my friends who made it possible to continue working on this project through the pandemic by helping me with the care of my toddler when our usual childcare support was unavailable. Thank you to Linda Leung and Mari Adan for playing with Alex at the park every day at the beginning of the pandemic so that I could have some hours to write. Thank you also to Kerry McKenzie and Reuven Brandt, Emily Munic, "Aunt" Monica and "Uncle" Chris Halliday, Jean Dimattia, Meredith and Mark Walton, and Heather and Erik Dorthé for taking Alex in the evenings and weekends, especially during the final days of dissertation writing. Thank you also to my dear parents Dale and Delvene Otto for your endless love and for letting me and Alex stay with you for months at a time when the pandemic upended all childcare. You are truly the best Oma and Opa my children could have. Thank you also to Corbin and Francisca Koepke for letting me take over your dining room table for weeks on end and for feeding me delicious food during this time. This dissertation is truly the product of family effort!

Most importantly, this dissertation and completing graduate school was only possible through the continued support and love of my spouse Geoffrey Davis. Thank you, Geoff, for continuing to encourage me in this endeavor long after it made much sense to continue. Finally, thank you to Alexander and Elliott for the happiness you have brought to my life and for providing me the motivation I needed to finish this project already so that we can go to the park.

Vita

2003   Bachelor of Arts, Northwest Nazarene University

2006   Master of Divinity, Emory University

2013   Master of Arts, Georgia State University

2022   Doctor of Philosophy, University of California San Diego

Abstract of the Dissertation

Responsibility and Addiction: A Defense of a Control-Based Theory of Moral Responsibility

by

Cami Esther Koepke

Doctor of Philosophy in Philosophy

University of California San Diego, 2022

Professor Dana Nelkin, Chair

Professor David Brink, Co-Chair

This dissertation in part aims to use the case of addiction to consider which theory of moral responsibility best accounts for and explains the features core to assessments of blameworthiness. In this project, I argue that *Reasons Responsiveness* theories provide a better framework for tracking and explaining the features necessary to responsible agency than *Quality of Will* theories. I also show how the condition of addiction can help refine theoretical commitments made by Reasons Responsiveness theories with regard to assessing

blameworthiness for wrongdoing. Further, I give direction as to how we should think about addiction in theorizing about responsibility, including how we might conceptualize the condition, which of its features are most important in assessments of responsibility, and whether, why, and how significantly the condition might undermine responsible agency.

After clarifying my assumptions about moral responsibility in chapter 1, in chapter 2 I argue against conceptualizing addiction as irresistibly compelling addictive substance use. I also offer some paradigmatic cases of addiction that have been influential in philosophical discussions about moral responsibility and introduce complications to these cases given the richer notion of addiction I defend.

In chapter 3, I challenge the claim that Quality of Will views alone rightly account for moral responsibility in the case of Frankfurt's so-called "willing addict." I contend that the stipulations about addiction, wrongdoing, and control in the cases invite implicit assumptions that impact how the cases are often sorted in terms of blameworthiness and excuse.

In chapter 4, I contend that structural Quality of Will views depend on the feature of effort to distinguish relevantly different cases of so-called "unwillingness" with regard to wrongdoing. Yet, I argue that this dependence on effort puts the views at a loss for explaining cases in which effort is absent because responsible agency seems undermined in some important way.

In chapter 5, I argue that non-structural Quality of Will views cannot explain how addiction can mitigate blameworthiness for addiction-explained wrongdoing. Finally, in chapter 6, I offer guidance for how capacitarian Reasons Responsiveness views might explain the relevance of culpability for addiction in assessments of blameworthiness for addiction-explained wrongdoing.

Chapter 1: Introduction

1. Overview

Addiction poses interesting puzzles for philosophical theorizing about moral responsibility. In many ways, the psychology and behavior of addicted agents look akin to those of any normal actor. Addicted agents are generally sane, sensitive to means-ends relations, capable of planning, responsive to incentives, and aware of the needs of others in many types of situations. So it would seem that we should hold addicted agents morally accountable when they act wrongfully on account of their addiction just as we would with any other typical, non-addicted agent. Yet, the mental life of some addicted agents, particularly those who are severely addicted, suggests that their responsible agency is degraded in important ways that might bear on our assessments of their blameworthiness for addiction-related wrongdoing.   The ability to weigh reasons for and against substance use can be especially muddled by addiction, influencing practical reasoning about moral actions. In the heat of experiencing severe cravings, actions that seem obviously wrong to a non-addicted agent can seem somehow justified to an addicted agent. Moreover, heightened motivation can make it difficult to resist substance use. In extreme cases, addictive motivation for drug use seems to occupy attention to the point that little else is noticed. Constant wanting and distracted attention wear down resistance even if substance use is not anticipated to be pleasurable or is even outright despised in a user's own evaluations.

To complicate matters, addicted agents can hold a variety of evaluative perspectives on their own addicted condition. They can hate their condition and struggle against it daily, doing everything they can to try to prevent succumbing to drug use. After struggling against substance use, but failing repeatedly, they can resign in shame to their condition, believing that they are

unable to be anything but addicted. Alternatively, they might identify with their addicted condition, embracing their status as an "addict" and finding meaning and purpose in this identity. Many addicted agents seem to fall between these extremes. Under the sway of addiction, evaluations of addictive drug use can oscillate tremendously and addicted individuals can have difficulty maintaining a consistent perspective.

All of these features suggest that addiction can impact responsible agency, perhaps undermining it in important ways relevant for our assessments of blameworthiness. Yet, when an addicted agent acts wrongly because of their addiction, which of the above features should matter most in our assessments of blameworthiness for addiction-explained wrongdoing and why? Should it matter most that a severely addicted agent seems to experience impaired control over resisting their cravings or appreciating the considerations against drug use? Or, alternatively, should considerations of control be eschewed in favor of considerations of whether the wrongful actions reflect an addicted agent's deeply held evaluative moral stance.

Taking these questions together, my dissertation aims to consider what the psychological and behavioral impact of addiction can reveal about the features of responsible agency. My main goal is to use the case of addiction to consider which theory of moral responsibility best accounts for and explains the features core to assessments of blameworthiness. In this project, I argue that *Reasons Responsiveness* theories provide a better framework for tracking and explaining the features necessary to responsible agency than *Quality of Will* theories that look to agential features such as an agent's deep self, evaluative judgments, intrinsic desires, or the structural fit between an agent's evaluative stance and their effective motivating desires. Without appealing to an agent's capacities to track and act upon their sufficient practical moral reasons, and focusing on how these capacities can be partly impaired by addiction, theories conflate

2

importantly different types of addicted agents, extend mitigation of responsibility to undeserving agents, fail to extend mitigation to deserving agents, or entirely neglect the importance of various ways in which addiction seems to impair responsible agency. But in addition to revealing which psychological features are central to responsible agency, I also aim to show how the condition of addiction can help refine some of the theoretical commitments made by Reasons Responsiveness theories of moral responsibility with regard to assessing agents as blameworthy for wrongdoing. As I fulfill these aims, I also attempt to give direction as to how we should think about addiction in theorizing about responsibility, including how we might conceptualize the condition, which of its features seem most important in assessments of responsibility, and whether and why the condition might undermine responsible agency to some degree.

With these aims in mind, my dissertation project has three main components. The first component is to determine which conception of addiction is best for theorizing about responsible agency and for addressing questions about responsibility and addiction. Addiction has often been conceived as featuring irresistible desires that compel action. Yet, this conception has significant limitations both for theorizing about moral responsibility and for considering whether addiction might be a condition that can mitigate blameworthiness. I present and defend a conception of addiction that is superior for doing both.

The second component is to determine which theory of moral responsibility best explains the features of addiction relevant for considerations of responsibility. Part of what makes a theory of responsibility theoretically plausible is that it tracks and explains intuitively important features in hypothetical cases.[1] In philosophical tradition, the condition of addiction has been

---

[1] Or, when a theory challenges these intuitions, it can provide compelling reasons for why the intuitions might be wrong. As Fischer and Ravizza (1998) note: an "approach [to moral responsibility] should be judged by its fruitfulness in sorting through and illuminating the puzzling and difficult problem to which it is applied" (40).

treated as illustrative for highlighting the features of these theories and for contrasting and comparing their approaches to responsible agency.[2] Considered intuitions about cases of addiction are thought to illustrate and support the explanatory robustness of positions regarding the features central to assessments of blameworthiness. So, how might intuitive differences between addicted agents and non-addicted agents, as well as between various types of addicted agents, be best accounted for and explained? Continuing in this tradition, in this project I will consider different paradigmatic cases of addiction to defend the thesis that a theory of moral responsibility that takes the ability to track and act in light of sufficient prudential and moral considerations to be a central feature in assessments of blameworthiness is better suited than competitors who focus on other agential features to both track and explain when and why the condition of addiction is relevant in such assessments.

Finally, the third component of my project is to use the case of addiction to fine-tune my preferred theory of moral responsibility. Specifically, because of its complicated psychological and behavioral characteristics, addiction can help Reasons Responsiveness theories refine their commitments to be even more explanatorily robust. Specifically, I will use the condition of addiction to show why capacitarian Reasons Responsiveness theories are explanatorily superior to mechanism-based theories and how these theories might best explain the relevance of culpability in assessments of blameworthiness for a partially incapacitating condition like addiction.

The remainder of section 1 is devoted to explaining the competing views of moral responsibility that I will be considering in this project. In section 2, I will present some key basic

---

[2] e.g. Frankfurt (1971); Watson (1977, 1987, 1999a, 1999b); Wolf (1993); Wallace (1994); Fischer and Ravizza (1998); Leon (2001); Smith (2004, 332); Arpaly and Schroeder (2013, 2014); Pickard (2015, 2017); Shoemaker (2015); Sripada (2016, 2017); Brink (2021).

assumptions about moral responsibility and explain some basic concepts and distinctions that

will play a role in the following chapters. I will also give a basic overview of responsible agency

as conceived by capacitarian Reasons Responsiveness theories of moral responsibility. Finally,

in section 3, I will provide a brief overview of each chapter.

1.1 The Competitors

Philosophical theories of moral responsibility might be broadly divided into two primary

groups: Quality of Will views and Reasons Responsive views.[3] In the most general terms,

Quality of Will (QW) theories of responsibility say that the foundational concern in ascribing

responsibility is an agent's evaluative orientation toward the well-being or rights of others. This

orientation reflects the agent's fundamental normative commitments, though QW views vary in

whether they conceive of this privileged orientation in terms of character, intrinsic desires,

fundamental cares, evaluative judgments, or a certain structural fit between evaluative and

motivating psychological attitudes, to name some approaches.[4] Being blameworthy for an action

or omission is just a matter of whether it reflects this moral concern. Broadly speaking, on this

approach an addicted agent is responsible for her wrongdoing iff and to the extent that the

wrongful action manifests her fundamental normative commitments as reflected in her evaluative

orientation.

While the title "Quality of Will" encompasses numerous diverse theories, in this project I

will focus specifically on two influential versions of this approach. First, *structural* QW views

---

[3] These groups are typically associated with *compatibilist* theories of responsibility. I will say more about this in section 2.5 below. Also, though I refer to this broad group of theories as Quality of Will views, they are given various names. For instance, Sripada refers to this general category as "deep self" views (2016, 1204-5).

[4] Representatives of these positions are held by or discussed in Moore (1990, 2010); Wolf (1993); Arpaly and Schroeder (2014); Sripada (2016); Smith (2005, 2007); and Frankfurt (1971), respectively. See Talbert (2016) for a summary of other positions broadly classified under the Quality of Will heading. These categories are not mutually exclusive.

look specifically at the structural fit between an agent's evaluative orientation and her motivating attitudes that successfully drive her action. Depending on the view, the evaluative orientation might be understood as a *higher-order desire*, an *evaluative judgment* about the goodness of one's desires, or even *a higher-order intention*. Agents are responsible for actions that result from the right structural fit, i.e. an agent's evaluative orientation coheres or meshes with what she is successfully motivated to do. These views consider this fit to be both a necessary and sufficient condition for an agent who acts wrongly to be assessed as blameworthy. The general structural QW view of blameworthiness and excuse might be captured as this:

> **Blameworthiness (Structural QW):** An agent is blameworthy for an action if and only if and to the extent that the action is an instance of wrongdoing and the agent identifies with the motivating desire that drives the action.

> **Excuse (Structural QW):** An agent's blameworthiness for an action is excused if and only if and to the extent that the action is an instance of wrongdoing and the agent does not identify with the motivating desires that drive the action.

Second, *non-structural* QW views look specifically at an agent's privileged evaluative attitude that is expressed in or motivates the action in question. The privileged evaluative attitude might be conative, such as fundamental moral cares or intrinsic moral desires. Alternatively, it might be cognitive, such as an agent's evaluative judgments. An agent is responsible for actions that reflect or are sufficiently motivated by this fundamental evaluative stance. The commitments of the view might be articulated as this:

> **Blameworthiness (Non-Structural QW):** An agent is blameworthy for an action if and only if and to the extent that the action is an instance of wrongdoing and it reflects or is sufficiently motivated by the agent's fundamental evaluative stance with respect to the rights or well-being of others.

> **Excuse (Non-Structural QW):** An agent's blameworthiness for an action is excused if and only if and to the extent that the action is an instance of wrongdoing and it does not reflect or is not sufficiently motivated by the agent's fundamental evaluative stance with respect to the rights or well-being of others.

Reasons Responsive (RR) theories, again in the most general terms, say that agents are responsible only if and to extent that they possess sufficient control with regard to the wrongdoing. Control is conceived in terms of a responsivity to reasons. As I will explain in more detail in section 3 below, RR theories tend to divide reasons responsiveness into two distinct but mutually supporting capacities: a cognitive capacity and a volitional capacity. *Capacitarian* RR views, the theory I will be defending in this project, treat these two capacities as involving numerous agent-based abilities by which an agent can do the right action for the right reasons. The basic commitments of the view can be captured as this:

> **Blameworthiness (Capacitarian RR):** An agent is blameworthy for an action only if and to the extent that the action is an instance of wrongdoing and the agent is reasons responsive with respect to that action.
> **Excuse (Capacitarian RR):** An agent's blameworthiness for an action is excused if and to the extent that the action is an instance of wrongdoing and the agent is not reasons responsive with respect to that action.

2. The Concept of Moral Responsibility

Broadly speaking, moral responsibility concerns being responsible for actions related to morality and moral expectations, such as fulfilling or neglecting moral obligations, acting rightly and wrongly, doing good or bad, or being virtuous or vicious.[5] Being *responsible* for a moral action generally means two things. First, it means that the acting agent satisfies certain conditions in relationship to a moral action, such as possessing and being able to exercise some key agential psychological ability or power with regard to the action.[6] Second, it means that certain moralized responses to the agent in question are justified as a reaction to the moral action. Paradigmatic responses include blaming and praising. The notion of *moral* responsibility is

---

[5] c.f. McKenna and Pereboom (2016, 11).
[6] Talbert (2019).

contrasted with other notions of responsibility, such as *causal* responsibility or *role* responsibility.[7] For instance, on the former notion, someone or something might be causally responsible if they play the right explanatory role in a chain of causal events, e.g. a fallen rock might be causally responsible for causing traffic delays. These other notions of responsibility do not alone justify *moralized* responses, such as blame in the sense that I will be discussing here.

2.1 Accountability and Attributability

Many theorists of moral responsibility assume there are at least two different notions of moral responsibility.[8] Following Gary Watson (1996), a commonly held distinction is between *attributability* and *accountability* notions. One way to understand this distinction is that each notion assumes different agential conditions for being responsible, *and* each assumes different blaming responses to be justified when the respective agential conditions are met. Being responsible in the *attributability* sense means that a moral action reflects some key agential feature of the actor such that the action can be said to reflect upon her agential self, such as her character, moral cares, or "real" self.[9] In turn, meeting this attributability agential standard licenses limited blaming responses, such as making a judgment about the agent's self or character in light of the action. In contrast, being responsible in the *accountability* sense requires a more robust agential standard, namely control over the action, and licenses blaming responses that hold the agent accountable for the action, including responses such as demanding a justification for the action or imposing sanctions. One approach to why more robust agential standards are required for accountability blame is that fairness demands that an agent must have

---

[7] See Hart (1968, 210-37) for a helpful description of various notions of responsibility.
[8] Some theorists propose many more notions. Shoemaker (2011) proposes three, while Fischer and Tognazzini (2011) propose even more. However, other theorists, such as Sripada (2015, 2016), reject the idea that there are distinctions between types of moral responsibility.
[9] Talbert (2016, 49).

control over a wrongful action in order for accountability blaming responses to be imposed.[10] Being the target of accountability blame can be unpleasant or personally costly, and it is only fair to hold someone accountable for a wrongdoing when they had the fair opportunity to avoid that wrongdoing.

The attributability and accountability notions of responsibility are taken to be conceptually distinct by many. Some theorists, though, hold that being responsible in the attributability sense is a necessary though not sufficient condition for being responsible in the accountability sense (e.g. Arpaly and Schroeder 2014). Further, some theorists clearly situate their view according to one notion of responsibility and not the other. For instance, in their attributability account, Nomy Arpaly and Timothy Schroeder (2013, 2014) explicitly seek only to offer the conditions for attributability both in their theory of responsible agency and in the sort of blame they think this agency can license. However, other theorists I discuss defend a view of responsible agency that traditionally aligns with attributability while also holding that the satisfaction of these agential conditions is sufficient for accountability blame.[11] Chandra Sripada (2016, 2017) and Angela Smith (2008, 2012), for instance, argue that meeting the attributability agential standard justifies the sort of responses often associated with accountability, such as sanctions.

As I will clarify below, in this project I am concerned with accountability blaming responses. The view of moral responsibility that I defend – capacitarian Reasons

---

[10] This fair opportunity view of accountability is supported by responsibility theorists like Wallace (1994); Brink and Nelkin (2013); Nelkin (2015) and Brink (2021). Alternatively, we might think that accountability blame requires agential control because agents who fail to meet the agential standards of accountability cannot meet the conditions for "intelligible moral address" (Nelkin, 2015, 374 summarizing Watson's (2011, 308) view.
[11] Such views are variously referred as "new attributionism" theories (Talbert, 2016, 52-61) and "pure or predominant attributionism" theories (Brink, 2021, 47).

Responsiveness – assumes a notion of responsible agency associated with accountability. However, I will be responding to views that defend a notion of responsible agency often associated with attributability.

2.2 Blame

Again, generally speaking, being morally responsible for an action means that one is a legitimate target for certain moralized responses, such blame and praise. Blame is thought to be justified when an agent who meets the standard of responsible agency acts viciously, wrongly, or badly, while praise is an appropriate response to acting virtuously, rightly, or good. In this project, I am primarily concerned with blame and blameworthiness because I think that the primary questions in understanding addiction's relationship with moral responsibility is whether and how the condition might impact our assessments of blameworthiness. Following this focus, from this point on I will only discuss blame and blameworthiness and set aside considerations of praise and praiseworthiness. When I use the term responsibility, I will have notions such as wrongdoing, blameworthiness, and blame in mind unless I specify otherwise.

There are numerous philosophical approaches to the concept of blame.[12] Blame might simply involve judgments about the responsible actor. The judgment might concern the quality of another's use of their evaluative capacities or how their action falls short of "the standards of excellence in human affairs" (Watson 1996, 230). Or, the judgment might merely be the intellectual observation that the wrongdoer exhibits "ill will" toward another (Arpaly 2006; Arpaly and Schroeder 2014). Alternatively, these judgments might aim to mark against or "add up" someone's moral ledger (Glover, 1970; Haji, 1998, Zimmerman, 1988). Beside, or in

---

[12] See Coates and Tognazzini (2013), Nelkin (2016) and Vargas (2013) for a thorough and helpful description of many of the views I mention here.

addition to, a judgment, blame might also consist of an experience or expression of reactive emotions, like indignation or resentment directed at the acting responsible agent (Strawson, 1962; Wallace, 1994, 2011). Alternatively, blame might involve "affective and behavioral" dispositions to act in particular ways toward a wrongdoer (Sher, 2006, 114). Other reactions might involve modifying or being disposed to modify the relationship that one has with the wrongdoer (Scanlon, 2008). Blame might also be primarily constituted by moral protest (Smith, 2013). Finally, following Dana Nelkin (2016), we might think that, "blame is to hold a wrong against someone, and this requires taking a certain stance toward the wrongdoer. In turn this requires making certain negative judgments and also being disposed in certain ways … Taking such a stance can be manifested in a variety of ways depending on the context" (605-606).

Again, whatever the exact components of blame, the notion of blame I will assume in this project is one that aims to hold the wrongdoer *accountable* for their action. As mentioned above, *accountability blame* goes beyond merely making a judgment and actually seeks to hold the wrongdoer accountable for their actions in some way. Accountability blame might involve reactive emotions like expressing resentment, demanding an apology or reparations, modifying the relationship, or sanctioning the wrongdoer.

Further, in this project I am specifically concerned with the conditions of responsible agency that are necessary to justify *accountability* blame. The reason for this focus is that, in response to addiction-explained wrongdoing, the general question in both academia and society seems to be whether holding addicted agents fully accountable is justified. Notably, while I aim to understand the agential conditions necessary for accountability blame, not all theorists I consider in this dissertation share this aim. Arpaly and Schroeder (2013, 2014), whose view I discuss in chapter 5, specifically aim to provide the conditions of responsible agency necessary

for attributability blame alone, though they think that being responsible in the attributably sense is necessary for being responsible in the accountability sense. In my challenge to their view, I will explicitly address the relevance of this difference in aims as it bears on the cogency of my own position.

2.3 Blameworthiness

Most straightforwardly, being blameworthy means that blame is a deserved response to an agent. While some views hold that the appropriateness of blaming is what makes someone blameworthy, making blameworthiness dependent on the appropriateness of blaming itself, here I will assume that blameworthiness is grounded in the independent condition of being accountable for a wrongdoing, which includes meeting the standards of responsible agency in regard to the wrongful action.[13] Again, I will defend the idea that being accountable is best understood in terms of the capacity for reasons responsiveness.

I will assume that justified blame should be proportional to an agent's *desert*, understood as the product of their wrongdoing and their responsibility for it.[14] While blameworthiness is a scalar notion, I will assume that there is a suitable *threshold* at which certain blaming responses become appropriate. This is especially the case for blaming responses that are themselves not scalar. For instance, sanctioning responses such as severing a relationship with someone is all-or-nothing. In this case, the specific threshold would set a standard for when the severity of the wrongdoing and the agent's responsibility combined are sufficient for the response. In contrast, a response such as mere emotional resentment can be scaled to the degree of blameworthiness. Finally, though blameworthiness may justify blame, I do not think that a blameworthy agent

[13] For a defense of this idea, see Nelkin (2011, 28-29 and 2016, 607) and (Brink, 2021, 137).
[14] I follow Brink (2021, 137-8) in this conception of desert.

must necessarily be blamed.[15] Other considerations might speak sufficiently against blaming.

For example, a spouse might be justified in blaming their addicted partner, but the personal

psychological cost of blaming might be sufficient reason not to blame.

2.4 Free Will and Compatibilism

Philosophical discussions of moral responsibility often take place within the context of

debates about free will since acting freely and acting responsibly are often thought to go hand-in-

hand. In this project, though, I set aside these broader metaphysical questions about free will and

am specifically interested in understanding the features of responsible agency as they bear on

considerations of blameworthiness. Admittedly, the competing views of moral responsibility

that I consider here all explicitly identify with the compatibilist tradition – namely, the approach

that holds that moral responsibility and free will are compatible with determinism. However, the

conditions of agency I consider can be taken as necessary for responsibility whether determinism

or indeterminism is true.[16] Thus, while I focus on compatibilist views, what I argue in this

project concerning responsible agency and addiction may be relevant for non-compatibilist views

as well.

2.5 Normative Theory

In this project I will be neutral as to the best moral theory by which moral obligations,

right and wrong, virtues and vices, good and bad, etc. are determined, with the caveat that I

assume that morality and the standards of morality are not merely relative to agential desires.[17]

To discuss violations of moral norms, I will typically use language such as "wrong" or

"wrongdoing" to describe the action, while being open to the idea that "bad" or "doing bad" may

---

[15] As Smith (2007) says, there is a difference between "being responsible" and "holding responsible."
[16] See Vargas (2013, 13) for a brief defense of this idea.
[17] I follow Fischer and Ravizza (1998, 76-77) and Arpaly and Schroeder (2014) in making this assumption.

be the more appropriate terminology depending on whatever turns out to be the best normative theory. I will assume that moral violations can take the form of both actions and omissions, though I will be neutral as to how best to explain the nature of omissions or accountability for omissions.[18]

3. Capacitarian Reasons Responsiveness

Capacitarian Reasons Responsiveness theories hold that meeting a certain standard of control over a wrongful or bad action is a necessary condition for being fully blameworthy for that action.[19] True to the name, *capacitarian* RR theories understand control as a capacity for reasons responsiveness. This capacity is generally divided into two sub-capacities – being able to recognize or track the sufficient moral considerations for or against a wrongful action and being able to act in light of these considerations. These two sub-capacities, which I will call *normative capacities,* are not single capacities or abilities, but rather distinctive categories that capture numerous abilities that can play a role in either tracking and appreciating moral reasons or acting in light of these reasons.[20] In this project, I will use various terms for each category. For the first category, I will use terms like "track," "understand," "appreciate," and "recognize." For the second category, I will use terms like "act on," "respond," "self-regulation," and "self-control." I will usually refer to agents possessing or lacking, in part or whole, each category of abilities as having or lacking "cognitive control" or "volitional control."

---

[18] Clarke (2014) gives interesting critique of various views of omissions and offers a compelling view of his own.
[19] Some representatives of this position include Wallace (1994); Nelkin (2011); Brink and Nelkin (2013); Vargas (2013); Brink (2021).
[20] Throughout this project, I use the terms "capacities" and "abilities" interchangeably.

As a modal concept, the capacity for reasons responsiveness is understood to come in degrees such that one can be more or less reasons responsive. *Moderate responsiveness* is taken to be the standard or threshold for being fully responsible for an action:

> **Moderate Reasons Responsiveness:** Where there is sufficient moral reason for an agent to act, she can regularly recognize the reason and conform her behavior to it.[21]

To put this standard into context, consider the two extremes. *Strong responsiveness* means having a perfect fit between the sufficient moral practical reasons in a situation and the reasons that an agent takes herself to have and can conform to, such that, "whenever there is sufficient reason for the agent to act, she can recognize the reason and conform her behavior accordingly." *Weak responsiveness* means "there is at least one situation in which there are sufficient reasons for an agent to act, and she can recognize the reason and conform her behavior accordingly." While reasons responsivity might be entirely scalar, ranging from no responsiveness to perfect responsiveness, full responsible agency is not thought to require perfect responsivity since, on such a standard, few would likely ever be fully responsible for any wrongdoing.[22] Conversely, being weakly responsive is too low of a standard. For instance, rarely being able to track and act on a type of reason in a particular type of context would be too low of a standard such that agents who intuitively lack responsible moral agency in some important way would be considered fully responsible.[23] The standard of moderate responsiveness captures the idea that, while perfection

---

[21] I follow Fischer and Ravizza (1998) in distinguishing these three levels of responsiveness. See also Brink and Nelkin (2013); Vargas (2013, 220-22); and Brink (2021).

[22] As Brink helpfully points out, if strong or perfect responsiveness is the standard, wrongdoing is per se excusing.

[23] Fischer and Ravizza (1998) are a notable exception here. While they argue for a moderate standard for tracking and understanding one's sufficient moral reasons, they support a weak standard for acting on these reasons. See Watson (2001); McKenna (2001); Brink and Nelkin (2013); and Brink (2021, 69-74) for convincing arguments against this asymmetrical weak standard in favor of a common moderate standard for normative capacities.

is not expected, an agent should be able to do the right thing for the right reasons on a regular

basis.

Meeting the moderate standard of responsiveness is a necessary condition for being assessed

as fully blameworthy for a wrongful action. We can capture the view as this:

> **RR Blameworthiness:** An agent is blameworthy for wrongful action A only if
> and to the extent that she meets the standard of moderate reasons responsiveness
> with regard to that action.

As the capacity for reasons responsiveness is typically conceived, most adults meet the moderate

standard in most situations, even if individuals sometimes fail to exercise their capacity to do the

morally right thing in those situations. This means that doing a moral wrong does not itself

indicate a failure in the capacity for responsivity. A capacity can remain unexercised.

However, some individuals do not meet even the moderate standard of reasons responsivity,

either in general or in particular with regard to the type of action in question. To the degree that

responsivity falls below the moderate standard, the responsivity-compromised agent is

considered worthy of mitigated blameworthiness. This idea can be captured as this:

> **RR Excuse:** An agent's blameworthiness for a wrongful action is excused if and
> to the extent that she fails to meet the standard of moderate reasons
> responsiveness with regard to that action.[24]

In terms of the capacity for responsible agency, blameworthiness can be mitigated in two

ways given the distinctive general capacities required for reasons responsiveness. First, it can be

mitigated when the agent falls below the moderate standard in being able to track or

appropriately weight their sufficient practical moral reasons. Second, it can be mitigated when

---

[24] This statement of excuse assumes that the agent is not culpable for falling below the moderate standard. I will say more about this in section 3.3 below.

the agent falls below the moderate standard to act on moral considerations even if they can

appropriately track them. The RR position on blameworthiness can be stated as this:

> **RR Blameworthiness*:** An agent's blameworthiness for wrongful action is
> excused if and to the extent that she fails to meet the moderate standard with
> regard to *either* tracking her sufficient moral considerations against that type of
> action or acting in light of these considerations.

An agent who is moderately responsive with regard to both abilities is said to have *normative

control* over her action.[25]

3.1 Capacitarian vs. Mechanism-Based Reasons Responsiveness

There are two general approaches to the metaphysical nature of reasons responsiveness. I

will not engage deeply in this debate and will briefly touch on it again in chapter 3, but it is

worth mentioning in some detail because one's approach matters in how reasons responsiveness

is assessed. It is worth mentioning which view I commit to here to more fully appreciate why I

defend an agent-based capacitarian approach to the Reasons Responsiveness view of moral

responsibility.

Following John Martin Fischer and Marc Ravizza (1998), one approach to responsiveness

is to understand it in terms of the specific psychological mechanism or etiological processes that

lead to the action in question (38). An agent is thought to have normative control over her action

only if the actual mechanism implicated in the action can regularly track and respond to

sufficient moral reasons to act otherwise as determined counter-factually via a consideration of

possible worlds. On this view, when an individual acts wrongly, the question is whether and to

what degree the *mechanism* itself that issued in the action is reasons responsive. The reasons

responsiveness mechanism-based view of blameworthiness might be captured like this:

---

[25] Brink and Nelkin (2013); Brink (2021).

> **Blameworthiness (RR Mechanism):** an agent is blameworthy for a wrongful action only if she acted from a mechanism that can regularly track the sufficient moral considerations against that type of action and respond in light of these considerations in relation to that action.[26]

By appealing to psychological mechanisms, theorists like Fischer and Ravizza hope to show that an *agent* need not have the ability to do other than they actually do in the circumstances to be morally accountable. However, note that identifying reasons responsiveness with a single psychological mechanism means that the focus of assessments is solely on the psychological process actually implicated in the action. To meet the standard of responsible agency as conceived by RR Mechanism, only the actual psychological mechanism at play in the wrongful act needs to be moderately responsive.

Yet, assessing responsiveness in this way excludes the consideration of the other abilities an agent could employ to do the right thing in those circumstances. To illustrate, take an example of the sort I will be discussing in the following chapters. In the moment of intense addictive cravings, the actual psychological processes implicated in resisting drug use might be considerably impaired. The agent might be experiencing significantly muddled reasoning, pain, and heightened wanting with the result that they can think about nothing other than using their preferred addictive substance. Importantly, though, in assessing whether this agent can respond to sufficient considerations to resist drug use, we naturally ask what other abilities they can exercise, even if self-control in the heat of temptation is too difficult for them to exert. For instance, we might ask whether the agent has the sufficient ability to avoid the situations that initiate the intense cravings. Could they have, without excess difficulty, skipped the party where

---

[26] Again, Fischer and Ravizza (1998) argue for a weak standard for responding to sufficient moral considerations. Given my assumption of a moderate symmetrical standard for normative capacities explained in section 3, I will set this complication aside.

it was obviously known that addictive substances would be present? Or does the agent have the

sufficient ability to acquire and regularly take a prescribed medication to dampen cravings? By

intentionally narrowing the focus of assessments to consider only the actually employed

psychological process, mechanism RR views exclude a consideration of all the various abilities

that agents can plausibly use to do the right things for the right reasons, even if an agent does not

actually exercise these abilities in the relevant situation. This analysis suggests that, to be

explanatorily robust, rather than focusing on the single specific psychological processes

implicated in an action reasons responsiveness should be understood in terms of the capacities of

an agent, thereby taking into account the responsivity "ability tool set" that an agent could draw

from to track and act upon their sufficient moral reasons.[27]

  With this major concern in mind, a second approach to reasons responsiveness looks

broadly at the *agent* in question and the various psychological abilities that she might exercise in

a situation to do the right thing for the right reasons. This agent-based approach articulates

blameworthiness like this:

> **Blameworthiness (Capacitarian RR):** an agent is blameworthy for a wrongful
> action only if and to the extent that she has the abilities required to regularly track
> the sufficient moral considerations against that type of action and respond in light
> of these considerations in relation to that action.

  I am not committed here to any particular metaphysical story about the nature of

capacities, but I will assume here that there is a reasonable metaphysical account. One common

way to draw the distinction between "won't" and "can't" is by appealing to how an individual

would act in relevantly similar counterfactual situations in which they have sufficient reason to

---

[27] Theorists have raised other concerns with the mechanism-based approach to reasons responsiveness as well. See
Watson (2001); Nelkin (2011, 64-79); McKenna (2013 151-83); Vargas (2013, 199-233); Brink (2021, 67-69).

act or refrain from acting. To have a capacity means that an agent is typically or regularly

responsive across a specified range of these counterfactual situations. Though, more needs to be

said to specify notions like "relevantly similar" and "typically."[28] Another way is to conceive of

abilities or capacities is in terms of what an agent could do in near possible worlds at the same

time that the ability is exercised in this world.[29] Alternatively, Nicole Vincent (2013) argues for

a diachronic approach in which the presence of a capacity is determined by how an agent acts in

actual relevantly similar scenarios encountered over time (188-90). I will use modal terms such

as" capacity," "ability," and "power" interchangeably though I recognize that some theorists

make fine distinctions between these terms.[30]

3.2 Situational Control

Across RR views, normative control is treated as a necessary and sometimes sufficient

condition of responsible agency. Some views, however, add additional necessary conditions for

an agent to be blameworthy for a wrongdoing. In particular, David Brink (2021) and Brink and

Nelkin (2013) add the addition of *situational control* as a necessary condition. Together,

possessing normative control and situational control are treated as sufficient for blameworthiness

for a wrongdoing. Possessing sufficient situational control means having the right control over

external factors in a situation. For instance, being threatened with serious harm for not engaging

in a wrongdoing, as in the case of duress, would mean that an agent lacks sufficient situational

control (Brink, 2021, 74).

---

[28] See Vargas (2013), Coates and Swenson (2013), or Smith (2003) for some approaches.
[29] Fischer and Ravizza (1998) and Vargas (2013). Though, of course, Fischer and Ravizza understand this in terms of mechanisms and not agents.
[30] Clarke (2009 and 2015); Whittle (2010); Cross (2012).

I am open to the idea that situational control is a necessary condition for being responsible for wrongdoing. Further, I recognize that discussing the features of addiction in the context of situational control is a fruitful topic.[31] In this project, though, I am specifically concerned with the conditions of responsible agency and will set aside the discussion of situational control. To leave open the possibility that situational control is a necessary condition of responsibility, I will treat normative competency as only a necessary condition of being responsible as well.

3.3 Culpability for Incapacity

As I will show in this project, addiction can sometimes sufficiently undermine responsible agency by affecting both cognitive and volitional control with regard to addiction-explained wrongdoing. To put it another way, addiction can *incapacitate* some individuals such that they fall below the standard of moderate reasons responsiveness with regard to addiction-explained actions. However, it is often assumed that this incapacity is not sufficient for mitigating blameworthiness when the acting agent is responsible for the incapacity. In the context of addiction, this means that even if addiction undermines normative control to a significant degree in regard to an addiction-explained wrongdoing, an addicted individual might still be blameworthy for that wrongdoing if they are culpable for their addiction.

Culpability for one's own addiction is an important issue. On the one hand, there is the question of whether addicted agents are actually culpable for becoming or remaining addicted. On the other hand, if the answer to the first question is "yes," there is the further question about the best way to account for this culpability in an instance where the condition directly explains a wrongdoing. I devote all of chapter 6 to considering these questions and addressing some of the

---

[31] See Watson (1999b) and Brink (2021, 352-354) for one such interesting discussion.

resources that capacitarian RR theories might employ to account for the possible relevance of culpability for incapacity.

In chapters 1-5, however, I will set aside the complication of culpable addiction and assume in all of the cases that I consider that the addicted agents are *not* culpable for their addiction. There are several plausible scenarios in which agents might be non-culpably addicted. For instance, an addicted agent might have acquired their addiction in childhood when they did not meet the moderate standard of normative control and might have since lacked the opportunity to overcome their addiction due to the unavailability of resources necessary for recovery. As such, in chapters 1-5 my primary questions will be whether addiction can mitigate blameworthiness by degrading normative control and which theory of moral responsibility best accounts for this mitigation.

4. General Outline of Project

In chapter 2, I motivate a move away from viewing addiction as a condition that irresistibly compels the use of addictive substances. I introduce and defend a notion of addiction assumed throughout the dissertation and explain the nature of the addiction explained wrongdoing with which I will be concerned. I also offer some paradigmatic cases of addiction that have been influential in philosophical discussions about moral responsibility, and introduce complications to these cases given the richer notion of addiction I defend. In chapters 3, 4, and 5, I use these paradigms of addiction to compare the merits of various Quality of Will theories against a standard capacitarian Reasons Responsiveness theory.

In chapter 3, I focus on *non-structural* QW views, which ask in assessments of blameworthiness whether an action reflects or is motivated by an agent's evaluative stance concerning the rights or well-being of others. Traditionally, the case of the so-called "willing

addict" has been thought to support QW views based on the idea that only these views offer the right extension with regard to the case. While QW views consider the willingly addicted agent to be blameworthy, RR views need not – a supposedly counter-intuitive result. However, I will contend that the stipulations about addiction, wrongdoing, and control in the cases invite implicit assumptions that impact how the cases are often sorted in terms of blameworthiness and excuse. I show that, when the cases are modified to better capture how control is often taken to be lost or exercised in addiction, capacitarian RR views not only get the right extension with regard to the case but can also track and explain important differences between cases of addiction that QW views miss.

In chapter 4, I focus on *structural* QW views. Again, these views look specifically at the structural fit between an agent's fundamental evaluative stance about the well-being or rights of others and her motivating attitudes that effectively drive her action. I contend that structural QW views depend on the the feature of effort to distinguish relevantly different cases of so-called "unwillingness" with regard to wrongdoing. Yet, I argue that this dependence on effort puts the views at a loss for explaining cases in which effort is absent because responsible agency seems undermined in some important way. I argue that capacitarian RR accounts can better track and explain why absent effort may be relevant for assessments of blameworthiness.

In chapter 5, I focus on *conative non-structural* QW views to argue that they cannot explain why addiction might be a mitigating condition for blameworthiness for addiction-explained wrongdoing. For these views, an action is blameworthy iff and to the extent that it reflects or manifests insufficient moral care or care for moral wrong or bad. Looking at a complicated paradigmatic case of addiction, I argue that conative non-structural QW views can only explain deservingness of excuse by making considerations of loss of control primary in

assessments of responsible agency. I will suggest that since capacitarian RR views already offer such an account, they should be preferred.

Finally, in chapter 6, I consider how capacitarian RR views might explain the relevance of culpability for addiction. Intuitively, even if addiction sufficiently undermines reasons responsiveness at the time of wrongdoing, whether an addicted agent is culpable for becoming addicted or remaining addicted seems to matter in our overall assessment of that agent's blameworthiness for the wrongdoing. RR theorists in general attempt to account for the importance of culpability for incapacity by calling upon either a " tracing strategy" that seeks to derive responsibility for a wrongdoing done while incapacitated at prior point in time when the agent is sufficiently normatively competent and can be reasonably expected to foresee the risk of wrongdoing or a "negligence/recklessness strategy" that seeks to hold addicted agents culpable for the act of unjustifiably risking an incapacity and its consequences.

I will show how culpability for addiction and addiction-explained wrongdoing presents a worry that these strategies double-count blameworthiness for a single wrongful outcome. In response, I will propose what I call a " circumvention strategy" that explains addiction-explained wrongdoing by reference to the proximate abilities that an agent could exercise to avoid the wrongdoing, though I will allow that the negligence/recklessness strategy can be called upon to explain culpability for addiction and the outcome of period partial normative incapacity even if the circumvention strategy best explains responsibility for addiction-explained wrongdoing.

To help the reader, I offer one final note about the ordering of the chapters. My ordering reflects my main aims for the dissertation. Again, my main aims are as follows: 1) to give direction as to how we should think about addiction in theorizing about moral responsibility and in assessing blameworthiness for addiction-explained wrongdoing; 2) to consider what the

condition of addiction can reveal about the features of responsible agency; and 3) to refine some theoretical commitments of control-based views of moral responsibility.

Chapters 2 and 3 motivate a move away from thinking about addiction as featuring irresistible desires and advocate for a view of addiction that features a partial loss of control. Specifically, chapter 2 advocates for a specific understanding of addiction while chapter 3 offers a comparison between non-structural Quality of Will views and capacitarian Reasons Responsiveness views to reveal the problems that arise with stipulating a complete loss of control in addiction for theorizing about moral responsibility.

Chapters 4 and 5 are devoted to explanatory puzzles raised by addiction to show which features are essential to assessments of responsibility and blameworthiness. While I address this aim as well in chapter 3, chapters 4 and 5 are devoted entirely to comparing and contrasting a capacitarian Reasons Responsiveness view with a specific type of Quality of Will view with regard to a particular explanatory puzzle that I think is especially challenging for the type of view on which I focus in the chapter.

Chapters 5 and 6 fulfill my more practical aim to demonstrate how addiction might be relevant in assessments of blameworthiness and excuse with regard to addiction-explained wrongdoing. Finally, in chapter 6 I fulfill my final aim to refine some commitments made by control-based views for assessing blameworthiness.

Chapter 2: The Concept of Addiction

1. The Cases of the Willing and Unwilling Addicts

In recent philosophical tradition, the condition of addiction has been treated as illustrative for highlighting and defending philosophical approaches to responsibility and free will. Considered intuitions about cases of addiction are thought to illustrate and support the explanatory robustness of positions regarding the features central to assessments of blameworthiness. Perhaps the most notable cases in this tradition are the cases of the so-called "willing" and "unwilling" addicts. First described by Harry Frankfurt (1971) in his seminal essay "Freedom of the Will and the Concept of a Person," these cases contrast two addicted agents who are both compelled by irresistible desires to use, but who take different evaluative stances on their addiction. The unwillingly addicted agent is motivated to use a substance, but he also wants to refrain from using. Yet, while he is split in his motivation, his higher-order evaluation is that he should refrain. It is this desire that the agent "wants to be effective and to provide the purpose that he will seek to realize in what he actually does" (12). And yet, despite this evaluation, he experiences an irresistible force to succumb: "he hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, helplessly violated by his own desires" (12).

The willingly addicted agent likewise experiences addiction as an "irresistible" thrust that irresistibly forces him to substance use. The primary difference in this case, though, is that he evaluates this motivation as overall good. In contrast to the unwillingly addicted agent, the willingly addicted agent "is altogether delighted with his condition" and "would not have things any other way. If the grip of his addiction should somehow weaken, he would do whatever he

could to reinstate it; if his desire for the drug should begin to fade, he would take steps to renew its intensity" (19).

It is promising to think that exploring whether and how addiction can impact agential abilities can shine a light on features core to being a free and responsible agent. These two cases in particular have had a long and influential life in such theorizing. As the cases are often used, they are meant to illustrate that features other than control over one's actions, such as an agent's evaluative stance toward their motivation and actions, are most central to assessments of responsible agency.[32] As the thought goes, an agent can completely lack control over an action and yet still be responsible on account of their evaluative stance toward that action.

However, as stipulated these hypothetical cases make questionable assumptions about both addiction and the notion of control.[33] One concern is that this conception of addiction is not consistent with how addiction is experienced, and thus the cases do not reveal anything useful about moral responsibility for addiction per se.[34] As the Frankfurt cases depict it, addiction is a purely physiological condition chiefly characterized by truly irresistible desires that compel an agent as a "force other than his own" to act (13).[35] Yet, as I will say more about below, this is not an accurate depiction of the condition. Rather, addiction is better characterized as involving a partial loss of control that implicates many psychological processes besides motivating desires, including the ability to sustain stable evaluations or recognize things of value that might compete with substance use. Still, despite this concern, some argue that even if the stipulated notion of

---

[32] Frankfurt (1971), Wallace (1994), and Sripada (2017). Philosophers use the cases to examine other interesting aspects about agency more generally as well. See Flanagan (2017) and Kennett (2013).
[33] Wallace (1994); Kennett (2014); Pickard (2015); Flanagan (2017).
[34] Pickard (2012).
[35] Philosophers are not alone in accepting this model of addiction. It is influential in psychology as well. See Leshner (1997) for an articulation of the idea that addiction features compulsive desires.

addiction in these cases does not accurately reflect addiction, or any actually experienced

psychological condition, the hypothetical cases can still be helpful in what they illustrate about

control and moral responsibility.[36] In response, a second concern that I defend at length in

chapter 3 is that these hypothetical cases as stipulated do not depict loss of control in a way that

gives helpful intuitive information about the features most central to assessments of responsible

agency and blameworthiness. The notion of control assumed in these case is so intuitively

implausible, it invites implicit assumptions about existing control over substance use despite the

stipulation that control is lost.

If addiction is not best characterized as featuring an irresistible physiological force that

compels action, how should we conceive of the condition in our theorizing about responsible

agency? The goal of this chapter is to answer this question by presenting a notion of addiction

that will serve as the basis for the chapters that follow. Modifying a conception of addiction

defended by Walter Sinnott-Armstrong and Hanna Pickard (2013), I will argue that, for

theorizing about responsible agency, addiction should be viewed as *a strong and habitual want*

that significantly reduces control and significantly risks considerable harm.

In this chapter, after first discussing some general assumptions that I will be making

about addiction, I will consider the notion of substance-related disorders presented in the

Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM–5; American Psychiatric

Association, 2013) as one possible candidate for a notion of addiction useful for theorizing about

responsible agency and responsibility for addiction-explained wrongdoing. I will reject this

notion on the grounds that it conflates harmful, self-defeating use with use that exhibits a loss of

---

[36] Sripada (2016, 2017).

control. I will then consider Sinnott-Armstrong and Pickard's (2013) definition of addiction.

While I agree with them that strong and persistent wanting is a core characteristic of the

condition, and will offer additional explanation of and support for this feature, I will clarify how

addiction might be thought to compromise control, especially in severe manifestations of the

condition. I will also consider the role that harm might play in the definition of addiction.

Finally, I will revisit Frankfurt's cases of the willingly and unwillingly addicted agents and

suggest some ways the cases might be modified to be more helpful in both depicting addiction

and in theorizing about the core conditions of moral agency.

2. Assumptions About Addiction

Discussing the nature of addiction has a rich and interesting history. Given the focus and

limitations of my project, I will bypass many of the historical and contemporary debates about

the concept and make some starting assumptions. Chiefly, I will assume that there is such a

thing as addiction and that the condition is comprised of a subset of psychological, physiological,

and behavioral features.[37] In this section I will briefly describe my assumptions concerning how

we talk about addicted agents, the addictive behaviors I want to focus on, the level of description

that I will be using to discuss the condition, and the nature of the wrongdoing associated with the

condition.

2.1 Describing Addiction

---

[37] As with other supposed psychological illnesses or conditions, there is a vocal group of thinkers who doubt there is
really anything "real" that is addiction (Davies, 2013). At the other extreme, others think that addiction is a very real
condition that transcends normative constructs. Here I am neutral as to whether the condition is socially
constructed, a natural psychological kind, or something in between.

One note about terminology. Recent editions of the DSM mostly eschew the term "addiction" and "addict" and instead use terms like "substance-related disorder" and "substance abuser," though the term "addiction" is used in the context of gambling, which is labeled as an "addictive disorder" (DSM-5, 481-590). Here, though, I will follow many contemporary addiction theorists and use the term "addiction," though I appreciate that this term often carries moral stigma.[38] Like the term "addiction," the term "addict" has been steadily replaced in both clinical and academic contexts with more person-centered references such as" person with an addiction or substance use disorder."[39] I will attempt to follow this new convention here by using terms like "addicted agent" rather than "addict." Note, however, that many contemporary researchers and addicted agents themselves continue to to use the term "addict."[40] Thus, the term may still appear when I am citing other research or biographical statements from addicted individuals.

2.2 Addiction: Substances and Processes

While the term addiction is commonly applied to behaviors that range from substance use to food consumption to exercising, in this project I focus on substance-based addictions only since these are commonly considered the paradigm of the condition. I will remain open to the possibility that there are non-substance behavioral or "process" addictions as well, such as an addiction to gambling.[41]

---

[38] I suspect that since the condition of addiction itself, no matter what we call it, is usually morally stigmatized, any new term used to describe the condition will eventually be problematic as well on these grounds

[39] Thank you to Mary Devereaux and Christiana Eltiste for drawing my attention to this important shift in in how we talk about addiction and the people who suffer from the condition.

[40] Pickard (2020).

[41] See the DSM-5 (2013, 585-595) and Ainslie (2013a, 2013b) for two interesting approaches to gambling as a possible addiction. See Dill and Holton (2013) for an interesting debate about the definition of addition and the inclusion of non-substance behaviors within the extension of this concept.

2.3 Addiction and Levels of Explanation

Another current controversy centers on whether addiction is best understood and explained as primarily a social condition, an agential condition, or a biological condition.[42] Here, I will treat addiction as an agential condition and focus on "person-level" descriptions. This means that I will mainly discuss addiction in terms of psychological attitudes, subjective experience, and behavior.[43] However, I will mention biological-based explanations where doing so might helpfully inform the understanding of person-level mental states, experiences, and behaviors.[44] Also note that I fully recognize that the experience of addiction can vary depending on the target substance. For instance, the experience of being addicted to an opioid is different than an addiction to alcohol or a stimulant like cocaine. These differences can extend to differences in qualitative subjective experiences of cravings, the experience of intoxication, and the experience of withdrawal, to name a few possibilities. In this project, my focus is on addictions to opioids like heroine or fentanyl, stimulants like cocaine or methamphetamine, and alcohol. While there are important difference between addictions to these substances, research

---

[42] See Alexander (1978, 2018), Hari (2015) and Graham, et al. (2008); Flanagan (2011) and Kalant (2010); Kalivas and Volkow (2005) for examples of each explanatory approach, respectively. See Levy (2007) for a defense of a mixed notion.

[43] Even with this person-level focus, I do not mean for it to capture the fullness of what is involved with addiction. As Flanagan (2011) plausibly argues, a full explanatory picture of addiction involves understanding both how addiction is experienced by individuals, what he called the first-person phenomenology, and the objective third-person scientific perspective of addiction, which includes perspectives from philosophy, evolutionary biology, psychology, psychiatry, anthropology, sociology, and neuroscience.

[44] The context of the debate here is that some vocal researchers who work primarily on biological mechanisms implicated in addiction assume that a biological mechanism-only approach is the best way to understand addiction. See Flanagan (2011) for an argument that a "person-level" description is most apt for addiction and Kalant's (2010) challenge to the idea that neurobiology can capture the whole of the concept of addiction. Plus, as Sinnott-Armstrong and Summers (2018) argue, even researchers working solely on biological mechanisms have a person-level conception of addiction in mind that guides their work. So a clear person-level definition is a helpful starting point for even those working at the descriptive level of biological mechanisms. Ainslie (2000, 77) makes a similar point.

suggests that there are some core similarities, such as heightened motivation for use and uncomfortable or painful withdrawal symptoms upon cessation of use.[45]

2.4 Addiction and Wrongdoing

I do not assume that the pursuit or use of addictive substances, even when one is addicted, is necessarily a wrong in itself.[46] Severe addiction, though, can be a primary explanatory cause in a host of more obvious moral wrongs. Even if most addicted individuals do not set out to hurt others, the significantly heightened and narrowed focus on the pursuit and consumption of addictive drugs can lead to wrongful actions and failures to do the right thing.[47] Many wrongs are minor, say, not giving others the full attention they deserve and need or misleading others about one's substance use. Owen Flanagan (2011), for instance, describes his own experience at the birth of his first child. He recognized it as one of the most meaningful moments in his life but also simultaneously resented it because he could not drink while at the hospital. His attention was constantly pulled away from where it should have been, and his mental absence was likely felt by those who needed him to be fully mentally present in the moment.

Other wrongs explained by addiction can be far more considerable. Take the case of Mandy McGowan who overdosed in a dollar store while shopping for diapers with her toddler.[48]

---

[45] See Pober's (2012, 2013) argument that there are many significant differences in the underlying biological mechanisms associated with addiction to assumed addictive substances as well. See also Nestler (2005).

[46] Schroeder and Arpaly (2013; 2014) and Brink (2021). See Husak (especially 2004 but also 1999 and 2000) for an interesting discussion on the possible moral and legal wrongdoing of drug consumption as opposed to the condition of being addicted or the wrongdoing that might follow from either. In chapter 6, I do suggest that using addictive substance can be a wrongdoing when an agent is aware of their genetic propensity toward addiction.

[47] See Crowley, Connell, Jones, and Donovan (2019) for an attempt to quantify the cost of opioid abuse on children via the number of children removed from their homes due to neglect and abuse directly explained by addiction.

[48] Following a recent trend by police departments attempting to communicate the impact of addiction on their communities, McGowan's overdose, captured by a bystander, was released for public viewing by the Lawrence

Prior to picking up her young daughter from a caretaker, McGowan consumed a lethal dose of fentanyl, a synthetic opioid, the effects of which paramedics were fortunately able to reverse after she collapsed in the store aisle with her daughter desperately crying at her side. Unfortunately, McGowan's situation is not that uncommon. Take David Carr's experience that he recounts in his autobiography *The Night of the Gun*.[49] Addicted to cocaine and experiencing cravings, Carr was torn between acting on his addictive desires for substance use and caring for his infant twins. Wanting to do the right thing, but mentally and physically overwhelmed by addictive desires, he left the twins in their family car alone in sub-freezing weather in order to use intravenous cocaine:

> I saw two sleeping children, the fringe of their hoods emerging in outline against the backseat as my eyes adjusted to the light. Teeny, tiny, itty-bitty, the girls were swallowed by the snowsuits. We should not have been there. Their mother was off somewhere, and I had been home looking after them. But I was fresh out. I had nothing … I could not bear to leave them home, but I was equally unable to stay put, to do the right thing. So here we were, one big, happy family, parked outside the dope house. It was late, past midnight. Then came the junkie math; addled moral calculation woven with towering need. If I went inside the house, I could get what I needed, or very much wanted. Five minutes, ten minutes tops … Sitting there in the gloom of the front seat, the car making settling noises against the chill the math still loomed. Need. Danger. A sudden tumbling? Naw. Nothing to it, really. In that pool of darkness, I decide that my teeny twin girls would be safe. It was cold, but not really that cold. Surely God would look after them while I did not … Leaving, I remember that… How long had it been, really? Hours not minutes. I walked toward the darkened car with drugs in my pocket and a cold dread in all corners of my being. I cracked the front door, reached around, unlocked the back, and leaned in. I could see their breath. God had looked after the twins, and by proxy me, but I realized that moment that I had made a mistake He could not easily forgive. [Carr, 2008, 162-165]

Police Department. See Woolsey and Willingham (2016) for a discussion of the motivation behind this policy. McGowan's case is discussed in news articles by Seelye (2016, 2018) and Germano (2016).

[49] Arpaly and Schroeder (2013, 2014) use this insightful example.

Whether we describe McGowan's and Carr's wrongdoings as inflicting unnecessary emotional harm on their children, taking unnecessary risks with their children's well-being, or neglecting to prioritize their children above their substance consumption, it is clear they acted wrongly.

In the cases I consider in the dissertation, my focus is on instances where addiction, especially the pursuit and use of an addictive substance as a consequence of addiction, is directly and immediately involved in the explanation of the wrongdoing. The simplest case of addiction-explained wrongdoing is one in which the wrongdoing directly results from a desire to use a substance or a failure to resist substance use. In Flanagan's, McGowan's, and Carr's cases, their failure to resist cravings or their substance use plays a considerable and direct explanatory role in why they do what they do or fail to do what they should.

However, there are many other cases of wrongdoing where addiction is a central explanatory factor in explaining why the wrongdoing occurs, but the condition is not the direct or immediate cause of the wrongdoing. Take for instance the fact that in 2015 alcohol was implicated in nearly 50% of all cases of death from liver disease.[50] It is plausible that it is morally wrong to unnecessarily and unreasonably increase competition for a scarce medical resource, like a transplantable organ, that is necessary to prevent untimely death (Glannon, 2009). Addiction frequently plays an important role in the explanation of why individuals develop alcohol-related end-stage liver disease, even if it is only indirectly implicated in the wrongdoing for unjustifiably increasing competition. Still, determining addiction's precise role in the wrongdoing is less straightforward than its role in, say, McGowan's or Carr's wrongdoings.

---

[50] Yoon and Chen (2018).

Further still, there are cases where addiction plays some explanatory role in the wrongdoing, but actions other than pursuing or using drugs constitute the wrongdoing. Such examples might be theft or assault where addiction explains why the actions are pursued, but there are many psychological elements not directly impacted by addiction that are more proximately involved in the wrongdoing. Such an example might be theft done in order to fund their purchase of an addictive substance. Assuming that addiction is a possible mitigating condition, if we want to understand whether it can mitigate blameworthiness for this act, we would need to determine whether and to what degree addiction is implicated in the various intentions, planning, actions, etc. related to the theft. With that said, everything I propose about the features of responsible agency and addiction apply to cases like this as well, with the note that the analysis would be more complicated than with cases where addiction is more directly implicated in the wrongdoing.

3. Defining Addiction: The DSM-5

There are many competing conceptions of addiction that could be a plausible basis for cases used in theorizing about the conditions of moral agency. One possible contender is the notion of a substance-use disorder offered in the most recent version of the Diagnostic and Statistical Manual (DSM-5).[51] As a clinical manual aimed at helping clinicians identify and treat a range of psychological disorders, the DSM's primary purpose is to identify appropriate subjects

---

[51] Note again that the DSM-5 does not use the term "addiction" but rather "substance related disorder." As the authors explain, the concern with the term of addiction is that it is especially vague and entails negative connotations: "Note that the word addiction is not applied as a diagnostic term in this classification, although it is in common usage in many countries to describe severe problems related to compulsive and habitual use of substances. The more neutral term substance use disorder is used to describe the wide range of the disorder, from a mild form to a severe state of chronically relapsing, compulsive drug taking. Some clinicians will choose to use the word addiction to describe more extreme presentations, but the word is omitted from the official DSM-5 substance use disorder diagnostic terminology because of its uncertain definition and its potentially negative connotation" (APA, 2013, 485).

for medical interventions like therapy or medication (APA 2013, 20). Yet, though the DSM is written primarily to guide clinical interactions, its conception of addiction is highly influential in addiction research and theorizing as well (Heyman, 2009).[52] Here, I will reject the DSM-5's notion of addiction because it does not make features associated with loss of control, such as heightened motivation for drug use, mental obsession with drug use, or difficultly with the cessation of use, necessary features of the condition.

3.1 Addiction as Harmful Use

A primary consideration in analyzing the strength of a conception of addiction is whether it captures paradigm cases of the condition. Another consideration is whether the characterization can demarcate those who have the condition from those who exhibit similar physical, psychological, or behavioral features but who do not have the condition. The DSM-5's concept of addiction adequately captures paradigm cases of addiction. However, as I will suggest here, the extension of the conception is too inclusive, identifying substance users who seem importantly different from each other with the same condition. As I will suggest, this confusion can be problematic for theoretical research that uses addiction to illustrate important features about human behavior or agency or for theorizing about responsibility and addiction.

The authors of the DSM-5 attempt to capture the core feature of addiction and then name a cluster of symptoms that exhibit this core feature. This "core plus symptoms" approach clearly draws a line between addicted and non-addicted individuals by setting a minimum threshold which has to be met, while also setting various thresholds to distinguish degrees of severity in

---

[52] For a recent example of where the DSM-5's conception is influential in philosophy, see Sripada (2022).

the condition. A *mild* disorder involves having two or three symptoms.[53] A *moderate* disorder

includes having four or five symptoms. A *severe* disorder involves six or more symptoms.

According to the DSM-5, the core feature of addiction is "the continued use of a

substance despite significant substance-related problems," which is further specified as "repeated

and significant distress or impaired functioning" (APA, 2013, 483, 496). This core feature is

indicated by eleven possible symptoms categorized in four general groupings:

I. Impaired Control

1. Taking a substance in "larger amounts or over a longer period than was originally intended."

2. Expressing "a persistent desire to cut down or regulate substance use" or reporting "multiple unsuccessful efforts to decrease or discontinue use."

3. Spending a "great deal of time" obtaining substance, using substance, or recovering from effects of use.

4. Experiencing "cravings for using substance, as manifested by an intense desire or urge for the drug" such that when one has an urge to take the drug "they could not think of anything else."

II. Social Impairment

5. Use results in a "failure to fulfill major role obligations at work, school, or home."

6. Continued "substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance."

7. Important social, occupational, or recreational activities" … "given up or reduced because of substance use" such as withdrawing from "family activities and hobbies in order to use the substance."

III. Risky Use

---

[53] The DSM-5 offers the following caveat. If an individual only has three symptoms and two of the three are only pharmacological, the individual should not be diagnosed with an addiction. This allows for the fact that many medications can cause tolerance and withdrawal symptoms though the symptoms are resulting in physical dependence, which does not in itself prompt treatment (APA, 2013, 484).

8. "Recurrent substance use in situations in which it is physically hazardous."

9. Failure to abstain from using a substance "despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance."

IV. Pharmacological

10. Tolerance as "signaled by requiring a markedly increased dose of the substance to achieve the desired effect or a markedly reduced effect when the usual dose is consumed."

11. Physical withdrawal symptoms indicated a decline of concentrations of the substance in the blood or tissues of an individual the frequently result in the individual using again to relieve.[54]

The extension of the DSM's core/symptoms approach nicely captures paradigmatic instances of addiction. Take McGowan's case again. The DSM-5's conception of addiction rightly identifies McGowan as having a severe addiction. She satisfies nearly every symptom and experiences all of the features we commonly associate with the condition, like conflicted motivation, unsuccessful efforts to stop use, harm to self, heavy and chronic use, and physical dependence and withdrawal symptoms. As communicated through various interviews, she obviously deeply valued maintaining a relationship with her daughter and staying out of jail, both goals contingent on her not using drugs, but she still required significant assistance from others to resist use. She tried desperately to stop use, including watching videos of her own humiliating overdose to try to muster the resolve not to use again, but she could not stop mentally obsessing about substance use. Shortly after her filmed overdose, volunteers saved McGowan from yet another overdose and organized rehabilitation for her. But within months, she experienced a series of relapses that contributed to several months in jail, homelessness, and eventual psychiatric care. She valued her relationship with her daughter above all else and

---

[54] APA, 2013, 483-485.

exerted significant effort to abstain from use, but it took years for her to start to maintain sobriety

for months at a time. Even then she achieved this success only by living in addiction-recovery-

support housing and attending daily meetings with clinicians, counselors, and groups.[55]

However, note that even if the DSM-5's approach to addiction accurately capture's

McGowan's case, the standard is problematic in that two relevantly different agents can satisfy

the criteria for a severe addiction even though the two agents have little in common beside

physical dependence on a substance. Take the difference between an agent who uses addictive

substances heavily and to the point of incurring harm, but who does not experience impaired

control over use, and an agent who does not experience substantial harm from substance use but

who experiences impaired control over use, as represented by cravings, mental obsession, and

difficulty with discontinuing use. To illustrate this difference, consider two agents who

exemplify each of these options:

Heavy, Harmful User:
This agent satisfies the criteria for *Social Impairment* and *Risky Use,* but not for *Impaired Use*.   She uses heavily daily because she likes a "party" lifestyle. Substance use enhances risky activities with her friends, such as off-road driving or cliff diving. She happily drinks while doing them, though the activities significantly risk people getting hurt as a direct result of combining such activities with substance use. She recognizes the danger of her behavior but enjoys the rush it brings. Her substance use has resulted in her getting kicked out of several living situations and her family refuses to talk with her until she stops her substance use and "grows up." She has failed out of college on account of neglecting her work to party and would often take her exams drunk. In general she found the course work unchallenging and thought it would be funny to be inebriated though all her lectures and tests. She works at a restaurant and her work colleagues use heavily as well. While she does experience hangovers in the morning, she does not obsess about using again but instead focuses on the parties and activities she will do with her friends. If an interesting enough opportunity came along, she could quit her substance use without much difficulty. She is considering moving for art school, but at the moment is just passing time and having fun.

---

[55] Seelye (2018).

Impaired Control User:
> This agent satisfies the criteria for *Impaired Control* but not for *Social Impairment* or *Risky Use*. She has tried on numerous occasions to stop substance use, but has relapsed each time. She vows to use only a "little bit" each evening to ease her cravings and stave off withdrawal symptoms, but she almost always uses more than she intends. Her mornings at work are shrouded with a hang-over from use. The rest of her day is marked by cravings for substance use and her mental focus is almost entirely on counting down the hours until she can return home to use again. Fortunately, her job is simple and she can maintain her employment despite thinking about drug use most of the day. She lives alone and thus does not neglect the care of others. She has effectively hidden her substance use from her family though she sees them often.

If both of these agents experience the pharmacological symptoms of tolerance and physical withdrawal symptoms, according to the DSM-5's criteria they are both severely addicted. Yet, the two agents have an importantly different psychological relationship with their substance use. Heavy, Harmful User uses primarily for fun and could change her using behavior without much effort if given a good reason to do so. The reasons she is able to act upon are much like the typical reasons that people regularly act upon, such as finding an interesting alternative to her current situation. It is quite plausible that cases like Heavy, Harmful User do exist. While heavy and harmful use can be evidence of loss of control, it is not always so. As Pickard (2012, 2018, 2020) observes, individuals commonly use addictive substances, even to the point of harmful excess, because it satisfies some interests even if it undermines other interests. Substance use can be pleasurable or instrumentally valuable, such as facilitating sexual behavior, expanding opportunities for new experiences, or finding a self-identity (2020, 13).[56]

In contrast, Impaired Control User does not seem able to respond to typical incentives to stop use. She continues using despite her efforts otherwise and her psychological relationship with substance use is focused more on *needing* to use than on pursuing it for the pleasure it

---

[56] cf. Flanagan, 2018.

41

brings or the richness it adds to her life. Impaired Control User has specific subjective mental experiences with regard to her substance use that shapes her behavior in important ways, such as making it difficult for her to discontinue use despite her evaluations against substance use and effort exerted to resist it.

I think there is good reason to distinguish these two cases with our conception of addiction. First, there is an important difference in the mental experience of their substance use. Heavy, Harmful User uses addictive substances to augment her enjoyment of life. In contrast, Impaired Control User seeks substance use for its own sake. She feels like she *needs* to use the substance no matter how pleasant or unpleasant any particular experience of use is anticipated to be. Watson (1999a) articulates a similar thought in distinguishing the two kinds of cases. Even if both of these individuals are highly motivated to use drugs, he suggests there is an important difference between an instrumental appetite for drugs and an appetite for drugs in themselves (12). The latter appetite seems more illustrative of addiction. Second, this mental difference results in different substance use-related behavior. An agent who uses for fun can cease to use when something more interesting comes along, while an agent with impaired control will have difficulty ceasing use even if she greatly values and wants to pursue something that is mutually exclusive of substance use. Third, the trajectory of their use is different. Again, Heavy, Harmful User is likely to stop use when other, better alternatives present themselves. This means that the agent with impaired control is more likely to continue substance use throughout their life no matter the overall personal cost of use. Finally, if the two agents commit wrong on account of their substance use, the substance use will likely play a different role in the wrongdoing as well.

All of these reasons are likely why communities devoted to addiction recovery likewise tend to distinguish such cases. Someone who does not experience features such as cravings or

mental obsession and can stop use with little effort is an unlikely candidate for being labeled as addicted, let alone labeled as severely addicted. Flanagan (2013) supports this distinction in his observation about the Alcoholics Anonymous's (A.A.) *Big Book,* that presents the A.A.'s program for resisting alcohol use. As he says, there is an important difference between someone who "'drinks like a fish,'" and as a consequence faces negative repercussions, but "who is is neither mentally obsessed, nor craves alcohol when he is not using, and could stop if he tried to stop," than someone who experiences all of these psychological features and has a hard time stopping even if trying (870 ft. note 6). While we might rightly note that the first individual is unwisely using a substance, they are importantly different from the second sort of user. [57]

The relevant differences between heavy, harmful substance users and control-impaired users provide a strong initial reason for distinguishing the two cases with our conception of addiction. Note that many substance users can experience impaired control over use and *also* use heavily and harmfully. My goal here is not to advocate for mutual exclusive categories. Rather, my point is that there are good reasons to think that substance users who only use substances heavily and harmfully but still possess sufficient control over use should *not* be included in the extension of the concept of addiction. In the next section, I will use a recent research example to illustrate why excluding merely heavy, harmful substance users is important for theorizing about agency, responsibility, and addiction.

---

[57] A possible defense of the DSM-5's conception relies on the fact that conception does a good job at achieving its *clinical* goals. By making the experience of problems, distress, or impaired functioning core considerations, it aptly identifies those who would benefit from medical interventions whether or not they experience the mental features associated with loss of control (Sinnott-Armstrong and Summers, 2018). However, I would contend that even in clinical settings, such extension diversity might be problematic. Someone who is experiencing loss of control in their regard to the substance use is likely better treated by different methods than someone who merely uses heavily and riskily.

3.2 The Perils of Conflating Cases

In addition to concerns with differences in experience, the distinction between heavy, harmful users and users who experience impaired control is important when a notion of an addiction is assumed for research or theorizing about things like behavior, agency, responsibility. A useful conception of addiction should allow for variations between cases given individual differences and degrees of severity. But because the conception of addiction impacts which agential or behavioral puzzles or features are presented by the condition, a conception that allows for substantial variation in key features can impact findings. For instance, if we want to know whether we can reasonably expect addicted agents to stop substance use when it risks addiction-explained wrongdoing, defining addiction in a way that includes both merely heavy, harmful users and impaired-control users will unsurprisingly lead to mixed answers to the question. To see what I mean, take the work of psychologist Gene Heyman.

In his work *Addiction: A Disorder of Choice,* Heyman (2009) argues that addiction is an instance of typical, voluntary behavior. He uses the condition as the basis for illustrating and defending a behavioral theory for understanding why people engage in harmful, self-defeating behavior, which he thinks is most clearly illustrated by addictive behavior. He employs the DSM-IV's criteria for addiction, which, much like the DSM-5, identifies the condition as a cluster of symptoms "indicating that the individual continues use of the substance despite significant substance-related problems" (DSM-IV, 176; Heyman, 2009, 28). Importantly, as with the DSM-5, the conception of addiction presented in the DSM-IV makes it possible to be classified as having a severe version of addiction without also experiencing features associated with loss of control over use, such as cravings, mental obsession with use, or unsuccessful

attempts at discontinuing use. In fact, there is not even a cravings criterion in the DSM-IV for substance abuse disorders.

Based on this conception of addiction, Heyman offers numerous cases that satisfy the criteria for severe addiction to argue that it is generally reasonable to expect addicted agents to resist addictive substance use. Yet, notably, the cases illustrate agents who use substances in a harmful or risky way, but are able to stop use with relative ease when they realize it is impinging upon their other life values. Though the subjects of the cases satisfy the criteria set forth by the DSM-IV for severe addiction, they are primarily merely heavy, self-defeating users who do not experience the primary features associated with impaired control. What becomes confusing is that Heyman uses these cases to support the idea that even severely addicted agents can stop use when it becomes too personally costly to long-term values and goals. However, if the focus is on cases of merely heavy and harmful substance use, it should not be surprising that such agents respond to the sort of rewarding incentives to which agents in general typically respond.[58] Yet, it still remains to be asked whether agents who experience features associated with impaired control can likewise respond to such incentives. McGowan, for instance, does not seem able to respond to incentives that even she recognizes as the overall most worthwhile things in her life.

For all these reasons, I think that for theorizing about moral responsibility we should prefer a notion of addiction that treats the mental states associated impaired control as necessary features of the condition. This approach will result in more consistency in the extension of the

---

[58] An additional concern about Heyman's research is that in addition to assuming the DSM-IV's conception of addiction in his own research, he in turn depends on research about addiction that also employs the DSM's criteria of addiction. See also Ahmed's (2010, 2012) concern that laboratory studies of addiction that primarily use non-human animals for research encounter similar concerns since all that is known is that the "participants" in the studies are heavy users.

term, while also putting a focus on the psychological and behavioral features that seem most interesting for discussing responsibility for actions.

4. Defining Addiction: Heightened Motivation

Sinnott-Armstrong and Pickard (2013) offer an influential notion of addiction that makes the mental features associated with impaired control over use central. They propose this definition: "addiction is a strong and habitual want that significantly reduces control and leads to significant harm" (862).[59] Here, I will consider each feature of their definition and offer some additional support and explanation for their contention that addiction features reduced control as a result of heighten motivation for drug use. However, I will offer two modifications to their definition. First, I will complicate their notion of control, arguing that the notion should explicitly capture the ways in which the ability for forming or maintaining evaluations over time might be undermined. Second, I will suggest that significant harm should not be included as a necessary condition of addiction, though the significant risk of harm usually marks the condition.

4.1 A Strong and Habitual Want

Sinnott-Armstrong and Pickard contend that the core of addiction is a "strong and habitual want." The notion of *wanting* is meant to capture a range of conscious and unconscious psychological features that motivate actions.[60] Following addiction researchers like Terry Robinson and Kent Berridge (1993) they employ the term "want" rather than" desire" to

---

[59] See Flanagan (2011) for a similar approach to addiction, though he focuses on the "mental obsession" aspect of addictive motivation and translates the "reduced control" criterion as "physical compulsion" (see especially pg. 282).
[60] Their focus on motivation is in line with the features of addiction identified by other philosophers. For example, see Wallace's (1999) claim that addictive desires are resilient, persistent, and have a distinctive subjective quality referred to as "cravings."

emphasize that addiction can involve desires in the absence of expectation of pleasure (855).[61] The term *habitual* emphasizes that heightened motivation for drug use can be more of a response to conditioned stimulus than a product of deliberate choice. Also, addictive behaviors are habitual in the sense that they are persistent and hard to change even when they are no longer pleasurable or condoned by the agent. The motivation involved in addiction is *strong* in that it is intense, consuming of attention, persistent, and resilient. Addicted individuals can experience "obsessional drug-related thinking" and difficulty with recalling and attending to evaluations, beliefs, and desires concerning non-drug-related activities (858). The hyper-focus of attention on the imminent reward of substance use can become all-consuming such that it renders thoughts, attention, and emotions solely focused on using.[62] Finally, unlike other appetitive motivation that will wane as the pleasurability of the activity wanes, addictive motivation is notable both for its persistence and resilience even if it is judged unwanted or unliked (859).

While Sinnott-Armstrong and Pickard touch on the details of addictive motivation briefly, it is worth elaborating on the research that supports their position to get a richer picture of the motivational nature of addiction since these features are especially relevant for understanding how addiction can impair control over substance use. Their approach to addictive motivation is supported by the neurologically-based *incentive salience theory* of addictive motivation developed by Robinson and Berridge (1993, 2000, 2003).[63] The core idea of the

---

[61] See Kennett and McConnell (2013) for some first-hand stories of this experience; see also, Berridge, Robinson, and Aldredge (2009); Holton and Berridge (2013); Robinson and Berridge (2017); Robinson, Robinson, and Berridge (2018); Volkow et. al. (2016).

[62] For a defense of the idea that addiction features an attentional bias, see Field, Mogg, and Bradley (2006); Field and Cox (2008); and Field, Munafò, and Franken (2009).

[63] See also Berridge and Robinson (1998); Lewis (2011); and Dill and Holton (2014). As Berridge and Holton (2013) note, a lot happens between the neurobiological incentive salience process and the actual behavior of drug consumption (260-261). However, the incentive salience model is helpful for appreciating both the cue-conditioned

theory is that appetitively pleasurable experiences initially trigger an increase in the level of

dopamine in the mesolimbic dopaminergic system, which *sensitizes* the brain to the connection

between certain "cues" that predict the pleasurable outcome and the outcome itself. More

specifically, when an appetitively pleasurable experience is interpreted at the neural level as

"rewarding," i.e. something that should be repeated, these subconscious brain systems "note"

connections between features correlated with the experience - i.e. cues - and the appetitive

experience itself.[64] Cues can vary depending on one's experience. Once sensitized, conscious

and unconscious attention is tuned to notice these "signs" that anticipate that outcome. When

cues are encountered, motivation to seek the outcome is triggered.

Addictive substances are especially potent rewards for cue-conditioning because they

offer, at least in initial uses, intense pleasure immediately following consumption.[65] The brain

subconsciously tracks cues related to this hedonic experience as predictors of repeating this

powerful experience. Cues can be anything that an addicted agent has come to unconsciously

associate with drug use, including psychologically internal or external stimuli such as certain

thoughts, feelings, sounds, smells, people, places, or even types of situations. They might have

no explicit connection to substances or substance use, like thoughts, memories, emotions, smells,

---

nature of heightened motivation for drug use and the persistence of cue-sensitive motivation even when an addicted person deems it best to stop using or even tries to stop using. While I think this particular incentive salience theory of addictive motivation is useful for understanding the nature of the heightened addictive motivation, I will not assume that this exact biological story is necessary for addictive motivation or addiction.

[64] While the incentive salience theory is highly influential, many neuroscientists discuss how the motivational features of addiction recruit other systems of the brain as well. Appetitively rewarding experiences increase in the production of the neurotransmitter *dopamine,* which then communicates with other central structures in the brain involved in motivation (nucleus accumbens) decision making (orbitofrontal cortex) and emotions (amygdala) (Lewis, 2011).

[65] See Berridge and Holton (2014) and Dill and Holton (2014) for a description of how particular addictive drugs act directly or indirectly on the reward system. Also, this is why the route of administration can impact the addictiveness of a drug. The quicker the drug reaches the brain after administration, the more addictive it is.

people, or even patterns of behavior experienced before, during, or after drug use.[66] Some examples of cues might be the notes of a song heard or the smell of something cooking that were at one point experienced while using a substance. Merely driving part of the route one takes to the drug dealer's house might also serve as a cue. Later, when an addicted individual encounters a cue, she experiences what is described as a "psychophysiological response" characterized by a drive to consume the substance (Lewis 2011, 152). Sensitivity to cues can even persist even long after ceasing substance use.[67]

As a result of *incentive sensitization* as predicted by the incentive salience theory, an agent will experience what Brendan Dill and Robert Holton (2014) refer to as immediate or "occurrent" motivation to engage in the rewarding activity when cued. Heightened motivation to engage in appetitively rewarding experiences, though, is not present at all times, even after sensitization. Instead, once sensitized, an agent develops a "dispositional" motivation to repeat the activity in the future.[68] This means that the sensitized user's awareness is heightened to both detect and respond to any cues, even if they are not actively experiencing heightened motivation.

Incentive sensitization is developed without the agent's conscious awareness and heightened motivation is typically cued without conscious awareness as well. However, there are important subjectively felt motivational aspects of incentive sensitization, namely the

---

[66] Lewis (2011) provides a detailed summary of how this neural process works (see especially p. 151).

[67] For support of this position, see also Dill and Holton (2014, 29); Lewis (2011, 2017); Schroeder and Arpaly (2013).

[68] I follow Dill and Holton, 2014 in using the terms "occurrent" and "dispositional" to describe the distinctive types of motivation involved in addiction. Also, note that not all addictive drugs impact the dopaminergic system in the same way. Berridge and Holton (2013) explain that the impact on the dopaminergic system be direct or indirect depending on the drug: "…addictive drugs have an impact on the mesolimbic dopamine system: either by stimulating the release of dopamine (in the case of amphetamine, nicotine, and caffeine); by reducing the release of substances like GABA that themselves reduce the amount of dopamine released (opiates and perhaps THC); by reducing the level of substances that break down dopamine (alcohol); or by reducing the activity of the system that reabsorbs dopamine (cocaine and perhaps amphetamine)" (245).

experience of what is usually referred to as *cravings*. Cravings are generally understood to involve negative emotions, heightened attention to the feeling of deprivation or denial, obsessive thinking, and a consciously experienced drive to engage in a particular behavior.[69] Others emphasize that in addition to negative emotions, anxiety, and feelings that one needs to act, cravings tend to be "irrepressible and intense" as well as psychologically intrusive (Auriacombe, et al., 2018).

The psychological experience of cravings is taken by researchers across many fields to be a central explanatory notion of addiction – even more central than features like withdrawal.[70] Moreover, many clinicians treat cravings as a necessary condition of addiction in clinical settings as well, despite its absence in the DSM-IV and circumscribed role in the DSM-5. In clinical practice, addiction is often distinguished from mere physical dependence to a substance on account of the presence of cravings.[71] Similarly, interventions to treat addictions often aim at lessening or stopping cravings as the first line of action. Prescribed drugs attempt to stop or dampen cravings and cognitive behavioral therapy helps to lessen the power of cravings or to modify behavior to avoid triggers that incite them.[72]

---

[69] Robinson and Berridge (2000); Lewis (2011); Arpaly and Schroeder (2013, 230).

[70] Skinner and Aubin (2010); Berridge & Robinson (2011); Flanagan (2020); Auriacombe, et al. (2018). Interestingly, others have suggested that because some supposed addictions, like caffeine addiction, lack intense or urgent cravings, it follows that cravings are not a necessary condition of addiction (Wallace, 1999, 625). In response, I would contend that what these people are describing would better be described as a mild addiction or merely physical dependence rather than addiction. The DSM-5 commits to the latter option with regard to caffeine use. See Franken (2003) for an attempt to integrate levels of explanation with regards to cravings.

[71] Cederquist, Lynette MD. (2019, Nov. 26) "Managing Pain in the Opioid Crisis: Balancing patient needs against the widespread threat of addiction" at UC San Diego Medical Center Hillcrest.

[72] Robinson and Berridge (2000). Interestingly, prescribed pharmaceuticals that seem most effective at combating addiction are those that directly aim to modulate the experience of cravings as opposed to another symptom like tolerance or withdrawal. Some pharmaceutical examples include methadone and buprenorphine used to combat cravings for opiates, naltrexone and acamprosate used to combat cravings for alcohol, and nicotine patches or chewing gum use to combat cravings for nicotine.

Other subjectively felt aspects of addiction indirectly feed addictive motivation as well by heightening sensitivity to cues and increasing motivation for use. Though withdrawal symptoms are no longer considered by most theorists to be the primary motivational driver of addictive drug use, physical symptoms upon cessation can range from physically unpleasant to life threatening, and drive addicted agents to seek relief in use. The satiation offered by addictive substances is often so brief and intense that it leaves "frustration, loss, and often depression in [its] wake" (Lewis, 2017, 9). An attempt for relief can build on cravings. Finally, many theorists argue that the experience of shame that often accompanies addiction and addictive drug use also motivates continued use by exacerbating "the need for resolution, regulation, or escape" (Lewis, 2017, 9).[73]

The incentive salience theory, as well as the other subjectively felt aspects of addiction, such as cravings, withdrawal symptoms, depression, and shame, help explain many of the notable characteristics of addictive motivation, such as its cyclical nature, dissociation from pleasure, intensity, and persistence.[74]

4.2 Modeling Addictive Motivation and Behavior

If the incentive salience theory is right, the motivational aspects of addiction help explain addictive behavior. Sinnott-Armstrong and Pickard (2013) explain the motivational upshot as including the following four features. First, drug use becomes increasingly habitual, meaning that it is "more wanted than liked, more automatic than deliberately chosen" (858). Second,

---

[73] There is copious recent work that studies how the attitude of shame, and similar attitudes, can perpetuate addictive drug use. See Flanagan (2013); Snoek, McGeer, Brandenburg, & Kennett (2021); and Matthews, Dwyer, & Snoek (2017). See also Burgess (2016).

[74] The incentive salience theory depends largely on research of the dopaminergic system. But other brain systems are likely involved in addiction as well. See Goldstein & Volkow (2002) and Volkow & Fowler (2000).

resisting addictive drug use requires extra effort in exercising "willpower," and this willpower

gets depleted with continued exertion.[75] Third, addiction impacts attention and cognition,

directing focus toward drug use and prioritizing drug use in evaluations and reasoning.

Addiction can lead to a neglect of other values, drastic though temporary shifts in preferences,

and an overestimation of the benefits of drug use paired with an underestimation of the benefits

of not using. Finally, motivational ambivalence can plague addicted individuals even after they

resolve to quit, with motivation for substance use persisting long after establishing the want to

quit.

While Sinnott-Armstrong and Pickard (2013) say little to elaborate upon these four

features, I think the features can be helpfully distinguished as aspects of two general models of

how addiction impacts evaluations and behavior. The first model suggests that addictive wanting

motivates substance use despite opposing all-things-considered evaluations. The second model

suggests that addictive motivation actually influences the content of all-things-considered

evaluations regarding substance use. These two models are often presented as the basis for

competing accounts of addiction, but here I will suggest that both are needed to accurately

capture the various ways in which addiction is experienced and contributes to loss of control.

Both models assume a two-part picture of human psychology that distinguishes

psychological attitudes related to evaluation from basic motivating desires or wants that do not

---

[75] Though championed by researchers like Baumeister et al., 2020, the idea that willpower is an easily depleted distinctive mental faculty is controversial. Amaya (2020) helpfully summarizes this debate and suggests a way to preserve the idea that willpower is limited without also taking on the more controversial aspects of will power theories. I will not engage with this debate here, but instead draw from Sripada (2019) to suggest another way in which this experience of depleted ability for resistance might be captured. See also Holton (2009) for a defense of willpower as a distinctive mental state.

have evaluative content.[76] The evaluative psychological category is comprised of evaluative attitudes involved in assessing the worthwhileness of action-related goals. The "evaluation system," to borrow a term from Watson (1975), involves abilities related to forming and holding values, preferences, and beliefs aimed at deciding and implementing what one deems she should do all-things-considered. The evaluation system can produce evaluation-dependent motivation, variously referred to as "cognitive" desires (Dill and Holton, 2014, p31) or "practical" desires (Sripada, 2017, 789). This evaluative category is contrasted with a second psychological category consisting of attitudes that produce motivation independently from the evaluative system. Variously referred to as "spontaneous desires" (Sripada, 2017), "blind" motivation (Watson, 1975), or "given desires" (Wallace, 1999), what makes this category of attitudes distinctive is that these psychological sources produce desires that can arise independently of what is deemed valuable or best all-things-considered.[77]

On this two-part picture of human psychology, evaluations and evaluative-independent motivation can work together. Evaluations can influence independent motivation by shaping appetites or influencing the formation of habits. Conversely, evaluative-independent motivation can influence evaluations. However, evaluations and evaluative-independent motivation can also work against each other. Desires stemming from appetites or habits can be highly motivating despite directly conflicting with evaluations about what is best all-things-considered.

---

[76] Some psychologists likewise defend a two-tier model of motivation. See Sripada (1999) for a description and brief defense of this "dual process" structure of motivation (4).

[77] A variety of motivational sources are associated with this category, such as appetites associated with bodily needs and functioning like hydration, nutrition, and sexual reproduction, as well as emotions and habits (Sripada, 2017, 789; Dill and Holton, 2014, 31; Watson, 1975).

With this two-part picture in mind, the first model of how addictive motivation influences behavior turns on the idea that addictive motivation is especially strong independently of evaluations and can lead to action despite an agent's all-things-considered practical evaluations against drug use. Though not exactly his own view, R. Jay Wallace (1999) nicely articulates the core of this model:

> Addicts, we tend to suppose, are subject to impulses or cravings that are peculiarly unresponsive to their evaluative reflection about what there is reason for them to do. As a result of this unresponsiveness, we further suppose, addicts are typically impaired in their ability to act in accordance with their own deliberative conclusions (1999, 621).

On this model, the strength and persistence of addictive wanting that leads to an evaluative/motivational conflict is what makes addiction distinctive from other appetitive desires. Other desires might result in such a conflict, but not to same degree as addictive wanting.[78]

However, note that this model of addiction as an evaluative/motivational conflict does not presume that addictive desires irresistibly compel action, as assumed in the classical cases of willingly and unwilling addicted agents.[79] Rather, theorists like Sripada (2019) and Noggle (2016) argue that addictive motivation is better understood as persistent but never irresistibly strong.[80] Addictive motivation ebbs and flows in intensity, but persistently reoccurs in a way that makes regulating this motivation in light of one's evaluations especially difficult. As

---

[78] Some theorists who emphasize this conflict as the distinctive attribute of addictive motivation include Dill and Holton (2014); Holton (2009); Berridge and Robinson (2011); and Arpaly and Schroeder (2013, 2014). For another defense of the idea that addiction involves a reduction of volitional control see Loewenstein (1999).
[79] Note that some theorists who support this model do use the term "compulsion," but they use it *not* to say that the desires are irresistible, but just that they can strongly conflict with evaluative attitudes (e.g. Robinson and Berridge, 2011).
[80] See Noggle (2016) for a defense of a similar claim about the persistence of addictive desires.

Sripada explains, addictive desires frequently reoccur, even if temporarily sated or successfully resisted on a single occurrence. This frequency makes addictive desires difficult to resist in two ways. First, he suggests that though an addicted individual might be able to regulate any one occurrence of an addictive desire, given their frequent reoccurrence there is a greater probability over time that the addictive individual will eventually succumb. This idea of *sequential fallibility* rests on the thought that an agent is more fallible in respect to the performance of a sequence of events as opposed to a single event. Second, Sripada suggests that the frequency of persistent addictive desires results in a *cumulative coercion* in that their frequent reoccurrence means that "top-down regulation" has to be exercised frequently and for long stretches of time, which is mentally costly in two respects. First, mentally regulating errant desires is costly hedonically in that" prolonged dysphoric subjective experience" typically accompanies mental regulatory efforts. Second, mental regulation is costly in that it consumes limited mental space for executive functions, like planning and deliberating, that is also required for other tasks or projects (2019, 11). Even if one deems resisting substance use to be best all-things-considered, the capacity for mental regulation can be over-extended with other tasks, making it especially difficult to maintaining resistance over time.[81]

The second model of addictive behavior proposes that addictive motivation leads to substance use by influencing the very evaluations that concern the all-things-considered worthwhileness of substance use. As Watson (1999a) articulates this view, addicted individuals are not "so much overpowered by brute force as seduced" by addictive desires (10). One way in

---

[81] To add to this idea, Watson (1999a) explains how regulating addictive desires has two levels: exerting mental focus to resist itself requires mental effort, but sustaining this focus over time requires effort as well. Watson refers to sustaining effort as "the predicament of self-control" (11).

which addictive motivation is proposed to influence evaluations is by temporarily causing a shift in all-things-considered practical evaluations.[82] In "cool" periods, when satiated or not experiencing cravings, an addicted individual might deem it best all-things-considered to avoid drug use. Then, when cued and shifted into a "hot" moment of cravings and mental obsession, the agent temporarily changes her all-things-considered evaluation of drug use to prefer substance use over other options – at least for a short time.[83] In this moment, the agent extremely discounts the future costs of use. When the agent returns to a cool period, however, they regret consumption. For some addicted agents, diachronic shifts in evaluations allow for days of reprieve between hot moments. Other addicted agents experience only moments of reprieve that are short and few and far between, allowing perhaps only hours between substance consumption and the heightened cravings that shift evaluations (Lewis, 2017, 9).

One impact of evaluative shifts is "deliberative narrowing," in which reasons in favor of substance use become especially salient in a way that biases deliberation in favor of substance use (Dill and Holton, 2014). When actively weighing one's practical options, the benefits of use are highlighted and exaggerated while the negative features are inaccurately construed or minimized. Flanagan (2013) suggests that, because considerations against substance use "matter less than they should," this biased reasoning can result in an "inability to draw normal inferences about the harms one is doing to oneself or others" (148). Deliberative narrowing can also cause an agent to bypass even considering the worthwhileness of drugs use by encouraging a hyper-focus on the logistics surrounding procuring and using drugs.

---

[82] For research evidence to support this position, see Lowenstein (1999); Heyman (2009); Berridge and Holton (2013); and Ainslie (2001).
[83] Ainslie (2000, 2001, 2017) refers to this temporary oscillation in overall preferences as "hyperbolic discounting." See also Levy (2012) for a defense of the oscillation model in terms of oscillating beliefs.

Beside inducing *temporary* shifts in evaluations, some who support this second model

argue that addictive wanting can result in a more *extended* influence on evaluations by impacting

the very ability for forming or maintaining evaluations. Monique Wonderly (2021) argues that

addictive motivation can undermine the capacity for valuing by fomenting a myopic focus on the

object of addiction, leaving little or no room for other things of value. Addiction "crowds out"

engagement with other valuable things and activities worth valuing (239).[84] Of course,

Wonderly acknowledges that an evaluative crowding out naturally follows from having many

values and attachments that require focus and attention. She offers the example of parenting to

illustrate this point. Loving one's child might leave little time and attention to attend to other

valued projects, like gardening, volunteering, or even working (239). Yet, what makes addiction

distinctive, according to Wonderly, is the extent to which it can involve a myopic focus on the

objects of addiction over and against other things of value. She writes:

> "in the case of addiction, the agent's focused attention on the object sometimes
> not only causes her to compromise other values, but in a sense, it *blinds* her to
> them. Her mode of engagement is bleary-eyed and narrow, and she is in no
> position to recognize properly what she forgoes on account of her addiction."
> [239]

Wonderly observes that addiction's undermining of the capacity for valuing is evident during the

experience of cravings, but can also be present more generally in the addicted agent's life. In

typical cases of addiction, an addicted agent might be momentarily blinded not to see the other

things they hold or would hold as valuable. However, in severe cases of addiction, an addicted

agent can be consistently less able to recognize or properly weight things of value that might

---

[84] cf. Elster (1999); Watson (1999a). See also Summers (2015) for a related yet distinctive argument that addiction
is distinguished from other so-called "passions" by a type of "misevaluation on which a pattern of behavior is
based" (37).

compete with substance use.[85] Even when not actively craving, the ability to attend to and

properly weight the value of other things of value can be minimized or undermined entirely.

Again, these two models of how addictive motivation influences behavior are often

proposed as independent models.[86] I would suggest, however, that they are not necessarily

conceptually nor psychologically mutually exclusive as models of addictive behavior.[87] It is

entirely plausible that at different points in time an addicted agent's behavior might be best

explained by one model or the other.

4.3 Addiction and the Loss of Control

Sinnott-Armstrong and Pickard offer the following standard for explaining how addictive

motivation reduces control over substance use given the influence of addictive motivation:

[Control] "…an agent has control over a type of action iff:

1. If they want overall to perform that type of action, then usually they do it; and

2. If they want overall not to perform that type of action, then usually they don't do it."

[856]

This account of control has many benefits. It sets an explicit general standard or threshold that

must be met to be said to "have control" over a type of action. The term "usually" captures the

---

[85] c.f. Yafee (2011, 128).

[86] Each of these approaches is related to an understanding of what self-control is and how it is exerted. Amaya (2020), for instance, helpful presents these two approaches in light of theories of self-control and nicely surveys the diverse theories captured under each approach. Henden (2018) makes a similar distinction between two models of self-control (46).

[87] Some instances of each model are proposed to explain human motivation and behavior in general to the exclusion of the other model. For instance, some advocates of the evaluation oscillation approach argue that at the time of consumption, an addicted individual *always* evaluates substance use at the best option such that there is never a conflict between evaluations and independent motivation in that moment (e.g. Heyman, 2009). However, as Kennett and McConnell (2015) convincingly argue, it is plausible to think that at least some addicted individuals can succumb to an addictive desire that conflicts with their all-things-considered evaluations even *at the very time of consumption.*

idea that perfect or near perfect motivational regulation is too high of a threshold for control. Also, it is commonly thought that control comes in degrees, and this account captures this intuitive idea.[88] Finally, "want overall" is an ecumenical approach to the notion of an agent's evaluative stance. "Overall want" might be understood in terms of desires, such as a second-order desire, or in terms of all-things-considered evaluative judgements.[89]

Yet, this notion of control has flaws as well, especially when viewed in light of the two models of addictive behavior explained above. First, this conception does not distinguish instances of weakness of will from instances of loss of control. For instance, I may overall think it is a good idea to get 20 minutes of exercise a day, but not usually do this. Yet, it is a further question of whether I have reduced control over this type of action. I may easily be able to exercise every evening but repeatedly decide to watch my favorite program instead.

Second, if addictive motivation can influence substance use behavior through influencing evaluations, at least in some instances, this notion of control does not fully capture how addictive motivation can undermine control over substance use, either through periodic oscillations in evaluations or through a more global reduced ability in the capacity for evaluative formation or maintenance. An addictive agent might be doing exactly what they want to be doing when they use a substance, but this is precisely part of the problem of addiction. Their very control over maintaining evaluative stability or recognizing the value of competing things seems importantly compromised by their addiction with the result that at the time of consumption they really do prefer consumption overall.

---

[88] c.f. Mele (1996); Coates and Swenson (2013); Nelkin (2016).
[89] c.f. Holton and Schute (2007).

A satisfactory notion of control should be able to capture the key ways in which an addicted agent might possess or lose control. In the following chapters, I will be defending the idea that, at least in the context of responsible agency, control is best understood in terms of reasons responsiveness – a notion of control that can accommodate both models of addictive behavior. However, at this point, in terms of understanding control in terms of agency more generally, I am open to notions of control that can capture my above concerns.

But no matter which conception of control is ultimately preferred, more needs to be said to specify how addiction might reduce or undermine control, which speaks to my first concern with Sinnott-Armstrong and Pickard's notion of control. I propose the following account: *addiction can reduce control over substance use by making it more difficult for an agent to resist substance use.* Following Nelkin (2016), the notion of *difficulty* might be understood in the sense of" requiring a great deal of effort" (357).[90] Addictive wanting can make it difficult to resist substance use by making it especially effortful to act on one's evaluations against drug use, maintain consistency in one's evaluations, or, in some cases, even recognize and weight things of value that might compete with substance use. The severity of an addiction, as determined by the severity of the features associated with the motivation to use substances, such as cravings and mental obsession, impact the degree to which resistance is difficult. It is plausible that an addicted individual might experience difficulty with resisting substance use in just one of these three ways and not the others. Or, they might experience difficulties with resistance in more than one way at difference stages in their addiction.

---

[90] Nelkin also suggests that the notion of difficulty can be understood in the sense of "requiring a great sacrifice of one's interests." Here I focus on alone effort, though I address the second sense elsewhere (Koepke, in preparation).

Throughout the following chapters, I will discuss the ways in which addictive motivation can make resisting drug use especially effortful. Even if addiction does not feature literally irresistible desires, successfully resisting substance use can be especially difficulty, especially for agents suffering from a severe addiction. Take the case of McGowan again. Most of her mental and physical energy throughout each day is spent in an attempt not to use addictive substances and to maintain her evaluations against drug use. Her continued success is a product of her effort and external resources that offer her constant support.[91] In cases of severe addiction, her experience in successfully resisting substance use over an extended period of time is not that unusual.

4.4 Harm

I have argued that, at least for discussions about agency and responsibility, we should assume a notion of addiction that considers the features associated with impaired control necessary for being identified with the condition. Still, there remains the question of whether significant harm should be a necessary feature of the condition as well even if harmful, self-defeating behavior is not the agential puzzle we are seeking to explain. Sinnott-Armstrong and Pickard argue that a harm condition is necessary to the notion of addiction because harm usually follows from the motivational features of addiction associated with loss of control. But also, its inclusion is necessary to distinguish addiction from other appetitive conditions such as love. I will address these arguments in turn to contend that though risk of significant harm is considerable in severe cases of addiction, actual harm should not be considered necessary to the condition.

---

[91] Seelye, Turkewitz, Healy, and Blinder (2018).

Understanding harm primarily in terms of "death, pain, distress, and dysfunction," Sinnott-Armstrong and Pickard argue that addiction leads to significant harm. While not every case of addiction obviously displays such harm, they argue that addiction introduces an "extreme fragility" in which the "risk of harm can itself arguably count as a harm" (861). As their argument goes, even though some addicted agents manage to find more benefit from addictive substance use than harm, harm is avoided "only in a very narrow environmental niche" (861). This situational fragility means that even the smallest change in circumstances can result in significant harm.

In response, I think they are right that situational fragility applies to many cases of severe addiction given both the motivational nature of addiction and the impact of chronic substance use itself, which tends to undermine long-term interests and projects of value that might compete with substance use. For instance, in addition to the considerable physical harm that addiction can cause to personal health and the pain and distress that can accompany cravings and recovering from use, Heyman (2009) offers the idea of "behavioral toxicity" to explain how addiction sabotages competing valuable activities, which contributes to the experience of dysfunction. The cycles of drug seeking, intoxication, and withdrawal tend to "poison the field" of fulfilling, long-term life projects and relationships, destroying these projects or preventing them from developing in the first place (145).

Still, it is plausible that there are some addicted substance users who have values and projects that align with their substance use such that they benefit overall from their addiction. Such agents manage to live long-term in the "narrow environmental niche" in which the benefit of their addiction outweighs the harm from addiction or substance use that they experience. One such individual might be rock star Keith Richards who has regular access to an endless supply of

clean substances and who deems the creative and exciting experience of substance use to be worth any harms that might accompany it. It would be odd to say that Richards experience significant harm from his addiction, but it would also be odd to exclude him as a "true" addicted agent if he also experiences features like impaired control over use.

However, even if we admit that addiction introduces situational fragility in which significant harm is only avoided narrowly, I am doubtful that this fragility is unique to addiction such that it will help distinguish it from conditions like love. Sinnott-Armstrong and Pickard's claim that addiction shares a motivational profile akin in many ways to conditions like being in love is supportable by a compelling argument. As with being addicted to a substance, when in love "one can become single-minded, obsessed, and devoted at the expense of many other goods" (860). To support a similar claim, Wonderly (2021) argues that, along with underlying neural similarities, both addiction and loving attachment share core motivational features, both at the personal and sub-personal level, including "intensely focused attention, cravings, obsessive thoughts, emotional dependence … recalcitrant desires for further contact" and anxiety and depression when denied contact (225).[92] Finally, addiction and love – even healthy, non-abusive love – introduce a "vulnerability to suffering and diminished self-control" (238). Absence from the loved one can be distressing, for instance.

However, even if addiction and love introduce a situational fragility of risk of significant harm, we still might distinguish addiction from love by both the type of harm that is risked and the probability of the risked significant harm. For instance, Wonderly argues that addiction lacks the natural safeguards that might protect against the extreme evaluation myopia that can be

---

[92] See also Burkett and Young (2012,16) and Fisher et al. (2010; 51-2).

experienced in severe addiction. Valued activities in general can require focus, time, and attention – sometimes to the exclusion of other things or pursuits of value. However, the myopic focus that attends severe addiction often resists the "checks" that attend other projects of value. For instance, Wonderly argues that healthy loving attachments usually include protections against evaluation myopia. Even if a lover experiences a diminished ability in noticing or attending to their other things or projects of value, the beloved can respond to a lover's needs and "prevent the suffering that would result from rejection and separation and reduce the negative impact of diminished self-control by increasing one's self-confidence and self-reliance" (240). Moreover, the lover and beloved can engage in valuable projects together, which would direct the focus onto other things of value. Finally, the beloved might support the lover in pursuing other things of value beyond the beloved. In this way, the condition of love might actually open the lover to recognizing and appreciating their other values or forming new values.

Addictive substances or behaviors, however, do not generally offer these protective features. If Wonderly is right, what might distinguish addiction from other appetitive conditions or behaviors is a complicated combination of the probability of risked harm combined with the type of harm and the degree of the severity of the harm being risked due to the lack of protective features. Still, it is not clear that everyone who is severely addicted, as understood in terms of experiencing the features of loss of control, also experiences such severe evaluation myopia. As such, we might want to identify a different feature of addiction that might distinguish it from other similar motivating conditions. One possibility is just to say that the motivational features of addiction tend to be more extreme than other similar conditions. The difference between

motivating conditions, then, is one of degree and not of kind.[93] I will consider one other

possibility in chapter 5 in my discussion of the hijack theory of addiction.[94]

In sum, the feature of significant harm might not provide the bright line that helps

distinguish the condition of addiction from similarly motivating conditions. Also, it is not clear

that the mere risk of significant harm is a harm in itself since many worthwhile pursuits and

projects in life can entail such a risk. Still, given the motivating features of addiction, the risk of

significant harm seems to be a common experience, especially in severe instances of the

condition. As such, the harm feature might be best described as a "significant risk of

considerable harm."

5. Revisiting the Cases of the Willing and Unwilling Addicts

I have argued here that for theorizing about responsible agency we should assume a

notion of addiction that views the condition as a strong and habitual want that significantly

reduces control and significantly risks considerable harm. The one important note to this

definition is that addiction likely comes in degrees and, plausibly, the less severe the experience

of addictive motivation the less severe the attending loss of control over substance use or the risk

of considerable harm. To capture the wide range of severity, the definition itself might be

modified. However, in this dissertation I am specifically concerned with severe cases of

addiction since they most clearly illustrate the important features relevant to considerations of

---

[93] Lewis (2017) and Butlin and Papineau (2016) defend this position.
[94] Berridge and Robinson (2011), Dill and Holton (2014), Arpaly and Schroeder (2013, 2014) all suggest that addiction is distinctive from other "natural" conditions because addictive substance hijack the brain's natural reward processes.

responsible agency. As such, I am mainly concerned with cases of addiction that satisfy this definition to a high degree.

I have also argued here that while impaired control is a necessary feature of addiction, the condition should *not* be characterized as featuring literally irresistible desires. Instead, the heightened and strong motivation typically only partially reduces control and does not compel addicted individuals like an "outside force" to substance use. Addictive desires can be hard to resist because of their sustained tenacity, but also because they influence evaluations against substance use, often just temporarily but sometimes more profoundly and diachronically. Given the concerns with how the willingly and unwillingly addicted agents are originally stipulated, the cases can be modified to illustrate loss of partial control, though more than two cases are needed to capture the various possible paradigms given the two models of addictive behavior and the possible extremes of willingness. Specifically, the cases can be described to either reflect the idea that addiction involves impaired control over substance use as a consequence of strong and persistent motivation that leaves evaluations untouched, or the idea that it involves impaired control as a consequence of addictive motivation that influences evaluations of substance use as well.

In light of the first model, it is plausible that some severely addicted agents are able to clear-headedly evaluate considerations for and against substance use even though they experience strong and persistent addictive motivation. Holding fixed this experience of impaired control, the cases can still be distinguished in terms of what the agents want most overall in regard to substance use.

> Unwillingly Addicted Agent: This agent is strongly and persistently motivated to use his preferred substance and as a result finds resisting use significantly

difficult. Yet, he recognizes that he has good reason to stop substance use given its toll on his health, family, and valuable projects. He exerts daily effort to try to resist use, both directly though mental resolve and through making use of external resources, such as counseling, groups, and medication.

Willingly Addicted Agent: This agent is strongly and persistently motivated to use his preferred substance and would find resisting use significantly difficult if he tried. He recognizes that there are considerations to stop using given the toll of his addiction and substance use on his health and family. However, he has assessed his values and holds that substance use is more interesting and exciting than maintaining his long-term health or fulfilling his family obligations.

Though these two cases are hypothetical, they illustrate how addiction can be experienced in some instances. Take the case of Mark who illustrates the unwillingly addicted agent as described above.[95] A former soldier deployed in Iraq and wounded in an I.E.D. explosion, Mark initially used the opioid oxycodone as he recovered from his injury. But the pain lingered and he eventually switched to heroin as a more cost effective pain killer. After recognizing he had become addicted to opioids and not wanting to be addicted, he devised a plan to move to a new state away from his current influences and detox by using small doses of heroin and Suboxone, a synthetic opioid that lessens the efficacy of other opioids by dampening cravings and withdrawal symptoms. While Mark implemented his plan, though, a dealer deceptively sold him heroin laced with the considerably more powerful fentanyl. Trapped by a level of pain and wanting he had never before experienced, with Suboxone doing nothing to ease the wracking aches and fierce cravings, Mark feels at the mercy of an addiction he does not want and cannot seem to overcome. Despite desiring substance use, he hates that he is motivated to keep using. What he wants and what he wants to want are at war.

---

[95] Percy (2018).

In contrast, Richards illustrates the willingly addicted agent described above.[96] Richards deems the creative and exciting experience of substance use to be worth any harms that accompany it. In his autobiography, he acknowledges the toll that his addiction has taken on his family and health but emphasizes his appreciation for the experience of using drugs. He delights in intentionally seeking out new types of drugs for pleasurable and novel experiences. If his autobiography is any indication of his abilities, he can think carefully about why he uses drugs, what drugs he wants to use, and the importance of the quality of the drugs to his experience. He has clear-headedly prioritized these experiences above his personal health or family obligations.

According to the second model of addiction and behavior described above, it is plausible that addiction can influence evaluations of substance use. An agent who experiences temporary shifts in favor of substance use might be considered unwilling given that their evaluation in in "cool" moments are against substance use. Such an agent can be captured as this:

> Diachronically Unstable Unwillingly Addicted Agent: This agent is strongly and persistently motivated to use her preferred substance and as a result finds resisting use significantly difficult. When she is not experiencing strong cravings, she deems substance use unfavorably and exerts effort to resist use. However, when her cravings and mental obsession are particularly strong, her evaluations of substance use shift to favor use, justifying it as okay because it is "only one time." After use, she regrets her actions and feels ashamed at losing her resolve.

Such an unstable agent is illustrated in the case of Carr mentioned earlier in this chapter. In his cool moments, he could appreciate the weighty considerations against substance use. But, under the sway of heightened motivation, his evaluations shifted to reprioritize substance use above his other values and obligations.

---

[96] Flanagan (2017) describes this case as a paradigmatic instance of a willing addict.

Finally, cases of especially severe addiction can feature a more diachronically consistent bias in favor of substance use. Such a case might be captured as this:

> Diachronically Stable Willingly Addicted Agent: This agent is strongly and persistently motivated to use her preferred substance and would find resisting use significantly difficult if she tried. She has little time of reprieve between using a substance and experiencing severe cravings again, and her withdrawal symptoms are severe. This brief reprieve is mostly consumed by immediate considerations such as dealing with pain and meeting basic needs. Her daily deliberative focus is almost entirely consumed by procuring and using her preferred substance. She reasons that substance use is not just something she wants but something that she *needs* to stay alive and continue existing.

Patrick Griffin illustrates this more substantial and consistent evaluative bias in favor of substance use.[97] Griffin started using heroin in his early teens, overdosing on heroin and its synthetic kin fentanyl 30 times by the age of 34, but saved each time by the quick response of his family or strangers. Over his nearly 20 years of addiction, Patrick has occasionally stopped using but only for short periods and only as a consequence of compelled time in jail or a medical facility. Despite the help offered by his family, many of whom have struggled with their own addictions, he has refused to stop using. When pressed to explain why he uses, he describes himself as sick and says that drugs are his "cure." He emphasizes that those who suggest he stop using or try to help him just do not understand his experience.

These four core cases will be referenced throughout the remaining chapters. In the next three chapters, I will argue that capacitarian Reasons Responsive theories of moral responsibility can better track and explain the importance of relevant differences between these cases to give a more satisfactory account of which features should be central in assessments of blameworthiness of wrongdoing than Quality of Will theories.

---

[97] Seelye (2018).

In the next chapter, I will consider the Frankfurt's original versions of these cases as modified by Sripada (2017) to argue that, when the cases are clarified to address concerns with wrongdoing and control, capacitarian RR views not only get the right extension with regard to the cases but can also track and explain important differences between cases of addiction that QW views miss.

Chapter 3: Stipulating the Features of Addiction: Non-Structural Quality of Will Views and Capacitarian Reasons Responsiveness

1. Introduction

Frankfurt's (1971) hypothetical cases of the so-called "unwilling and willing addicts" are often treated as illustrative for revealing the features most fundamental to assessments of responsibility. The idea is that the cases, as Frankfurt originally describes them, supposedly distinguish the assessments of two types of views: those that see control over actions to be the core consideration in assessments of responsibility and those that see an agent's evaluative stance to be the core consideration.[98]

Again, when assessing whether or to what degree an agent is blameworthy for a wrongdoing, RR theories ask whether the agent can track their sufficient practical moral reasons with regard to that action and can act in accordance with these reasons. In contrast, QW views hold that blameworthiness for a wrongdoing turns on whether the action manifests or is sufficiently motivated by some privileged evaluative attitude held by the agent with regard to the rights or well-being of others. Non-structural QW views, in particular, consider whether a wrongdoing reflects the agent's *conative* or *cognitive* fundamental evaluative attitude, such as the agent's fundamental cares or evaluative judgments, respectively. Notably, these views reject the idea that control should be a consideration in assessments of blameworthiness.

The hypothetical case of the willingly addicted agent is thought to pose a particular challenge for control-based views of moral responsibility, revealing their limitations while illustrating the extensional adequacy of their QW competitors. As the thinking goes, control-based views wrongly excuse the willingly addicted agent of blameworthiness though this agent is

---

[98] Sripada (2017) makes this explicit observation (781).

clearly intuitively fully blameworthy. This extensional failure is taken to indicate a flaw in the idea that control should play a key role in assessments of responsibility for wrongdoing.

In this chapter I will argue that the cases, as originally stipulated, are less illustrative of the key features of responsible agency than they might initially appear.[99] As I will show, despite Sripada's (2017) recent attempts to clarify the details of the case of willingly addicted agent to make it a better tool for illustrating the features of concern in assessments of responsibility, many relevant features of the case remain indeterminate. I will contend that even when we offer further description to more clearly stipulate the nature of the wrongdoing and the loss of control, it is not always clear which features intuitive reactions are tracking in assessments of blameworthiness. Yet, when the cases are stipulated in a way that more plausibly tracks everyday notions of control, and so does not invite implicit assumptions that defy the stipulation, we can see that capacitarian RR views are more extensionally adequate than non-structural QW views.

In section 1 of this chapter, I present Sripada's basic argument for thinking that non-structural QW views, such as that offered by him (2016, 2017) or Smith (2005, 2008), better assess the cases of the willingly and unwillingly addicted agents in accordance with intuitive reactions than mechanism-based RR views. In section 2, I summarize Sripada's clarifications to the cases and argue that, even with his clarifications, the cases invite implicit assumptions that complicate the determination of which features are being tracked by intuitive assessments. I offer three different descriptions of the cases that attempt to hold fixed the relevant features to make it more apparent which features assessment are indeed tracking. These descriptions

---

[99] As noted in the introduction, here I assume in all of the cases that the addicted agent is not culpable for their addiction. Sripada also explicitly makes this assumption. I take up this topic in more detail in chapter 6.

attempt to account for the nature of the wrongdoing committed and the two main ways in which

self-control might be lost. However, as I offer these revised cases, I will argue that further

clarifications do not support the idea that non-structural QW views are more extensionally

adequate than capacitarian RR views. Rather, the cases either reveal significant problems with

trying to stipulate a complete loss of control or they reveal that capacitarian RR views can better

track and explain important differences between the cases. In section 3, I conclude by

suggesting that these highly stipulated cases not be treated a litmus test of which view of moral

responsibility better tracks the features most central to assessments of moral responsibility. I

propose instead that the views be tested by whether they can track and explain the relevance of

intuitively important features in cases based on more realistic conceptions of addiction and

control.

1.1 The Addicted Agents

As discussed in chapter 2, Frankfurt's describes the hypothetical willingly and

unwillingly addicted agents such that both are irresistibly compelled to act on their motivating

desires to use addictive substances but they differ in their higher-order evaluations as to the

merits of these motivating desires. While the willingly addicted agent endorses his motivating

desires, the unwillingly addicted agent does not. The unwillingly addicted agent does not merely

disapprove of his addictive desires but is psychologically distraught. His profound disapproval is

exhibited both psychologically and behaviorally. He "struggles desperately" and "tries

everything" to resist, but his ultimately "helplessly violated" by his addictive desires (12). In

contrast, the willingly addicted agent regards his addictive motivation with plumb joy. He is

"delighted with his condition" and "would do whatever he could to reinstate" his addicted

condition if he desires waned (19-20). On account of their differing evaluative attitudes, the two

addicted agents are thought to be worthy of differential assessments of responsibility as well, despite the fact that both are utterly compelled by irresistible desires to use an addictive substance. While the unwillingly addicted agent is intuitively not morally responsible for his substance use, the willingly addicted agent is. Plus, the fact that one agent is intuitively morally responsible for his substance use though neither agents possess *any* control over their action is meant to illustrate that control over an action is not necessary for being morally responsible for that action.

1.2 Competing Responses to the Willingly Addicted Agent

Sripada contends that best theory of moral responsibility is the one that accurately captures the willingly addicted agent as fully blameworthy in contrast to his unwilling counterpart. He argues that this extensional condition is where RR views fall short in comparison to non-structural QW views. His concern is this: control-based views are committed to assessing the degree of blameworthiness of the two hypothetical addicted individuals exactly the same despite the intuitive plausibility that they deserve differential assessments. Intuitively, the unwilling addicted agent *is not* morally responsible while the willingly addicted agent *is* fully responsible (783).[100] The result, he suggests, is that control-based views are overly permissive in their extension of excuse, which indicates a larger theoretical issue with theories that take control to be a primary concern in assessments of responsibility.

---

[100] As noted in chapter 1, I use the terms like "responsible," "culpable," and "blameworthy" interchangeably unless I am specifying a distinction. In his paper, Sripada distinguishes between being morally responsible and being either praiseworthy or blameworthy. If an agent is blameworthy, they are morally responsible, but it does not follow that if an agent is morally responsible, then they are either blameworthy or praiseworthy. Sripada's commitment here is rather confusing since, in this very same paper, he advocates for a "single kind" or "unity" of moral responsibility that does not distinguish between notions of moral responsibility (797). I say more about his responsibility/blameworthy distinction below when I discuss addiction and wrongdoing.

To show that RR views are indeed committed to this extensional account, Sripada characterizes the core claim of RR views as this:

> " [T]o be morally responsible for A-ing, the person must be in control of A-ing itself" (792).

Because the willingly addicted agent lacks the ability for self-regulation, the consequent of the statement is false and it follows that the antecedent is false as well. To demonstrate why RR views are committed to holding the consequent false, he specifically considers Fischer and Ravizza's (1998) mechanism-based Reasons Responsiveness (RR Mechanism) account, which he describes as committed to this statement:

> **(RR Mechanism)** An agent is morally responsible for A-ing only if, holding fixed the mechanism on which S acts, in a suitably broad range of scenarios in which there is sufficient reason to do otherwise, S recognizes the reasons to do otherwise and does otherwise.

The same motivational mechanism always mediates substance use and so" in every possible scenario in which there is sufficient reason to do otherwise than he does, the addict still uses the drug" because the mechanism cannot be self-regulated (791). Thus, under this understanding of control, and how it is lost in the case, the willingly addicted agent completely lacks control over his substance use.

In contrast to mechanism-based RR views, Sripada argues that non-structural QW views offer the right extensional account of the willingly and unwillingly addicted agents. The willingly addicted agent is fully morally responsible because his substance use is sufficiently motivated by the practical judgments based on his fundamental evaluative stance. QW views contend that blameworthiness for a wrongdoing turns on the degree to which an action reflects an agent's privileged evaluative attitude:

(Non-structural QW) An agent is morally responsible for A-ing iff and to the extent that the agent's fundamental evaluative attitude provides sufficient motivational support for A-ing.[101]

Sripada defends a conative version of the non-structural QW account on which the attitudes that comprise an agent's cares serve as the privileged evaluative attitude by which an agent should be assessed (785). Whether the wrongdoing is sufficiently motivated by an agent's most fundamental cares determines whether the agent is morally responsible for that action.

Sripada suggests that other non-structural QW and structural QW views likewise have the right extension in regard to the willingly addicted agent, even if these views might be rejected for other reasons. While he does not explicitly name other views, one such cognitive view might be that defended by Smith (2005, 2008, 2012) who posits the idea that an agent's fundamental *evaluative judgments* serve as the privileged evaluative attitude by which an agent might be assessed on account of their actions. Sripada contends that either non-structural QW views will appropriately sort the cases of the addicted agents since the views rightly track the differences in their respective fundamental evaluative attitudes.

2. Rightly Assessing the Addicted Agents

In section 2, I will complicate the claim that the right extensional account of the cases is one that sorts them as Sripada assumes. I will argue that there are two major points of indeterminacy in the description of the cases that impact how we might assess the agents. These points concern the presumed moral wrongdoing in the cases and the various ways in which self-control might be stipulated as lost. I will argue that when these two matters are specified, the

---

[101] Sripada does not give a more formal statement of the non-structural QW position. I build this statement from what he says about evaluative cares and motivation in Sripada (2017, 785) and Sripada (2016, 1216).

descriptions either introduce new indeterminacies or capacitarian Reasons Responsiveness views can offer a better account of features in the cases that seem important for assessments of responsibility.

2.1 Addiction and Wrongdoing

Presumably, the cases of the addicted agents illustrate the features most key to assessments of blameworthiness as discerned by intuitive responses to their substance use. But what exactly do the agents do wrong that might make them worthy of being deemed morally responsible? In other words, what exactly is the action that intuitions are tracking in the cases? To see why specifying the wrongdoing is so important, consider that the cases of the willingly and unwilling addicted agents are meant to match in all relevant details except for their evaluative stance with respect to their addictive substance consumption. Holding all other variables consistent shines a light on whether an agent's evaluative stance is the central feature in assessments of responsible agency. Yet, without specifying the type of wrongdoing, it is easy to unwittingly fill in the details of each case such that different wrongdoings are ultimately compared. As a result, it is unclear whether assessments are tracking differences in the nature or severity of wrongdoings or differences in evaluative stances. This possible confusion is especially relevant for cases of addiction, which tend to be morally charged and invite presumptions about control.

To see the potential for confusion, consider again the basic descriptions of the cases. The willingly addicted agent's willingness is constituted by the fact that he values substance use above his other interests and obligations. With this basic description, one can easily visualize a flagrant partier who cares most about substance use and generally neglects other competing

values. Yet, the more we elaborate on the acts that constitute or manifest an evaluative

willingness to pursue and use addictive substances, the more likely implicit assumptions about

wrongdoing are imported into the details of the case. When an agent prioritizes something like

hedonic gratification received from narcotics above all else, it is not hard to imagine the

numerous possible wrongdoings that are directly explained by this prioritization.[102] Even if this

agent has no children that might be neglected on account of his substance use, living a life

"imbued throughout with hedonic satisfaction" directly tied to irresistible addictive desires is

bound to directly result in wrongdoing to others (793). Risky, substance-related behaviors can

endanger the safety or well-being of others, such as driving while inebriated or using substances

at work.

In contrast, what exactly does the unwillingly addicted agent do wrong? To

communicate the strength of his unwillingness, the description of the case implies that this agent

spends considerable time and energy planning and implementing his failed attempts at resistance.

Yet, it is not at all clear that his substance use is implicated in any wrongdoing or that substance

use is even an action for which he might be held morally accountable. Moreover, some features

of the case suggest that this agent might even be deserving of admiration and possibly moral

praise. Even if his planning and efforts are not ultimately successful, his perseverance and

dedication devoted to resisting an action he evaluates as not worthwhile suggest that he is worthy

of our esteem. Even assuming that his substance use is a moral wrong for which he might be

---

[102] For instance, consider that theorists who elaborate on the willingly addicted agent's willingness tend to describe the neglect of moral obligations or general bad behavior (e.g. Wallace (1994, 172); Flanagan (2017, 72)).

assessed as responsible, our intuitive assessments of this action might be balanced in aggregate by our assessments of his resistant activities.[103]

If these factors at all influence our intuitive reaction to the cases, then very different cases are being compared. While the willingly addicted agent wrongly neglects his obligations as he prioritizes substance use, the unwillingly addicted agent admirably exerts himself to take on the difficult task of resisting use. This point is important because even if the cases are rightly sorted as Sripada suggests, with one deserving mitigation while the other is fully blameworthy, there are two plausible explanations for this categorization of the cases. On the one hand, reactions to the cases may be tracking differences in the addicted agents' evaluative stances toward their substance use. On the other hand, reactions might just be tracking these fundamental differences in the moral quality of their actions.

Interestingly, Sripada does recognize that the action done by the willingly addicted agent needs specification to illustrate that the agent is a worthy target of assessments of responsibility. But rather than specifying the nature of the wrongdoing in the cases to make the relevant details parallel, he offers an alternative case of willingness in which the action is flagrantly and obviously wrong. He presents the case of the "willing exploiter" whose desires to watch exploitive pornography are directly motivated by his "narcissistic kind of self-love" and his attraction to a "position of dominance over others" (802). Sripada makes the case that while the willingly addicted agent is only *morally responsible* for his substance use, the willing exploiter is actually *blameworthy* for his actions. On the assumption that being blameworthy is sufficient for being morally responsible, and given that the actions of the willingly addicted agent and the

---

[103] Thank you to Dana Nelkin for raising this insightful point.

willing exploiter have the same etiology in that both are directly motivated by evaluative

judgments driven by their fundamental evaluative attitudes, Sripada argues that the willingly

addicted agent is *morally responsible* for his action even if not blameworthy.[104]

There is much to say in response to this analogical appeal to a different case.[105] Here,

though, I will set aside these concerns because my contention is that the cases of addiction

already do implicitly incorporate actions for which an agent might be blamed (or even praised).

So, rather than relying on a supposedly analogous case that carries with it its own confounding

details that invite implicit assumptions, we should instead consider a plausible wrongdoing

directly explained by addictive substance pursuit or use that would make the target action both

readily apparent and parallel between the willingly and unwillingly addicted agents. The

obvious candidate for wrongdoing might be the act of drug consumption itself. However, it is

not at all obvious that drug consumption in and of itself should be considered a moral wrong, and

this focus invites in the implicit assumptions and disparities I mention above.[106]

---

[104] If the willing exploiter's action is aetiologically identical and his action is more clearly blameworthy, why does Sripada not just focus on this type of case rather than the case of substance addiction? I would suggest that it is because we would run into the same intuitive difficulties with comparing the case to an unwilling exploiter. For instance, it is far less intuitive that an agent who unwillingly experiences and fails to resist an irresistible desire to view exploitive pornography is worthy of being fully excused of blameworthiness for the wrongdoing of seeking out and viewing such images.

[105] At this point, we might also question what exactly makes the willingly addicted agent's substance use an action for which he might be *morally responsible* if not also *blameworthy*. Sripada claims that the etiology of the actions of the addicted agent and the exploiting agent are the same and so both are morally responsible. It just also happens that the latter agent does something obviously morally wrong, which also makes him blameworthy. But, again, what does the addicted agent do such that he is even *morally* responsible. Is it just that the agent values substance use such that he cares about it, and this evaluative care is what makes the act a moral act? If this is the explanation, this categorization does not communicate anything interesting about how we might sort the cases according to moral responsibility because it is not clear that the unwillingly addicted agent's action is one for which he might be held morally responsible since the action does not share the same etiology as the willingly addicted agent's action or the action of the willing exploiter. With all of this pondering aside, I do think that the willingly addicted agent is often assumed, at least implicitly, to have done something morally wrong, thus making him a potentially appropriate target of blame. As such, I will set aside these worries about the finer points of how Sripada is trying to distinguish being morally responsible from being blameworthy via an action's etiology.

[106] Arpaly and Schroeder (2013) make the observation that there are many situations in which substance

To see a plausible wrongdoing that any addicted agent, willing or not, might do when experiencing irresistible desires for substance use, recall the experience of Carr summarized in chapter 2. Again, though his infants are not physically harmed by being abandoned in a car in sub-freezing weather while he uses cocaine, they are unjustifiably put into great peril. This example is a clear instance of how addiction, and acting on addictive desires, can be directly implicated in moral wrongdoing.

With this wrongdoing in mind, consider the addicted agents again but imagine that both agents act similarly to Carr. They are both irresistibly compelled to substance use and their addictive cravings play a significant explanatory role in why they both gravely risk the well-being of their young children by neglecting their interest of safety while pursuing and using addictive substances. The unwillingly addicted agent acts on his irresistible desires to pursue and use an addictive substance despite valuing the well-being of his children higher than substance use. In contrast, the willingly addicted agent is also irresistible compelled to seek and use an addictive substance, but this action is also motivated by his value of substance use above his children's well-being.

With the wrongdoing clearly in mind, is there a clear intuitive answer to how the cases should be sorted with respect to blameworthiness? Some still might say that the unwillingly addicted agent is worthy of being excused for blameworthiness for the wrongdoing while the willingly addicted agent is not, but it is less apparent that this categorization of the cases is

---

consumption seems morally neutral on many moral theories (p. 217). See also Husak's (2000) survey of the goods that substance consumption can offer. Someone might say that addictive substance use is a moral wrong in the sense that an addicted agent violates duties to their own selves even if no one else is wronged. The idea here is that the agent fails to meet his obligation of self-improvement or self-care. While I think that this thought is worthy of exploration since it underlies some aspects of the popular discussion of addiction and wrongdoing, my focus here is on interpersonal moral responsibility.

intuitively the "right" one.[107] Moreover, even if this is the right intuitive categorization, there are still competing plausible explanations for this intuitive response. Again, even if both agents wrongly risk the well-being of their children, this wrongdoing is not the only action that seems important. The unwillingly addicted agent may be intuitively blameworthy for the wrongdoing and yet also praiseworthy for taking every effort not to succumb to his addictive desire, which again raises the aggregate assessment point made above.

But even if it were granted that the blameworthiness of these re-described cases be sorted as Sripada suggests, I would contend that it is still not obviously apparent that intuitions are tracking only differences in the agents' evaluative stance. In the next section, I will show how assumptions about control likely play a role in intuitive assessments of the cases as well, complicating further exactly what qualifies as the "right" extension with regard to the cases.

2.2 Sripada's Clarifications About Control

Sripada (2017) argues that, as Frankfurt's addiction cases are originally described, key details concerning control are indeterminant in a way that invites confusion in both intuitive reactions and explanations for these reactions. One apparent point of confusion is that the vagueness in the descriptions of the original cases invites the intuitive reaction that perhaps something is amiss with regard to the willingly addicted agent's evaluation of substance use, suggesting that his control over these evaluations is impacted by his addiction as well. This is a worry for the use of the cases as a test to clearly sort views of moral responsibility because an impaired ability for evaluating substance use would call into question whether the willingly

---

[107] I tested these cases in a 2021 seminar with undergraduate and graduate students. While many did sort the cases according to blameworthiness as Sripada suggests, many others either considered both or neither to be blameworthy. The reasons offered by the students for their positions, though, often did not track the evaluative stances of the agents in the cases.

addicted agent possess the sufficient ability for tracking the considerations against substance use.

Anticipating and responding to such related worries, Sripada offers key clarifications of the cases

to forestall "certain misreadings" that contribute to the persistence of the position that

considerations of control are an important factor for evaluating responsibility in these cases.

These clarifications are meant to show that the cases can plausibly illustrate a complete lack of

control over substance use.

Wallace (1994) offers one such "misreading" that Sripada might have in mind. In his

reading of the willingly addicted agent, Wallace offers the consideration that the total loss of

control stipulated in the case plausibly impacts the willingly addicted agent's practical reasoning

as well as his self-control. This impaired practical reasoning might be experienced long-term or

episodically, depending on the ebbs and flows of addictive desires. For the first sort of case,

Wallace explains:

> addiction is pictured as functioning (as it were) continuously through a person's
> life. It is not merely a source of desires that are practically irresistible. Rather,
> the condition of being addicted is taken to affect the agent's capacities for
> practical reasoning, depriving the addict of the power to think clearly about those
> aspects of his life affected by the addictive desires. On this model, it may well be
> the case that the addict is substantially and persistently deprived of the ability to
> choose to refrain from engaging in the activities that the addictive desires lead
> him to find attractive. Thus, the addict's inability to think clearly about his
> addictive activities may make him unable to accept the reasons that count against
> those activities, and so unable to make choices in accord with those reasons."
> [Wallace, 1994, 171][108]

As Wallace interprets the case, the willingly addicted agent is intuitively worthy of some degree

of excuse because his supposed willingness is suspect in a way that suggests his reasons

---

[108] The sort of agent Wallace seems to have in mind here is what I call the "diachronically stable willingly addicted agent" in chapter 2 section 1.5.

responsiveness if undermined. Specifically, the agent lacks sufficient control over his evaluations regarding substance use to be fully morally responsible for his addiction-explained actions. As Wallace clarifies, "even if the agent is "willing" to be subject to his addiction, the addictive circumstances in which he came to identify with this condition are circumstances in which the agent lacked the powers of reflective self-control" (172).

Again, anticipating such readings, Sripada offers a series of clarifications about control and how control is lost in the cases. His first clarification is to show how control over an action might be plausibly undermined by irresistible desires. He offers a familiar psychological motivational architecture and explains how one might experience a total loss of control on this picture (789-791).[109] As he presents it, actions are motivated through either practical reason or some other psychological source like "spontaneous" desires or emotions. These two sources of motivation can work together. Practical reasoning can augment spontaneous motivation, or vice versa. However, spontaneous motivation can also run counter to practical reason, as seen when an agent is highly motivated to act in opposition to their best judgment. Sripada proposes that when practical judgments and motivation conflict, the errant motivation can usually be reigned in through the agent's psychological "capacity for resistance" or "self-regulation." Sometimes, though, this capacity fails because the spontaneous desire is too strong. A desire is considered "irresistible" if the capacity for resistance *cannot* overcome it and the desire leads to action. This capacity for resistance can be compromised piecemeal. An agent might effectively regulate some types of errant spontaneous desires but fail to regulate others.

---

[109] See chapter 2 for a detailed summary of this commonly referenced motivational architecture.

Applying this psychological architecture to Frankfurt's cases of addiction, he explains that the addicted agents experience constant irresistible desires to use addictive substances because the motivational strength of their substance-directed desires is consistently stronger than the strength of their capacity for resistance or self-regulation (789). This loss of control over drug use is straightforwardly exhibited in the case of the unwillingly addicted agent. His practical judgments and spontaneous addictive desires are at odds, but his capacity for resistance is too weak, as depicted by his failed efforts. Sripada suggests that the willingly addicted agent's loss of control can be appreciated counterfactually. Were he to judge it best to resist substance use, he could not act on this judgment due to the strength of his spontaneous addictive desires.

Sripada's second major clarification is that an agent's capacities related to practical judgments can remain completely intact even while the agent experiences significant impairment in their capacity for self-regulation. As he explains, a deficit in the capacity for self-regulation does not imply a deficit with other key capacities implicated in behavior," for example deficits in [one's] capacities for deliberation, practical judgment, desire, or anything else" (790). This stipulation makes it plausible that addictive desires, though impossible for the two addicted agents to regulate, are *not* influencing and undermining other psychological capacities, such as judgment or beliefs concerning addiction or the worthwhileness of substance use.

Sripada's third major clarification concerns how to understand the willingly addicted agent's "willingness" for substance use, especially in light of the motivational fact that he motivated to substance use by irresistible desires. As Sripada characterizes willingness in the case, the agent "has narcotic-directed desires and he genuinely loves using narcotics; he truly would not change a thing" (798). He can assess moral considerations and appreciate and appropriately weight his practical options, but he ultimately considers substance use to be the most important value above

"family, friendships, and satisfying work" (794). While the independent strength of the spontaneous motivating addictive desires is the most proximate motivating factor in the psychological etiology that leads to substance use, it is this practical judgment that favors substance use. Sripada argues that, even as the distal cause, the practical judgment is the most explanatorily important aspect in the mechanism that motivates the agent's substance use. As he explains:

> the direct cause of the willing addict's using the drug is his desire to use the drug, a desire that as it happens is irresistible. This desire arises spontaneously and is sufficient by itself to ensure he uses the drug. But as it turns out, this desire does not operate alone. In addition, there is also an indirect factor that is part of the aetiology of his action, namely, the addict's practical reasoning. Its effect on action is mediated by its effect on the direct factor, that is the addict's drug-directed desire. [Sripada, 2017, 790-1]

The willingly addicted agent's practical judgment "augments" the irresistible spontaneous desire. As such, the substance use is overdetermined on account of two independently originating motivational sources that work together to effectively motivate substance use. On this picture, the agent can be said to willingly use addictive substances while also acting, first and foremost, on an irresistible desire to use the substance.

All together, Sripada argues that these clarifications to the cases show how the addicted agents can experience total loss of control in regard to substance use while also maintaining distinctive and clear-eyed evaluations of their actions. Further, these clarifications are meant to show that the willingly addicted agent's willingness is not itself a product of undermined control due to his addictions. As he stipulates the details, the willingly addicted agent's ability for practical judgments are *not* impacted by his addicted condition. Rather, the cases only exhibit a

profound but isolated loss of *self-control.* He contends that, with these clarifications in place, only QW views rightly assess the responsibility of the addicted agents.

2.3 The Many Abilities Involved in Self-Control

In this section, I will contend that Sripada's assumptions concerning loss of control over substance use in the cases is counter to common assumptions about how control might be exercised or lost in addiction. This difference complicates intuitive responses to the cases by inviting presumptions that the agents possess control over their addictive behavior despite stipulations that such control is absent.

In Sripada's clarifications, he focuses on a very specific way in which control is lost in the cases. Namely, the agents' substance consumption results from acting on a motivating desire that they cannot self-regulate. As described by Sripada, self-regulation seems to be a synchronically performed mental act. In the moment of experienced addictive desires, the addicted agents are unable to mentally regulate the psychological mechanism by which the agent experiences and acts on the desires. The narrow focus on the implicated psychological mechanism in substance use, and the characterization of lost control in these terms, is in part catered to the notion of control assumed by mechanism-based RR views, which approach reasons responsiveness in terms of the specific psychological mechanism implicated in the action in question. If indeed an addicted agent's substance consumption results from the exact same mechanism every time, and indeed the mechanism defies self-regulation in the moment every time, then on the mechanism-based RR view, the *mechanism* is not reasons responsive.

However, as I suggested in the introduction, understanding reasons responsiveness in terms of mechanisms has notable explanatory limitations compared to understanding it in terms

of capacities that belong to an agent. To see this, consider how we tend to assume that control is exercised or lost in daily life. When we set out to control the actions that might follow from our own errant desires, or in our expectations and assessments of others, we commonly take into account the host of different abilities that can be called upon to help with this regulation. If I want to resist my desire to sleep in in the morning, I do not solely rely on my mental effort to single-handedly resist this desire in the moment when I am tempting to press the snooze button again. Instead, I take steps in advance to make sure that I do not fall back asleep. I place my alarm clock across the room; I train my dog to expect an early morning walk; or I request help from my spouse to pester me if I sleep past a certain point. An addicted agent might use similar methods for resisting errant desires. If an agent cannot consistently regulate their addictive desires that regularly lead to them drinking an entire six pack of beer in the fridge against their own better judgment, we do not simply concede that the individual must lack the ability for self-control. Rather, we tend to ask questions that consider all of the reasonable ways in which the agent might avoid drinking the beers. Could they without much difficulty avoid stopping at the store to buy the pack of beer? When they start thinking about drinking the beer, could they call a friend to come take the beer or dump it out? Could they leave the house to meet the friend and avoid the beer until the craving past? Resisting a desire in the "hot" moment of temptation using sheer mental effort alone can be hard for most everyone. However, if our goal is to resist a temptation, sheer mental fortitude in the face of a "hot" desire is not our only self-control resource for resistance.

This common approach to how control is exercised suggests that rather than focusing on the single specific psychological process implicated in an action and whether that mechanism can be mentally regulated for understanding whether someone can control an action, the relevant

notion of control – and the one that is likely at least implicitly assumed when assessing the cases – for assessments of moral responsibility should take into account the abilities" tool set" that individuals can draw from to do the right thing for the right reasons. This is precisely what capacitarian RR views propose to do. In considering whether an agent has control over a specific wrongful action, the assessment takes into account the specific abilities an agent can wield relevant to the action in question.

Recalling what I discussed in chapter 1, capacitarian RR views assume that agents have a host of abilities that comprise reasons responsiveness that can be generally organized under two categories that I will refer to as *cognitive competence* and *volitional competence,* and collectively call *normative competence*. Again, these categories are general headings for capturing what the abilities accomplish. Certain psychological abilities are not indexed or solely confined to one category or the other and various individual psychological abilities can be relevant to both types of competences. For instance, the abilities one might use to track which moral considerations they should act upon might also be recruited in planning how to act in light of the considerations.

As capacitarian RR views understand cognitive and volitional competence, both capacities are required at a moderate degree in order for an agent to be assessed as fully blameworthy for a wrongdoing. This means that if either capacity is significantly compromised, the agent is deserving of some degree of excuse. Being *cognitively competent* means being able to appreciate which moral requirements bear on a situation and how to appropriately weight the strength of competing moral considerations. This category of competence involves abilities related to moral and prudential reasoning and practical judgment, but it can also recruit a range

of other cognitive, emotional, and attentional abilities, like comparative reasoning, empathy, and self-directed focus.[110]

Volitional competence is the capacity to implement the reasons that one recognizes into action. Importantly for this discussion, this category involves a range of psychological abilities that can be exercised *in the moment* or *over time* related to acting on one's practical judgment. Such psychological abilities include developing, guiding, and regulating motivation over time "despite distraction, temptation, and other forms of interference" (Brink, 2021). Volitional competence exercised in the moment recruits abilities to resist or *conquer* an errant motivating desire or to *summon* motivation to do something that one feels unmotivated to do. Conquering an errant desire or summoning motivation in the moment both involve exerting sheer conscious mental effort. An addicted agent might be able to mount the motivation in the moment to resist acting on a motivating addictive desire to pursue and consume a substance. However, when confronted with persistent errant desires, this sheer mental effort is often difficult to mount and hard to sustain in repeated situations over time. Resolve fails and mental regulation requires mental space and attention that is required for other tasks and can be emotional or hedonically costly.[111] Using only mental effort to self-regulate has limited efficacy in producing or guiding action over time if this is the only psychological ability used.

Given the limitations of conquering errant desires or summoning motivation, acting on one's practical judgments in day-to-day life typically involves volitional abilities other than sheer

---

[110] Nelkin (2011, 21-27) provides an especially good discussion on the role of emotions in cognitive competence. Vargas (2013, 203 ft note 6) also mention emotions as well in his defense of the idea that responsiveness is not overly rationalistic such that it excludes emotions or affective attitudes.
[111] Sripada (2019).

mental exertion.[112] One such ability is circumventing errant desires either by avoiding inciting

the desires in the first place or by having a plan in place to resist acting on the desires even if

they are incited.[113] Accomplishing or avoiding certain actions is often successfully done

indirectly through setting intentions, pre-planning, establishing personal rules, setting up external

support structures, and committing to not reconsidering resolutions.[114] Circumventing errant

desires or flat motivation can involve making use of resources like supportive friends, family,

medical providers, programs, or institutions. Together, these abilities work together to either

avoid inciting errant desires in the first place or avoid finding oneself in a motivational slump, or

to figure out strategies for dealing with errant desires that may be incited or with flat motivation

once experienced.

These distinctions in how self-control might be exercised bear on the discussion of

addiction in our hypothetical cases as well. An addicted agent can lack the ability to directly

mentally regulate their desire for opiate use, but they might possess the ability to motivate the

action to take their daily dose of the pharmaceutical Suboxone that dampens their cravings to the

point that they can regulate their errant desire. Similarly, rather than using sheer mental effort to

summon motivation in the face of a "hot" craving, an agent can structure activities in advance to

nurture motivation to avoid the mental confrontation. An addicted agent may not be motivated

to attend an AA meeting or call a medical office to schedule a visit with a doctor to refill a

pharmaceutical that can lessen the experience of addictive cravings, but they may be able to

---

[112]See Kennett & Smith (1996) for a defense of the idea of diachronic control. See also Snoek, Levy, & Kennett (2016) for a defense of the idea that recovery from addiction involves diachronically developed and implemented strategies.

[113] See Bermúdez (2021) for a compelling argument that possessing an advanced skill of self-control results in having to utilize self-control less because the experience of conflicting motivations is more skillfully avoided.

[114] Holton (2009, 112-136); Ainslie (2001). Consider also the many techniques for resisting unwanted addictive habits carefully explained by Dill and Holton (2014).

muster enough motivation to have their AA sponsor visit them at their house or to enter a program that rewards them with a voucher if they show up at pre-set time at a medical office to meet with a clinician and pick up their prescribed pharmaceuticals.[115]

Notably, implementing these techniques requires forethought. An agent needs to reflect, plan, and act on these plans to successfully comply with their practical judgments. This means that self-control, as the notion is commonly conceived, is not just something that is exerted in the moment of addictive temptation with sheer mental effort, but instead involves a host of other abilities related to deliberation, practical judgment, and emotional regulation. These various abilities are called upon to do things like avoiding inciting the desire or putting oneself into a situation in which one cannot act on the desire even if it is incited.

If I am right, then the cases can be described in several ways depending on the notion of self-control assumed to be undermined. The ability to resist an errant desire using only mental exertion can be undermined and the ability to circumvent the errant desires can be undermined. In the next section, I will consider two versions of the cases with these notions of self-control in mind. I will argue that in one version implicit assumptions about control still heavily influence intuitive responses. But I will argue that when we consider a version that better aligns with how people tend to lose control, capacitarian RR views can readily sort and explain the cases better than non-structural QW views.

2.3 Conquering, Circumventing, and Cognitive Control

---

[115] The second suggestion here is based on the idea of contingency management programs that help structure motivation by focusing attention on small rewards based on day-to-day behavior rather than large rewards that are achieved only through long-term accomplishments comprised of a series of behaviors. Elsewhere, I defend such programs against those who question whether they are morally justifiable (Koepke, In Progress). See Lussier et al. (2006) and Ginley et al. (2021) for reviews of the efficacy of contingency management programs.

Given the distinction between conquering and circumventing an errant desire, the cases can be stipulated to emphasize a *complete loss of volitional control* with respect to substance use. This means describing the cases so that both agents are stipulated to lack the capacity to either conquer or circumvent their addictive desires. Moreover, to truly lack volitional control over the wrongdoing, it must also be supposed that the agents lack the ability to place their children out of harms way while they pursue and use their addictive substance. Again, the two agents are different in respect to their evaluations of substance use and the wrongdoing. The unwillingly addicted agent recognizes that the risk he takes with his children in order to use substances is morally wrong. He tries to overcome or resist his desires for substance use because he overall values their well-being more than satisfying his additive desires. In contrast, while the willingly addicted agent recognizes, or at least is able to recognize, the wrongness of the risk he takes with his children in order to use an addictive substance, he greatly values substance use. He does not try to resist his desires because he values them above the welfare of his children.

Yet, despite these dissimilarities, the two addicted agent share the following features:

- their abilities related to practical judgment, like conceptual thinking, attention, concentration, practical judgment, and empathy are fully exercisable and untouched by their addiction;

- both agents *are unable to resist substance use either by mental exertion or by circumventing their addictive desires*. This means they are unable to avoid inciting the desires in the first place by avoiding situations, people, or psychological states that "trigger" their addictive desires. Further, they are unable to effectively arrange their

circumstances to make the substances less accessible, e.g. by committing to a treatment facility;

- both agents gravely risk the well-being of their young children by neglecting their interest of safety while acting on addictive cravings; and

- both agents are *unable to figure out how to effectively avoid risking the well-being of their children*.

These specifications are meant to illustrate a more profound lack of control to prevent implicit presumptions that either agent might be able to effectively avoid substance use or the wrongdoing. Akin to the original Frankfurt addiction cases, the parallel loss of control between the cases is meant to highlight the agents' respective evaluations of their actions. The unwillingly addicted agent seems obviously unwilling, as evidenced by his attitudes and actions. He explores all methods of conquering and circumvention. In contrast, the willingly addicted agent is truly willing. He pursues substance use and does not even consider resisting use, preferring instead to prioritize it over the well-being of his children.

As described, intuitive reactions may still tend to hold that the willingly addicted agent is blameworthy for the wrongdoing, while the unwillingly addicted agent is not blameworthy, or at least is not *as* blameworthy as his counterpart. Yet, are intuitions tracking the difference in evaluative stance or something else? Here, I will suggest that presumptions about existing volitional control can *still* play a role in these cases despite the more detailed stipulation that such control is absent. Here is why.

While it is quite imaginable that an agent has extreme difficulty resisting a strong and persistent errant desire using only their direct mental effort alone, a complete absence of ability

to circumvent acting on this desire is harder to fathom. Take the unwillingly addicted agent. Despite the above stipulations, it is hard to accept that *all* of his circumventing efforts are bound to fail. His practical reasoning is stipulated as untouched by addiction. Presumably, then, he is fully aware of the needs of his children and the fact that these needs should be prioritized. And yet, despite this clarity in practical reasoning, and despite all of his effort, he cannot figure out how to effectively ask a neighbor for help or call the police to come take his children because he is unable to care for them. Surely at some point he can make an effective plan to avoid risking his children's well-being while using his preferred addictive substance, even if he cannot make an effective plan to avoid substance use.

With these thoughts in mind, intuitive reactions to the case might be tracking one of two things: either the unwillingly addicted agent has some ability for circumvention control, though perhaps not enough to be considered fully blameworthy, *or* his practical reasoning is impaired such that, if only while experiencing intense cravings, he is sufficiently unable to determine what he has sufficient reason to do regarding substance use and caring for his children. If the former is the case, the intuitive reactions might be tracking the presence of control over the wrongdoing rather than the agent's evaluative attitude. But if the latter is the case, intuitions might be tracking *sufficiently impaired cognitive control* in addition to impaired volitional control, which would make this agent worthy of at least some degree of mitigation for their wrongdoing. To see what I mean, consider that many of the abilities regularly recruited for volitional competence are recruited for cognitive competence as well. It is hard to reconcile the idea that the agent's abilities are so utterly compromised that he could never effectively avoid substance use or avoid recklessly endangering his children for substance use with the idea that his ability for recognizing and appropriately weighing his practical considerations is completely untouched.

Thus, even with this further stipulation about the loss of control in this case, intuitions might still be tracking either a presence of circumvention control or a sufficient loss of cognitive control. Either way, there is reason to question whether intuitive reactions are straightforwardly tracking the agent's evaluative stance alone or even at all.

A similar dilemma is encountered in regards to intuitive reactions to the willingly addicted agent. On the one hand, it is hard to imagine that the agent's volitional control with regard to the wrongdoing is truly lost, given that there is no evidence in the description of his actions that might suggest such a profound loss in abilities. It is hard to imagine that he is truly unable to avoid placing his children in harm's way given how much else he is able to figure out to promote his substance-use related values. On the other hand, as with the first case, the more that is said to explain the complete loss of volitional control, the less plausible it is that the agent's cognitive competence remains untouched. A further emphasis on the loss of volitional control makes it quite plausible that this agent experiences more profound problems with their practical reasoning as well, akin to Wallace's worry mentioned above. This brings us to a similar dilemma as the one articulated in the case of the unwillingly addicted agent: even with these further stipulations about loss of volitional control, intuitions might be tracking either a presence of circumvention control or a sufficient loss of cognitive control. Either way, it is difficult to know whether reactive intuitions are tracking what Sripada supposes they are tracking, i.e. the difference in the agents' evaluative attitudes. It is simply hard to imagine the loss of control of volitional control as stipulated without assuming either some control is still possessed or assign that practical reasoning is sufficiently impaired. Either way, presumptions about control likely still play a role in why we react as we do to the cases. And when the cases are revised to try to clarify how volitional control is lost, it becomes hard to maintain the idea

that the agents' cognitive competence is not impacted by their addiction as well, which complicates how the cases are rightly sorted according to blameworthiness.

As a last ditch effort, the cases might be stipulated in a way that limits the loss of volitional control to only the ability for *conquering* addictive desires. Given common experiences, it is more plausible to imagine that someone might have a sufficient ability for practical reasoning while also having great difficulty regulating an errant desire using only sheer mental effort. As such the cases will share the following features:

- their abilities related to practical judgment, like conceptual thinking, attention, concentration, practical judgment, and empathy are fully exercisable and untouched by their addiction;

- both agents *are unable to resist substance use either by mental exertion alone;*

- both agents gravely risk the well-being of their young children by neglecting their interest of safety while acting on addictive cravings; and

- both agents are *able to circumvent their addictive desires or figure out how to avoid risking the well-being of their children* when they pursue or use addictive substances. For instance, they are able to arrange their circumstances to make it more difficult to access drugs or they are able to place their children in the care of others to make sure they are not harmed by their pursuit or use of addictive substances.

As described, the details of the cases can be easily imagined without inviting major implicit assumptions about existing control that might confound the comparisons since the description aligns with common notions about how control is lost. It is also more intuitively plausible that the agents' ability for cognitive control is untouched.

Yet, as stipulated, intuitive reactions to these cases defy sorting as Sripada suggests. The unwillingly addicted agent seems blameworthy along with the willingly addicted agent. If this is right, capacitarian RR can easily track the features that straightforwardly explains why both agents should be assessed as blameworthy for their wrongdoing. Namely, both agents possess both cognitive and volitional control with regard to the wrongdoing. They may be unable to resist their addictive desires through sheer mental exertion, but they can appreciate the moral wrongfulness of placing their children in danger to act on their addictive desires and they are able to circumvent either using their preferred addictive substance or subjecting their kids to harm on account of their substance use. Both can recognize that their substance use directly entails wrongdoing and both are able to avoid this action.

Do non-structural QW views likewise get the right extension in categorizing the cases? Maybe. They rightly assess the willingly addicted agent as blameworthy on account of not prioritizing his children in his evaluative stance. However, the assessment of the unwillingly addicted agent is less straightforward, with the assessment going one of two ways. First, the agent might be excused for the wrongdoing on account of his unwillingness. When he uses his preferred substance, he is acting against his practical judgments that speak strongly against substance use. He hates experiencing or acting on his addictive desires. If this hatred constitutes full unwillingness, he might not be blameworthy on the non-structural QW account. Second, the unwillingly addicted agent might not be considered *truly* fully unwilling because he is not pursuing every possible avenue to try to avoid acting on his addictive desires or risking the well-

being of his children. As such, he might deserve some mitigation to match the degree of his unwillingness.[116]

Note that on both accounts, non-structural QW views are more lenient in their assessments of the unwillingly addicted agent than capacitarian RR views. If indeed the right categorization according to intuitive reactions is to consider both agents fully blameworthy, then non-structural QW do not get the right extension in regard to the cases.

3. Testing Views of Moral Responsibility

In this chapter, I have offered many reasons to reject Sripada's claim that control-based views of moral responsibility get the wrong extension in how they assess the hypothetical cases of the willingly and unwillingly addicted agents. If my argument is compelling, then these cases are not the challenge to control-based views as they are thought to be. Capacitarian RR can sort the cases in intuitively plausible ways, and track and explain such assessments.

But I have also shown that when we stipulate the loss of abilities in ways that are not actually experienced, either in general or in addiction, the cases invite implicit assumptions about abilities in intuitive assessments that make it unclear what exactly intuitive reactions are tracking as important for assessments of responsible moral agency. As such, the oddly stipulated cases do not serve as ideal tests for comparing the importance of evaluations in contrast to control in assessments of moral responsibility.

In the next chapter, I turn to consider cases of addiction that more plausibly align with how control is possessed or lost in the condition in an effort to see how QW and capacitarian RR

---

[116] I consider this sort of case in chapter 4 where I discuss the importance of effort in evaluative identification.

views might explain the relevance of apparently important features for considerations of

blameworthiness and mitigation.

Chapter 4: Explaining the Importance of Effort: Structural Quality of Will Views and Capacitarian Reasons Responsiveness

1. Introduction

As discussed in the last two chapters, Frankfurt's cases of the two addicted agents tend to be highly stipulated in philosophical discussions. Yet, messier versions of such cases plausibly exist, and even these messier cases can serve as good "tests" for theories of morality responsibility. The preferred theory is the one that can best track and explain the relevance of apparently important features in the cases for considerations of responsible moral agency and blameworthiness, but can also explain why features that seem important are not central to such assessments.[117] Consider again the cases of Mark and Richards described in the closing of chapter 2. Both agents experience considerable difficulty with resisting their addictive desires for substances. Yet, they take different evaluative stances toward their addictions and substance use. If both agents committed a similar addiction-explained wrongdoing, which of their psychological features seem most important in our assessments of blameworthiness and why?

Structural QW views contend that in assessments of responsibility what matters most is the structural fit between an agent's evaluative orientation and her motivating attitudes that successfully drive her action.[118] Depending on the view, the evaluative orientation might be understood as a higher-order desire, an evaluative judgment about the goodness of one's desires, or even a higher-order intention.[119] Agents are responsible for actions that result from the right structural fit, i.e. an agent's evaluative orientation coheres with what she is successfully

---

[117] As explained in the introduction, in this chapter I will assume that the agents are not culpable for acquiring their addictions. I take up this issue in more detail in chapter 6 of this dissertation.

[118] Fischer and Ravizza (1998, 184) refer to these theories as "mesh" views of moral responsibility. This sort of view is also referred to as a "structural" or "hierarchical" theory of responsibility (Talbert, 2016, 85) and are sometimes classified among the broader category of "identificationist" views (Jaworska, 2016).

[119] These views are defended by Frankfurt (1971), Watson (1975), and Bratman (2004), respectively.

motivated to do. Structural QW views consider this fit to be both a necessary and sufficient condition for agents to be assessed as blameworthy for wrongful actions.

Structural QW views are thought to give a compelling common-sense explanation for the blameworthiness of willingly and unwillingly addicted agents. In the popular imagination, addicted agents who do not want to be addicted but continue to use despite their psychological protests are pictured far more sympathetically than those who experience no such conflict. In a nutshell, Richards' higher-order evaluative attitude "fits" or "identifies" with his addictive motivation that drives his behavior, and so he is responsible for what he does. In contrast, the dissonance that Mark experiences between what he wants to want and the motivation that drives his behavior explains why he deserves at least a partial excuse in our assessments. So, even if two addicted agents like Richards and Mark are equally motivated to act on their desires to use drugs and commit the same addiction-related wrongdoing as a result of this motivation, their differing identifications with their effect motivation distinguishes them in terms of responsibility. Mark's internal psychological conflict makes him less blameworthy than willing Richards.

However, in this chapter I will challenge the idea that evaluative orientation viewed in terms of psychological identification with or against a motivating desire should be central in assessments of blameworthiness. As I will contend here, while structural QW views depend on the feature of effort to reveal the presence and strength of an agent's identification for explanatory robustness, they encounter a significant problem with explaining the relationship between effort and identification. When we consider how the two can come apart, we see that making an agent's evaluative stance in respect to their motivation a central consideration in assessments of responsibility has considerable explanatory limitations in distinguishing relevantly different agents and in explaining the importance of these distinctions for assessments.

I will argue that a capacitarian RR framework of moral responsibility is better equipped to both track and explain the importance of relevant differences between agents who are supposedly unwilling and between unwilling agents and akratic agents.[120]

To defend these claims, I will first present three structural QW views and explain how the case of the unwillingly and willingly addicted agents might be thought to support them. Second, I suggest that an appeal to the presence or absence of effort is essential for structural QW views to distinguish certain cases of unwillingness from seemingly superficial unwillingness. However, I argue that making effort an essential feature of identification means that structural QW views cannot explain the relevance of attitudes like resignation in assessments of responsible agency. Further, I argue that making effort only a signal of identification will not give structural QW views the explanatory power they need to distinguish cases of unwillingness from cases of mere akrasia. Finally, I will argue that capacitarian RR views are better equipped to track and explain the relevance of the important differences between various cases of unwillingness and between unwillingness and akrasia. Further, I will also show how capacitarian RR views can explain why resignation in the face of strong and persistent desires can matter for assessments of blameworthiness. I will close by suggesting that the concerns I raise here might apply to non-structural QW views as well.

2. Structural Quality of Will Views and the Cases of Addiction

In this section, I will consider three structural QW theories and explain how the cases of the willingly and unwillingly addicted agents have traditionally be thought to support them. In

---

[120] As noted in chapter 1, I assume in all of the cases in this chapter that the addicted agent is not culpable for their addiction. I take up this topic in more detail in chapter 6.

light of what I say in previous chapters, I will also briefly clarify the nature of the wrongdoing

done by the agents in question and the nature of addiction I am assuming in the cases.

2.1 The Case for Structural Quality of Will Theories

In Frankfurt's seminal essay "Freedom of the Will and the Concept of a Person,"

addiction is used to show how one can act freely and responsibly despite acting because of

irresistible, compulsive desires. As Frankfurt argues, no matter the source of motivation for a

behavior, whether an action is free and responsible rests solely on the details of an agent's

internal psychological structure. Addiction's assumed irresistible nature makes it illustratively

apt to show relevant psychological variations in the relationship between higher-order attitudes –

what an agent wants to want – and his effective motivating desires – what the agent is actually

motivated to do in action. Even if an addicted agent is irresistibly compelled to pursue and use

an addictive substance, he can identify with or withhold his identification from addictive

motivation to use drugs. As noted in the last chapter, in the original iteration of the cases of the

willingly and unwillingly addicted agents, Frankfurt stipulates that the addictive desires are

utterly irresistible. No matter what the hypothetical agents do to resist the desires, they are

forced to substance use. As I argued in the last two chapters, this is a misguided understanding

of how loss of control occurs in addiction and stipulating loss of control in this way introduces

problematic concerns for using the cases in theorizing about the role of control in responsible

moral agency.[121] Here, I will assume that while addictive motivation is not literally irresistible,

the addicted agents in the cases do experience strong, persistent addictive desires to pursue and

use substances that make it significantly difficult for them to consistently resist substance use

---

[121] See Pickard (2015) and Mele (2004) for additional concerns about stipulating irresistible desires as a key feature of addiction.

through conquering or circumventing. This stipulation makes room for their identification to be at odds with their addicted motivation, while also avoiding the worries about the cases that I raised in chapter 3.

Frankfurt contends that an agent acts freely and responsibly if and only if he identifies with his effective motivating desire, regardless of whether he is able to resist acting on his first-order desire for substance use. We can state his notion of free and responsible action as this:

> Free Action (F): A person acts freely iff and to the extent that the action results
>
> from a motivating desire that is endorsed by her a higher-order desire.

Key here is that the identification via a higher-order desire explains – or fails to explain – why the agent acts as he does.[122] The addictive desires are sufficiently difficult to resist, but the higher-order desire plays the primary role in explaining why each agent pursues and uses an addictive substance.[123]

On Frankfurt's structural QW view, the agent's evaluative orientation is seen as a *desiderative* attitude, and the addicted agents are described with this sort of psychological architecture in mind. The cases of the addicted agents, however, might also be thought to illustrate and support a cognitively-based structural QW theory as well, such as that articulated by Watson (1975).[124] Watson argues that the act of identification in the form of *evaluative*

---

[122] When identifying what explains the willingly addicted agent's action, we would say that he acts because of his identification with the effective desire to take the drug as well as because of the compelling nature of this desire. Sripada explains one way to conceive of this motivational over-determination, which I summarized in chapter 3. As Frankfurt (1971) offers this explanation: "His will is outside his control, but, by his second-order desire that his desire for the drug should be effective, he has made this will his own. Given that it is therefore *not only because of his addiction that his desire for the drug is effective*, he may be morally responsible for taking the drug" (20, emphasis mine).

[123] Sripada (2017) gives an extended explanation as to how the agent's evaluative stance plays the primary explanatory role in the action of substance use.

[124] It should be noted that Watson moves away from this structural account in his later work.

judgments more accurately represent an agent's fundamental moral concern. Evaluative

judgments are comprised by beliefs and values held in a "cool and non-self-deceptive moment"

and concern the practical ends an agent thinks make for a good and fulfilling life. The two can

work together, i.e. an agent can favorably evaluate what she desires or can come to desire what

she favorably evaluates. But the two can also come apart, such as when one is effectively

motivated to do something she desires but does not evaluate favorably. Free and responsible

action depends on the structural fit between the two psychological states:

> Free Action (W): An agent acts freely iff and to the extent that the action results
> from a motivating desire that is endorsed by her all-things-considered evaluative
> judgment.

A third style of structural QW can be constructed from Michael Bratman's theory of

agency as developed by philosophers such as Richard Holton (2009). On such a view, the

evaluative orientation of an agent might be comprised by what Bratman calls "self-governing

policies," which are general intentions to give weight to certain considerations in practical

reasoning about what to do or think in normative contexts (32). Self-governing policies –

hereafter referred to simply as "policies" – tend to be stable over time and tend to resist revisions

in motivations or beliefs. Importantly for our discussion, these policies can operate

independently of an agent's motivating attitudes, which can be driven instead by other mental

elements like emotions and bodily appetites that conflict with one's general policies. On this

view, free and responsible action can be captured as this:

> Free Action (B-H): A person acts freely iff and to the extent that the action
> results from a motivating desire that aligns with her self-governing policy.

The various virtues of each of these views in comparison to each other has been debated extensively.[125] Since my concern is with the broader claim that the *structural fit* between an evaluative stance with effective motivating desires is the central feature in assessments of blameworthiness, I will not make a case for which approach to the notion of evaluative stance is best. Instead, I will speak inclusively of an agent's *identification* with her motivating desires, leaving open whether this identification or endorsement is best understood in terms of higher-order desires, evaluative judgements, or self-governing policies. The notion of identification here will be broadly understood as the agent's evaluative orientation in the form of a second-order attitude directed toward her first-order motivating attitudes that result in action.

2.2 Identification and Blameworthiness

The notion of identification and the distinction and possible conflict between what one is motivated to do and the sort of motivating desires one identifies with serve as key notions for the structural approach to free action. But these notions can also provide a foundation for theorizing about the conditions of blameworthiness. On such a view, assuming that an agent meets the minimum standards for agency by being able to identify with motivating desires, an agent's free action is an action for which they can be assessed as responsible as well.[126]

On a structural QW theory of responsibility, an agent is blameworthy for wrongdoing that is motivated by desires that she endorses with her evaluative stance. Her blameworthiness,

---

[125] For examples, see Watson (1975); Bratman (2004); Sripada (2016).
[126] It is worth emphasizing again that the structural approach to moral responsibility developed here is only inspired by Frankfurt (1971), early Watson (1975), Bratman (1997, 2004) and Holton (2009, 2011), and so do not reflect their exact commitments or later revisions to their own views. Frankfurt (1971) and Bratman (1997), though, suggest that their theories might apply to considerations of responsible agency as well.

though, is mitigated to the extent that she does not identify with the motivating desires. These

ideas might be stated as this:

> Blameworthiness: an agent is blameworthy for an action if and only if the action is an instance of wrongdoing and the agent identifies with the motivating desire that drives the action.

> Excuse: an agent's blameworthiness for an action is excused if and only if and to the extent that the action is an instance of wrongdoing and the agent does not identify with the motivating desires that drive the action.

On this account, the willingly addicted agent identifies with his motivating desires to use

an addictive substance and is therefore more blameworthy for an addiction-explained

wrongdoing than the unwillingly addicted agent who also acts on his motivating desires but who

does not identify with these desires. Consider again the cases of Richards and Mark. Both act

on addictive desires they have significant difficulty resisting. But while Richards identifies with

his desires for substance use, Mark does not. If both Richards and Mark do something morally

wrong as a direct result of acting on their addictive desires, the structural QW theory considers

Richards to be more blameworthy than Mark, assuming that Mark is blameworthy at all. For

instance, if both were to act negligently or recklessly by risking the well-being of a loved one

while they pursued and used a substance, the theory would consider Richards more

blameworthy.

3. Unwillingness and Effort

In this section, I will argue that while the unwillingly addicted agent's *effort* to resist his

errant desires does essential work in illustrating his unwillingness, structural QW views cannot

provide a good explanation for why effort matters in assessments of responsible agency and

blameworthiness. I will consider three possible explanations of effort's relationship with identification, but argue that all three explanations are unsatisfactory accounts of its role.

3.1 The Essential Role of Effort in Distinguishing Identification

The idea that the unwillingly addicted agent is worthy of excuse because of his psychological unwillingness has initial appeal in part because most people have felt a degree of alienation from a desire and its related behavior. It seems intuitive that I am less blameworthy for a wrongdoing driven by desires that I do not want to be *my* desires. Hating one's own addictive desires are taken to be a potent illustration of this feeling. In making the case for unwillingness as compelling as possible, Frankfurt emphasizes this alienation. The unwillingly addicted agent *hates* his addiction and experiences addictive desires as a violation of who he really is (12).

Yet, it is not mere *psychologically* disapproval that most clearly illustrates the degree of the agent's identification against his motivating desire or its resulting behavior. Rather, his effortful *behavioral* struggle is key to communicating the extreme degree of his unwillingness. This psychological and behavioral exertion of effort are essential for distinguishing him from the willingly addicted agent. To see this and to appreciate the essential role that effort plays in illustrating unwillingness, compare the description the original unwillingly addicted agent to an agent whose unwillingness is merely psychological:

> Aspirational Unwillingly Addicted Agent: This agent is compelled to use drugs but hates these motivating desires. She always regrets succumbing to her addictive desires, feeling deeply ashamed and vowing never to use again, though these vows are short lived. Yet, though she struggles internally against her desires to use drugs, she does nothing to try to avoid using drugs or to mitigate her cravings. She turns down offers for counseling, group meetings, medical assistance, or long-term rehabilitation, though these options are all available to her.

Both this agent and the original unwillingly addicted agent psychologically identify against their addictive desires. Yet, the aspirational unwillingly addicted agent's lack of behavioral effort seems to make her relevantly different from the original unwillingly addicted agent. The aspirational agent seems unwilling in terms of her evaluative identification but her lack of effort to *do* anything seems an important feature for assessing her case. This suggests that effort plays an important role in explaining differences in the strength of psychological identification. But can the theory provide a plausible explanation of the role that effort plays in identification?

One possible response on behalf of structural QW views is that behavioral effort reflects the degree of an agent's identification against a motivating desire either by constituting or by signaling the degree of her unwillingness. As such, the more effort an addicted agent exerts to resist acting on an errant desire, the more psychologically unwilling she is when she performs the action anyway. As this response might go, while both the original and the aspirational unwillingly addicted agents are worthy of some degree of excuse on account of their unwillingness, because the original unwillingly addicted agent is more unwilling, as constituted or signified by her effort, she deserves a greater degree of mitigation for blameworthiness in our assessments.

In fact, a response of this sort might be just what structural QW views need to respond to a related worry commonly raised against structural views that pertains to an overly permissive extension of excuse.[127] Specifically, this complaint concerns how structural QW theories handle akratic agents. The standardly conceived akratic agent can be pictured as this:

---

[127] Talbert (2008, 6), Fischer (2012, 129-131), Jaworska (2017, 19), and Brink (2021) mention a worry of this sort.

Standard Akratic Agent: This agent acts on a motivating desire to do an action X, but does not fully identify with this motivating desire.

As the extension of excuse worry goes, if the unwillingly addicted agent is worthy of some degree of excuse on account of her lack of identification, then akratic agents are equally worthy as well. Yet, this result strikes many as a promiscuousness mitigation of blameworthiness because, as they are typically conceived, akratic agents are able to act in accordance with their evaluative judgments without much difficulty.[128] Assuming that action X is wrong, just because one does not fully identify with their motivation to do action X does not mean that they should be excused for the doing it. For instance, I might fully recognize that I am being a bad parent when I leave my children in a cold car to use intravenous cocaine and feel reluctant and guilty to use the cocaine or leave my children, but this lack of identification alone with my effective motivating desires does not at all seem sufficient for deservingness of some degree of excuse for the action. Structural QW views need a way to distinguish when a lack of psychological identification with a motivating desire can really be excusing.

Appealing to the presence or absence of effort might be just the theoretical tool that helps the views make this distinction. If an agent's expended effort to resist a motivating desire signals or constitutes her degree of psychological unwillingness, then structural QW theorists have a straightforward answer to the extension of excuse worry. Effort reveals the strength of identification and how unwilling or willing an acting agent really is. The original unwillingly addicted agent is deserving of a significant degree of excuse because he strongly identifies against motivating addictive desires, as revealed by both his psychological stance and his effort to act in light of this stance. In contrast, akratic agents are usually conceived as not trying

---

[128] Mele (2002, 2004).

113

particularly hard to act against their errant desires. If effort reveals the degree of the strength of identification, then this explains why we are not inclined to think such an agent is worthy of excuse to any substantial degree. Likewise, if the aspirational addicted agent's middling effort reveals only a half-hearted identification against her errant desires for drugs, then she too is worthy of a lesser degree of excuse than the original unwillingly addicted agent.

3.2 Effort as a Constitutive Part of Identification

This structural QW response looks initially promising. After all, effort seems to be an important component in measuring the strength of one's identification. If someone deeply identifies against a desire, should they not be trying with all of their might to fight this desire with every effort? Yet, I think this appeal to effort as an essential tool for determining strength of identification is less compelling when we look more carefully at the possible nature of the relationship between effort and identification. One option is to say that exerted effort is a *constitutive part* of identification such that the two are positively correlated. As such, the greater the degree of effort expended against an errant desire, the greater the degree of identification against it.

However, there are good reasons to think that effort is not a constitutive element of identification. Consider cases in which the two seem to come apart such that they are negatively correlated, such as cases in which a *lack of effort* belies an agent's evaluative stance toward their own motivating desires. Take Flanagan (2017) or Jeanette Kennett's (2013) case of the "resigned addict." As Flanagan describes this addicted agent, she has genuinely tried to stop using, accepted all offers of help, etc., but after repeatedly failing has finally given up actively

resisting the addictive motivating desires that she feels and believes she is unable to overcome.

We might picture her like this:

> Resigned Addicted Agent: This agent is highly motivated to use drugs but identifies against these desires. She has tried everything she could think of to resist using, but after years of repeated failure has finally given up trying. She still always deeply regrets using addictive substances and hates being addicted. Every time she uses, she feels deeply ashamed and is decidedly against her drug related desires and behaviors. Yet, she has become motivationally depleted when the cravings surge and no longer takes action to resist the desires. She hates who she is and what she does but feels unable to even try to try.

The resigned addicted agent has a complicated internal life but represents one plausible reaction to a continued experience of defeat. Her lack of self-efficacy, the attitude that one is able to exercise their agency to effectively meet their goals, results in a sense of impotence that inhibits her resistance to errant motivating desires. To illustrate this case, consider Flanagan's description of his own experience as a resigned addicted agent:

> If I was awake, I was drinking (and taking my pills). Sleeping involved becoming unconscious from drinking, and waking was coming to because I needed to drink, akin to the way normal sleep can be interrupted by needing to pee–not at all like waking up to meet a new day. I would come to around 6:15a.m., swearing that yesterday was the very last time. This was convenient because I was always out of booze by morning. So I would self-announce "I will not drink today, perhaps never again, but certainly not today" … But by 6:56 – every time, failsafe, I'd be in my car, arriving at the BP station–which meant beer to me … By the last months, I did not care, not at all – I even sought the sad consolation that the sweet kind souls who sold me my morning brew provided, knowing as they did how pathetic, how wretched, how lost I was. I guzzled on beer in the car. Car cranking, BP, a beer can's gaseous earnestness – like Pavlov's dogs, when these co-occur, Owen is off, juiced. And then gone … It, the second beer, was usually finished by the time I pulled back up to the house, the house on whose concrete porch I now spent most conscious, awake, time drinking, wanting to die. But afraid to die. When you're dead you can't use. I lived to use and to die. The desire to live was not winning the battle over death. The over-whelming need – the pathological, unstoppable – need to use, was. Living was just a necessary condition of using. The third and fourth morning beers were ingested more slowly than the first two. I knew that this was not a breakfast of champions. Not at all. But I was hardly a champion. Perhaps I had been once. But now I was a

pathetic loser. Ashamed and alone … My morning fix did temporarily suppress the self-loathing that was my constant morning companion. It mitigated briefly, maybe for 30 seconds or a minute, the inglorious shame of being a pathetic wretched being I was, that I had become … I was a wretched, worsening train wreck of a person – a whirling dervish contaminating, possibly ruining, the lives of my loved ones. What if they weren't strong enough to escape the harm? This was always the worst thought. And it was my constant torturous companion if I was awake and not using, which by the end was not often. [Flanagan, 2011, 277-8]

Cases like this show that effort can have a complicated relationship with identification. While someone like Flanagan is not exerting effort to resist his addictive desires, it would be odd to say that he does not identify strongly against them, as well as against the addiction-explained wrongdoing that follows from acting on them. Something seems amiss with the resigned addicted agent *not* exerting effort, which suggests that this agent's responsible agency is undermined in some way relevant to assessments of blameworthiness. How might structural QW views explain this, though, if effort is a necessary condition for identification?

A possible response from structural QW views is to double down on the appeal to a constitutive relationship between effort and identification and say that a resigned addicted agent is not actually unwilling, or at least not *that* unwilling, because their failure to exert effort just shows that they are not wholly identified against substance use. If an agent truly and absolutely identified against their addictive desires, they would put in actual behavioral effort rather than indulging in shame and self-hatred. A structural defense of this claim is perhaps most plausible when identification is conceived in terms of resolutions or policies about future actions. On such a view, exerted effort seems to go hand-in-hand with identification. As Holton writes, "sticking with a resolution requires us to *do* something: we are not passive spectators" (2011, p177). Of course, it is plausible that an agent might evaluative something favorably and not be constantly

effortfully engaged in supporting it.[129] But in the case of identifying against a highly motivating errant desire, exerting time and energy to resist it seems a necessary aspect of identifying against it.

Yet, even advocates of policy-based structural theories allow that an agent might be strongly motivated to hold a policy against acting on certain desires, but then understandably lack motivation for acting in light of this policy. To put it another way, the motivation involved with psychologically disavowing an errant desire and the motivation required to implement the actions of this disavowal can come apart. One plausible explanation for this disconnect is that while personal policies can be motivating, an agent needs to employ additional mental effort as well to stick to the policy, especially when that agent is already strongly motivated to act *against* that policy. Mustering and maintaining this additional mental effort, though, is vulnerable to all sorts of factors like stress, fatigue, and a sense that one's own efforts will not be successful.[130] Empirical work has shown the role of this last factor in particular can play a substantial role in an extended deflation of motivation to act on one's evaluative stance and against one's errant motivating desires. Kennett (2013), for instance, argues that the greatest "internal threat" to overcoming addiction is not so much the persistence and resilience of addictive wanting, though of course this plays a role, but the motivationally deflating attitudes fomented by these failures in self-regulation that undercut the motivation to act against addictive wanting, as well as

---

[129] Arpaly and Schroeder (2013, 2014) make this case in their explanation of intrinsic desires in their own non-structural QW view.

[130] Holton (2009, chap. 6) nicely offers and defends a picture of how intentions and mental effort or will power can work together or come apart. Holton analogizes these sort of cases to someone doing a physically exhausting task and who cannot continue the task any long due to muscle fatigue. This motivational picture offers an account that is compatible with the account I consider here.

undercutting the very attempts to deliberate about practical options or set intentions for actions.[131]

All this is to say that it is plausible to think that one can be deeply identified against an errant effective desire and yet not exert effort to act on this identification, lending support against the claim that effort is constitutive component of identification.

3.3 Identification and the Importance of Past Effort

Still, effort seems to play a vital role in how structural QW views might handle the extension of excuse worry. How else might the views try to explain the relationship between effort and identification in cases where the two come apart? Another possible response is to say that exerted effort is a constitutive part of identification but current identification can only be appreciated with an agent's history of effort in mind. In the case of the resigned addicted agent, her *past* effort tells us why her current lack of effort should not be taken as indifference or even willingness toward her addictive desires. As such an explanation might go, the resigned addicted agent is significantly unwilling because her *past* is explanatorily central to understanding why she fails to exert effort now. What distinguishes her from the aspirational unwillingly addicted agent or the willingly addicted agent is that she *had* exerted effort, and this historical effort continues serving as an aspect of her identification.

There is an appeal to this approach. What an agent evaluated favorably in the past seems important for revealing both the content and strength of their current stance. For instance, it is

---

[131] Cheshire Calhoun (2008, 194) offers a similar argument for how these beliefs and attitudes regarding one's own agential abilities can undercut motivation to do what one values or to resist doing what one does not value. See also McConnell (2016). Recent empirical work on addiction and shame has shed a light on how this one particular attitude can undermine motivation to resist use. See Matthews, Dwyer, & Snoek (2017); Snoek, McGeer, Brandenburg, & Kennett (2021).

more plausible that a person prioritizes their children in their evaluations when they have displayed obvious effort to make their children a priority in the past. However, when an agent no longer seems so strongly identified in a particular way it is not at all clear that her current identification is still constituted by psychological states or behaviors that she no longer experiences or acts on. An agent's lack of effort following a period of effortful exertion seems important, but appealing to effort as a constitutive part of identification cannot explain its importance. Rather, a structural QW theory would need to appeal to some other explanation.

One plausible explanation is that a current lack of effort when one has exerted effort in the past reveals something important about a change in a person's agential abilities. When someone stops trying to do something that has greatly mattered to them in past, it can tell us that their identification has shifted to value it less. But it can also tell us that their ability to exert this effort has been undermined in some important way. For instance, beliefs and attitudes fomented by past failures may be contributing to motivational depletion that is difficult to overcome.

While I will propose an explanation of this latter sort in section 3, I do not think that structural QW views can appeal to such an explanation. This is because this sort of reply smuggles in an appeal to diminished control over motivating desires as a fundamental aspect of the explanation as to why someone like the resigned addicted agent is worthy of a greater degree of excuse than someone like the aspirational unwilling agent. But, if the constitutive part story does not adequately explain the relationship between effort and identification, even when past history is considered, how else might a structural QW view explain this relationship? In the next section, I will consider one more possible explanation before offering my own account in section 3.

3.4 Effort as a Signal of Identification

It is still open for structural QW views to say that the presence of effort *signals* an agent's identification. As such an explanation would go, effort, while not a necessary component of identification, still serves to convey information about an agent's evaluative stance. In the case of the unwillingly addicted agent, his effort expended to fight acting on his addictive desire signals the degree of his unwillingness. In contrast, the willingly addicted agent's expended effort to encourage his addictive desire and to pursue and use substances signals his identification in favor of substance use. Importantly, on this explanation, a *lack* of effort does not necessarily signal absent or counter identification. A resigned unwillingly addicted agent, for instance, may still be considered significantly unwilling in light of her effective addictive motivation, but this identification would be explained by appeal to other psychological elements, such as shame or regret.

If this story is right, then this would explain why effort seems so important while comparing the identifications of the unwillingly and willingly addicted agent. However, this explanation would also mean that the presence or lack of effort is not a useful tool for distinguishing the original unwillingly addicted agent from someone like the aspirational unwillingly addicted agent. Unfortunately, this is a significant problem for structural QW views because, if effort is not a reliable tool for distinguishing the strength of identification for agents like these two, then it is likewise an unhelpful tool for distinguishing an unwilling agent from a merely akratic agent. This would mean that an appeal to the importance of effort cannot help respond to the promiscuous extension of excuse concern.

Structural QW views seem to rely on the feature of effort to be explanatorily robust, yet this reliance also results in explanatory deficiencies with regard to explaining identification in agents such as the resigned addicted agent. These explanatory limitations motivate a second look at the cases of the unwillingly addicted agents, the akratic agent, and the willingly addicted agents to see what else can do the explanatory work that identification cannot.

4. Capacitarian RR Views, Volitional Control, and the Importance of Effort

In this section, I will contend that capacitarian RR views can better account for the seemingly important differences between the various unwillingly addicted agents than structural QW views.

4.1 Explaining Distinctions between the Unwillingly Addicted Agents

As I argued above, structural QW views rely on the the feature of effort to distinguish the original unwillingly addicted agent from the aspirational unwillingly addicted agent. Yet, I contended that these views cannot offer a good account of why effort matters for identification, and this explanation is key for distinguishing relevant differences in identification relevant to assessments. Capacitarian RR views take a very different approach to the above considered cases, specifically taking into account the abilities of each agent to resist their motivating errant desires that they recognize that they have sufficient moral reason to resist. In such cases where an agent is able to track the sufficient considerations that weigh against their continued substance use, capacitarian RR views ask whether they also have a sufficient volitional capacity to act in light of these considerations by resisting their addictive motivation.

Again, as explained previously, volitional capacity is comprised of two primary sub-abilities: the ability to conquer an errant desire and an ability to circumvent an errant desire.

This important distinction, and the fact that many people can lack the first ability to a significant degree with respect to a specific errant desires while retaining the second, can explain why we might be inclined to excuse the blameworthiness of the original unwillingly addicted agent but not the aspirational unwillingly addicted agent. Both seem to lack a sufficient ability to *conquer* their addictive motivation using sheer mental effort alone. Both set intentions to stop use and exert mental energy that might directly counter the errant desires. Despite their persistence both fail to succeed in this resistance.

Yet, it is unclear whether both agents also experience a similar loss in their ability to *circumvent* their addictive desires. It is plausible to think that the original unwillingly addicted agent's ability to circumvent his addictive desires is significantly compromised. Continuing to succumb to substance use despite exerting considerable effort over an extended period to resist it is a good sign that something is amiss volitionally. In contrast, it is unclear whether the aspirational unwillingly addicted agent likewise experiences such a compromised ability for circumvention. Since one can possess an ability without it being exercised, it is unknown whether the aspirational agent could successfully resist her addictive desires by circumventing them through different tactics much like what the original unwillingly addicted agent tries but fails to do.

This capacitarian RR explanatory framework, then, can neatly explain the thought that these two unwillingly addicted agents might be importantly different with respect to responsible agency based upon a possible difference in volitional competence. Unexercised aspiration alone does mean one's volitional abilities are compromised to any extent, let alone to a significant extent. More needs to be said about the cases to discern whether the aspirational addicted agent's ability for circumvention is sufficiently impaired or if she is simply not exercising and

ability that she has. But also, this framework can explain why a standard akratic is not deserving of a significant degree of excuse, if they are excuse worthy at all. Presumably, the standard akratic can, without much difficulty, act on the considerations that he evaluates as practically sufficient. This is what sets him apart from agents like the original unwillingly addicted agent. Both may experience a conflict between their evaluative stance and effective motivating desires, but only the latter's abilities to act in light of the considerations that speak against a wrongful action are sufficiently undermined to make him deserving of any degree of excuse.

4.2 Effort and the Resigned Addicted Agent

The capacitarian RR framework provides a compelling explanation for the importance of differences between various types of unwilling agents, as well as between unwilling agents and akratic agents, that bear on assessments of blameworthiness. But importantly, it can provide this explanation without requiring that an agent be actively exerting effort to resist their errant desires. This positions the view for being able to explain assessments for cases where effort is suspiciously absent.

Consider again the case of the resigned addicted agent again. It is intuitive that something is amiss with such an agent's responsible agency. But how might capacitarian RR views explain what is amiss in a way that might bear on assessments of responsible agency? Here is one possible plausible explanation. As I suggested above, a cessation of effort after an extended period of unsuccessful effortful exertions can reflect that normative capacities have been undermined in some important way. The explanation that I suggested above is that the beliefs and attitudes fomented by significant past failures can contribute to motivational depletion. The capacitarian RR framework can explain the importance of this depletion by

reference to volitional abilities. The strong and persistent desires that characterize severe

addiction can take extensive mental and physical effort to resist, which can lead to mental

depletion. But this depletion can be further exacerbated by attitudes and beliefs that one is

unable to effectively resist errant desires. Once these attitudes and beliefs take hold, they

undermine the needed motivation for practical strategizing.

Whether a resigned addicted agent is deserving of some degree of excuse in light of this

mental depletion depends on whether the beliefs and attitudes that fuel mental depletion can be

conquered or circumvented without significant difficulty. For some, it might not take much

effort to pull out of this mental motivation depriving slump. Though not attempting to model

resignation per se, Holton (2009, 2016) offers an explanation for why some addicted individuals

might succumb to addictive temptation even though doing so conflicts with their evaluative

judgments. As Holton explains, one response to the conflict between evaluations and especially

hard-to-resist strong and resilient errant desires is to embrace the belief that one is unable to

resist them. Holding such a belief can ease the psychological discomfort experienced from the

on-going psychological struggle, while also offering the addicted agent an explanation and

justification for their failures in self-regulation. Embracing this belief when faced with a

difficult task that one expects to fail might be described as a form of willful denial, especially in

the face of ample accessible evidence that most addicted individuals can and do eventually

successfully stop or significantly reduce substance use. Yes, resisting strong and persistent

addictive desires day after day can be incredibly hard and there are bound to setbacks in the form

of lapses and relapses. But unless an addicted agent is in a context in which they do not have

access to basic circumvention resources, say in prison, or they suffer from a more profound

cognitive disorder that makes self-regulation especially difficult in general, it would seem that they can reasonably be expected to exert effort.

For others, though, the motivation-depriving beliefs and attitudes that characterize resignation may be exceedingly difficult to avoid or overcome. When these beliefs and attitudes are undergirded by shame, for instance, they can be especially hard to subdue. Shame might motivate recovery since it drives "the addict to want to get a grip" (Flanagan, 2013b, 149). But sometimes shame can also be experienced as disabling and "self-stigmatizing," fostering self-hatred and hopelessness (Matthews, Dwyer, and Snoek, 2017). This latter kind of shame undercuts motivation to confront addictive desires and instead motivates an escape into substance use, prompted by the idea that this is one's only viable option given their profound failures. This in turn cultivates more shame, resulting in a self-perpetuating cycle. Considering that even addiction therapists can struggle to determine the best methods for treatment in such cases, it can be unreasonable to expect that such an agent can circumvent the attitudes that undermine the motivation to try to resist errant addictive desires without the help and perspective of an experienced professional counselor, to which they may not have access.[132]

These two models show how resignation might bear on the volitional ability to resist substance use. Notably, capacitarian RR views can helpfully distinguish they types of cases of resignation even without appealing to effort. What role then does effort play in responsible agency according to these views? As shown in the case of the original unwillingly addicted agent, the presence of effort can be helpful for empirically identifying who has a substantial volitional capacity. But effort is not a perfect signal. As demonstrated in the case of the

---

[132] Snoke et al. (2021).

resigned addicted agent, lack of effort does not necessarily mean a lack of volitional capacity. Likewise, the presence of effort does not mean an agent necessarily has the requisite cognitive control. However, for capacitarian RR views, it is theoretically okay that effort is such an imperfect signal because they do not rely on effort as an explanatory lynch pin for making basic distinctions between cases that seem importantly different in order to explain responsible agency or blameworthiness.

4.3 Conclusion and Implications for Non-Structural QW views

I have argued here that structural QW view cannot track and explain the importance of differences between various types of unwillingly addicted or akratic agents. The reason for this is that such a distinction depends on the existence of effort to reflect the presence and degree of identification, yet effort does not seem to either constitute or reliably signal the presence or strength of an agent's identification with their effective motivating desires. I have argued that capacitarian RR views can track and explain important distinctions between the various cases of addiction considered in this chapter by making loss of control an essential consideration in assessing responsible agency and blameworthiness.

While I have focused on *structural* QW views in this chapter, my argument has implications for the explanatory robustness of *non-structural* QW views as well. For non-structural QW views, the presence of effort to promote what an agent evaluates as fundamentally morally important is an important signal of the content and strength of their evaluative stance. As such, these views will have explanatory difficulty accounting for cases like the resigned addicted agent as well. However, as I will argue in the next chapter, these views face a bigger

explanatory challenge in simply accounting for why *any* addicted agent might deserve mitigation

to any extent for an addiction-explained wrongdoing on account of their addicted condition.

Chapter 5: Explaining Worthiness for Excuse: Conative Quality of Will Views and Capacitarian Reasons Responsiveness

1. Introduction

In its more severe form, addiction is associated with a puzzling mix of features that together suggest it might undermine responsible agency to some degree. Consider again the case of Carr from chapter 2. He clearly acts wrongly when he unjustifiably risks the well-being of his children to pursue and use cocaine. Yet, something seems amiss with aspects of his responsible agency. He values the well-being of his children, as perhaps evidenced by the effort he put into properly dressing them for the weather and refusing to leave them home alone. But, overall, his actions undermine this value. His addiction seems to influence how he reasons about his choices and their consequences to the point that, in the "hot" moment of experiencing cravings, he tries to justify acts that most parents would find abhorrent. He cannot seem to resist acting upon his strong desires to use drugs even as he clearly realizes that what he is doing is wrong. In the moment of action, he experiences shame at what he feels impelled to do, and immediate deep regret after doing it.

Together, these features suggest that his addicted condition might mitigate his blameworthiness for his wrongdoing at least to some degree. To see the pull of this intuition, compare Carr's case to one in which a non-addicted agent commits a relevantly similar wrongdoing, such as father leaving his infants alone in a car in freezing weather in the middle of the night so that he can hang out with friends over pizza and video games.[133] Something seems amiss with Carr's responsible agency due his severe addiction that suggests his case should be treated dissimilarly in our assessments from a non-addicted agent who acts similarly.

---

[133] Arpaly and Schroeder (2013, 2014) make a similar comparison.

If the intuition that addiction can at least partly undermine responsible agency is right, this suggests that the condition can mitigate blameworthiness for wrongdoing directly related to the condition.[134] But how might deservingness for this mitigation be best explained, if indeed it is deserved? Capacitarian Reasons Responsiveness theories would ask whether Carr has control with regard to the action in terms of being able to appreciate the reasons for prioritizing the well-being of his children above substance use and to act in light of these reasons. Alternatively, non-structural Quality of Will views would ask whether Carr's action reflects or is sufficiently motivated by his fundamental evaluative stance toward the rights or well-being of his children. *Conative* non-structural QW views, in particular, would ask whether his action reflects what he most fundamentally cares about or intrinsically desires morally.

Here I want to argue that addiction poses a particular explanatory challenge for these conative QW views that they cannot meet in regard to explaining how addiction can be a mitigating condition for blameworthiness for addiction-explained wrongdoing. I will contend that these views must make considerations of control the most fundamental aspect in assessments in order to successfully explain the possibility of mitigation. To make this argument, in section 1 I will survey the basic commitments of two conative QW views defended by Arpaly and Schroeder (2013, 2014) and Sripada (2016, 2017) relevant to considerations of blameworthiness and excuse, and identify the explanatory challenge presented by addiction. In section 2, I will present and respond to a series of arguments in defense of the claim that conative QW accounts can explain how addiction can be a mitigating condition in assessments of blameworthiness. The most promising of these arguments appeals to the idea that addiction can hijack motivation,

---

[134] As explained in the introduction, in this chapter I will assume that the agents are not culpable for acquiring their addictions. I take up this issue in more detail in chapter 6 of this dissertation.

causing a mismatch between the strength of intrinsic desires or fundamental cares for substance use and the motivational strength to act on those desires. However, I will contend that the appeal to hijacking is fundamentally an appeal to the importance of self-control in assessments of responsibility. In section 3, I will push further on the conative QW explanation of mitigation by presenting the case of the diachronically stable willingly addicted agent. This case illustrates how addiction can impact the capacity for forming evaluations, diachronically blinding some severely addicted agents to values that might compete with drug use. I will argue that conative QW views lack the resources to explain what exactly is concerning about this evaluative blinding. I will conclude by suggesting that capacitarian RR views of moral responsibility can readily track and explain why evaluative blinding is relevant to assessments of blameworthiness and should be preferred for this reason.

1.1 Conative Quality of Will Views

Calling their view *Spare Conativism*, Arpaly and Schroeder (2013, 2014) argue that assessing someone as blameworthy for a wrongdoing is a matter of assessing the quality of that person's will, which they understand in terms of *intrinsic desires* – pre-deliberative desires for something being the case or not being the case for its own sake. Though agents are not always consciously aware of their intrinsic desires, these desires directly or indirectly underly actions, thoughts, emotions, and practical judgments, as well as organize what people care about as revealed in their priorities. An agent is blameworthy for a wrongful or bad action when it reflects that they intrinsically desire the morally wrong or bad *or* that they lack an intrinsic desire, or the right degree of intrinsic desire, for the morally right or good. Intrinsically desiring or caring most about the wrong or bad is called *ill will* and neglecting to care about the right or

the good to the degree it should be cared about is called *moral indifference.* Arpaly and

Schroeder capture their view succinctly as this:

> **Blameworthiness (Spare Conativism):** An agent is blameworthy for a wrong or
> bad action *A* iff and to the extent that *A* is done out of ill will or moral
> indifference.

For Arpaly and Schroeder, desiring the wrong or bad is generally more blameworthy than failing

to desire the right or good when one acts. To illustrate, someone who intrinsically desires to hurt

their child is more blameworthy than someone who cares about their child but prioritizes

something of less moral importance above their child's well-being.

Sripada's (2016) *Self-Expression* view takes a similar conative approach to responsible

agency and blameworthiness, identifying the privileged attitude in question as an agent's *cares*

that are fundamental to her "practical self". Fundamental cares, as I will call them, are what an

agent cares about most for their own sake. Sripada's view might be expressed as this:

> **Blameworthiness (Self-Expression):** An agent is blameworthy for a wrong or
> bad action *A* iff and to the extent that *A* is motivated by their fundamental cares.

Sripada argues that fundamental cares are identified functionally by their "syndrome of

dispositional effects." They provide intrinsic motivation; are experienced as a loss when

changed or altered; are connected to evaluative judgements such that the object of care and

actions that promote the object of care are evaluated favorably; and are causally related to a "rich

and distinctive profile" of emotional responses dependent on the object of care (1209-1211).

Spare Conativism and Self-Expression views differ in some important ways. Perhaps

most important for this discussion is that while Arpaly and Schroeder take intrinsic desires to be

a natural psychological kind with a distinctive physical etiology in the brain, Sripada identifies

fundamental cares functionally by their dispositional effects on other motivational, evaluative,

and affective psychological states. Still, both accounts take a similar approach to the role and function of intrinsic desires/fundamental cares in agential psychology and behavior. They both hold that the evaluative attitude at stake in assessments of blameworthiness is what an agent intrinsically cares most deeply and this attitude functions by shaping an agent's priorities, evaluations, motivations, emotional life, and actions.

Given the significant overlap between the views, I will treat them together, referring to the fundamental evaluative attitude in question using terms like "intrinsic desires," "fundamental cares," or "intrinsic cares." However, as I consider individual arguments concerning whether conative QW views can explain addiction as a possible mitigating condition of blameworthiness, I will note and respond to relevant unique aspects of each view.

1.2 The Challenge of Explaining Addiction as a Mitigating Condition

Addicted agents like Carr seem to fundamentally care for substance use more than anything else, including the well-being of their children. Using Arpaly and Schroeder's terminology, Carr does not manifest ill will toward his children in his action, nor is he morally indifferent to their well-being. However, he does lack *sufficient* concern for them and is blameworthy for not prioritizing their safety above his substance use. Yet, it is intuitive to many that Carr should be blamed to a lesser degree on account of his addicted condition than someone who acts similarly but does not suffer from such a condition. Yes, Carr does something terrible by leaving his infants alone and vulnerable to the severe cold. But on account of his severe addiction, something still seems importantly different between his case and a non-addicted agent who does the same. This initial intuitive reaction suggests that addiction might provide some

degree of mitigation for blameworthiness. But, given that Carr's action seems to accurately reflect his prioritization of fundamental cares, how might QW views account for this mitigation?

A straightforward way to explain the possibility of addiction as a mitigating condition is to say that it can reduce an agent's ability to control their addictive substance use, making them also less able in some circumstances to control a wrongful action that directly results from the pursuit and use of addictive substances. Capacitarian RR theories take this approach, offering two ways in which a condition like addiction might mitigate blameworthiness for addiction-explained wrongdoing. First, addiction can make it more difficult to recognize or appropriately weight the sufficient moral considerations that speak against substance use. This is plausible given the evidence that addiction involves motivational features that can significantly influence evaluations about the worthwhileness of substance use. Second, addiction can make it significantly more difficult for an agent to act on considerations against substance use, especially on repeated occasions.[135] This latter ability is often simply called "self-control" or "self-regulation."

Conative QW theorists, however, want to reject this approach to mitigation, refusing all appeals to control or its loss as central considerations in assessing either blameworthiness or worthiness of mitigation. Instead they contend that what matters for assessments is whether and the extent to which the wrongful action reflects an addicted individual's fundamental cares. QW theorists take it to be an explanatory advantage of the view that it does not appeal to considerations of control.[136]

---

[135] Sripada (2018).
[136] I challenge this assumption in chapter 3.

Yet, how might conative QW views account for mitigation in cases where the addicted individual seems to be doing exactly what they most care about at the moment of the wrongdoing? Here's the puzzle. On conative QW views, an agent who acts wrongly is blameworthy iff and to the extent that wrongdoing reflects their fundamental cares. This means that to be excused for a wrongdoing, the act must not reflect or be sufficiently motivated by that agent's fundamental cares. Yet, if only during cravings and consumption, severely addicted agents *do* fundamentally care about pursuing and using drugs more than anything else, including what is morally good or right. Addiction-explained wrongdoing that happens during these periods accurately reflects the agent's prioritization of cares. Thus, addicted individuals must always be assessed as fully blameworthy on conative QW accounts for wrongdoing committed during periods of addictive cravings or consumption. The primary tool for QW views to explain deservingness of mitigation is to show that a wrongful act does not sufficiently reflect an agent's fundamental evaluative stance. Since the act in question reflects the addicted agent's evaluative stance, to what else might QW views appeal beside considerations of reduced control over the action?

2. QW Arguments for Mitigation

In section 2, I consider a series of arguments that conative QW views might offer to explain how addiction can mitigate blameworthiness, drawing attention to the resources of different conative QW approaches where relevant. Their general strategy to explain mitigation is to show that addictive desires do not accurately reflect an addicted agent's fundamental cares, despite seeming like they do. As such, when an addicted agent does wrong on account of their addiction, the wrongdoing might not actually reflect the content or strength of that agent's moral evaluative stance. This general strategy depends upon the idea that addictive desires can be

distinguished from "typical" desires, an assumption I will argue is not convincing unless the main explanation for the distinction is that addictive desires are more likely to impair self-control undermine control to a greater degree than typical desires.

2.1 Functional Difference Argument

Sripada's conative QW account holds that fundamental cares should be defined functionally. Assuming this, to explain how addiction might be mitigating, one might simply claim that addictive desires are functionally different from fundamental cares. So, in cases where addiction is the primary explanation for a wrongdoing, the fact that the wrongdoing reflects a strong desire to use and pursue an additive substance does entail that it necessarily reflects the agent's fundamental cares. One plausible version of the argument could contend that addictive desires are more like appetitive urges that just do not have the full range of dispositional effects as intrinsic desires.[137] As such an argument would go, addictive desires are only strong, subjectively felt desires aimed at pleasure or relief. Such urges or desires are easily functionally distinguishable from the broader evaluative and emotional life of an addicted agent, which is the purview of fundamental cares.

However, as even some conative QW theorists note, there is good reason to think addictive desires can function the same as fundamental cares, especially in cases of severe addiction.[138] Especially in cases of severe addiction, there is good reason to think that the addictive desires "do" what fundamental cares do. Fundamental cares are what agents most deeply care about intrinsically. They arrange an agent's instrumental hierarchy of other cares.

---

[137] Smith (2004, 332) makes a comparison of this sort to explain why an agent might not be accountable for addictive desires.
[138] Arpaly and Schroeder (2013, 232)

They motivate actions and serve as the foundation for myriad of other instrumentally motivating attitudes. They are experienced as a loss when altered. They incline evaluative judgments that favor the object of care and incline practical judgments to promote actions that express the care. They support "an array of emotions and emotional responses directly responsive to the object of one's caring. Addictive desires can do all of these things. To see this, consider the nature of addictive desires and how addiction, especially severe cases, shapes the motivational, evaluative, and emotional life of addicted individuals.

In some contexts, addictive desires are compared to other appetitive desires, and thus it is tempting to see them as mere urges that are easily distinguishable from what an agent really cares deeply about. Strong, persistent, and resilient motivation to engage with or consume an addictive objective is a central symptom of the condition.[139] Most current theories of addiction agree that addictive motivation is characterized by subjectively felt aspects and unconsciously experienced psychological tendencies that strongly incline the addicted agent to pursue and use. Cravings are one explanatorily central feature of addictive motivation, commonly described as a combination of negatively valenced emotions and a consciously experienced drive to engage in a particular behavior to resolve a felt deprivation.[140]

Yet, addictive motivation, of which cravings are a central aspect, can be experienced as more than just merely appetitive urges. Rather, the desire to pursue and consume an addictive substance can become foundational in an addicted agent's life, shaping work life, relationships, hedonic experience, evaluations, and emotions. Friendships are developed primarily because

---

[139] Sinnott-Armstrong and Pickard (2013); Dill and Holton (2014, p29); Lewis (2011, 2017); Schroeder and Arpaly, (2013); Arpaly and Schroeder (2014); Berridge and Robinson (2011, 2016); Robinson and Berridge (1993, 2000, 2003); Wallace (1999); Watson (1999a).
[140] Arpaly and Schroder (2013, 230); Berridge and Robinson (2011); Robinson and Berridge (1993, 2000, 2003).

they guarantee a source for drugs or provide the company of other addicted drug users. Jobs are

sought or rejected based on how they might accommodate substance use. Relationships are

broken because they challenge substance use. Plus, addiction is thought to make competing

activities less pleasurable.[141] Eating, sexual activity, hobbies, work, or relationships seem less

interesting or worthwhile in comparison to substance use or activities directly related to use.

Some theorists hypothesize that one of the most difficult aspects of ceasing addictive drug use is

learning to find value again in activities that do not involve drug consumption.[142] Some

philosophers go so far as to describe the connection that addicted individuals can develop with

an addictive substance as "existential dependence."[143] Wonderly (2021), for instance, notes that

addiction can feature the same sort of attachment to an addictive object that love does to a loved

one. She describes this attachment as featuring a "recurrent, recalcitrant" desire to engage *with*

the addictive substance. The addictive object is non-substitutable and the engagement with the

addictive objective provides a sense of security, which she describes as "a kind of confidence in

one's well-being and agential competence" (233). While engagement with the addictive object

through consumption can provide a sense of comfort, purpose, and confidence, lack of

engagement can cause anxiety, hurt, and a sense of loss. As a consequence of how addictive

substances are desired in addiction, both in terms of the strength of motivation and in terms of

attachment, the pursuit and use of addictive substances is prioritized in an addicted individual's

hierarchy of values.

---

[141] Koob (2013, 2017); Koob & Le Moal (2005); and Heyman (2009).
[142] Heyman (2009); Snoek (2017); Pickard (2020).
[143] Watson (1999a); Pickard (2020); Wonderly (2021).

All this is to say that there is good reason to think that desires for addictive substances can be functionally the same as fundamental cares. This would mean that, especially in more severe cases of addiction such as Carr's, the pursuit and use of addictive drugs can reflect an addicted agent's fundamental cares. The attempt to distinguish addictive desires from fundamental cares by function will not be successful in cases of addiction most deserving of mitigation.

2.2 Hijacked Desires Argument

A more promising approach for conative QW views to explain mitigation is to admit that in cases of agents like Carr, addictive desires do reflect an addicted agent's fundamental care for substance use, but perhaps not as much as the motivational strength of the desires suggest. This is because addictive cravings nefariously heighten the motivational power of these cares. One version of this argument is offered by Arpaly and Schroeder (2013, 2014). They claim that though cravings function as intrinsic desires and have the same biological source in the brain's reward system as intrinsic desires, their motivational power is a product of "hijacking," and so their motivational power is not necessarily a true reflection of the strength of the addicted agent's fundamental cares. Here, after reviewing their argument, I will contend that the hijacking claim, if we accept it, is actually just concern about loss of control, specifically a concern about a compromised ability to self-regulate in light of one's evaluations.

As mentioned above, cravings are often thought of as a consciously felt state of deprivation that motivates actions to address the deprivation. Arpaly and Schroeder argue that cravings functionally act like intrinsic desires:"The person [experiencing cravings] feels highly motivated, feels good about the prospect of doing the thing in question and bad about the

prospect of not doing it, and would describe herself as very much wanting to do what she craves doing" (2013, 232). They also think that cravings have the same biological source in the brain's reward system. This is important because Arpaly and Schroeder contend that intrinsic desires are a natural psychological kind that are likewise produced in the brain's reward system – a group of neural structures implicated in processes like motivation, associative learning, memory, and emotions.[144]

However, despite their etiological and functional similarity, Arpaly and Schroeder argue that addictive desires are distinctive from "typical" intrinsic desires in the strength of their motivational power. This argument rests in part on a biological story that is worth explaining here. According to Arpaly and Schroeder, intrinsic desires form through a process of association that the brain makes between "cues" and "outcomes." Cues are anything perceived or thought that come before and reliable predict the presence of an outcome, an event or experience. When a person unconsciously or consciously makes a connection between a cue and an outcome and the outcome is perceive as "positive," i.e. should be repeated, the brain releases dopamine. Dopamine is neurotransmitter broadly implicated in processes related to learning, motivation, memory, and possibly pleasure.[145] Following a positive event, the release of dopamine functions to connect the occurrence of the cue and the outcomes memory. When the cue is encountered again, dopamine is released at the cue in anticipation of the outcome.

However, Arpaly and Schroeder emphasizes that once the connection between the cue and outcome is established, the brain stops producing a high level of dopamine and the

---

[144] See Arpaly and Schroeder (2013; 2014) and Schroeder (2010) for the details of their theory of addiction. See Berridge and Robinson (2011) and Holton and Berridge and (2013) for challenges to the "learning" approach to addiction as defended by Arpaly and Schroeder.
[145] Robinson and Berridge (1993, 2008); Berridge and Robinson (1998).

individual becomes unconsciously "jaded" to the connection. Namely, the specific outcome is simply expected to follow the cue and motivational features like eager anticipation fade. But if ever there is a "prediction error" between a cue and an outcome, where the actual outcome being more positive or negative than predicted, the amount of dopamine released will be modulated accordingly. When the actual outcome is more positive than predicted, more dopamine is released. When the actual outcome is more negative than predicted, the amount of dopamine will drop. These reward system processes happen unconsciously, but they work to help shape behavioral learning through influencing feelings, desires, memories, and pleasure.

This biological story is central to Arpaly and Schroeder's argument because they contend that it explains how addiction hijacks of reward process. Addictive drugs cause a release of dopamine that does not decrease over time no matter how the actual outcome of drug use is perceived or evaluated unconsciously or consciously. Addictive drugs act to continually reinforce the drug use experience as positive, even though the actual experience is never as positive as expected or may actually be perceived as quite negative. What makes this an instance of *hijacking*, according to Arpaly and Schroeder, is that addictive desires are especially resistant to an individual's evaluation of their worthwhileness. An addicted individual might come to hate drug use or even anticipate their use as unwanted or unpleasant. And yet, they will still strongly and persistently desire to use them. As they explain:

> This is where the metaphor of hijacking comes from. There is a normal process by which dopamine release leads to reward learning, and that process goes via perceiving or conceiving things that have come (through natural development or through learning) to be promoters of dopamine release—that have come to be intrinsically desired things. States of affairs are perceived or grasped cognitively, and if those represented are ones that are intrinsically desired, and are in excess of expectations, there is an increase in dopamine release, leading to unconscious reward learning. Addictive drugs circumvent this process. They promote unconscious learning, and all of its consequences, regardless of how their use is

perceived or conceived, regardless of whether or not using them satisfies any
intrinsic desires, regardless of expectations. A person who hates Imelda will not
have dopamine release promoted in her by seeing Imelda. But a person who hates
cocaine in just the same way, with just the same neural connections between the
idea of cocaine and the reward system as her counterpart has between the idea of
Imelda and the reward system, will nonetheless get a strong surge of dopamine-
type effects if that person consumes cocaine. This is the way in which addictive
drugs hijack the reward system. [2014, 278]

The hijacking metaphor is an attempt to capture the idea that the biological processes and

effects of addiction are not happening as they should.[146] Even if the same biological and

behavioral systems are involved, the motivational effects of addictive desires are normatively

unique from that of typical desires. This normative uniqueness is just that addictive desires are

formed and held independently of an agent's counter-evaluations. It is this supposed uniqueness

that allows Arpaly and Schroeder to claim that addictive desires are a nefariously amplification

of the motivational aspects of intrinsic desires. Someone like Carr may fundamentally care

about drug use, but, because of the hijacking, the motivating effects of his addictive desires are

far stronger than the actual strength of his care.

While the hijacking metaphor is a normative attempt to capture the problematic nature of

the motivational features of addiction, there is resistance to the aptness of this metaphor for

addiction. The metaphor communicates that an "outside" or "artificial" force has taken over, a

claim that some want to resist. Neuroscientist Marc Lewis (2011, 2017), for instance, argues that

---

[146] Araply and Schroeder are not the only advocates of the hijacking metaphor for addiction. There are other
versions of this argument as well, and they all share the goal of attempting to separate out addictive desires as
wholly unlike other "typical" desires. Dill and Holton (2014), for instance, argue that addictive drugs hijack the
reward system due to the fact that they "artificially stimulate" the motivation system (250). Interesting, they say
little to explain what exactly constitutes artificial stimulation except that addictive drugs can stimulate the
mesolimbic dopamine system without involving an experience that "also gives rise to pleasure" (2014, 29). This
claim is puzzling since drug use is often or even usually pleasurable, at least during use. See the use of the hijacking
theory also in Watson (1999a); Flanagan (2013b); Elster (1999), and even the Harvard Mental Health Letter (2011).
Marc Lewis (2017) attributes the use of the term "hijacking" in reference to the brain changes involved in addiction
to PBS correspondent Bill Moyer, who used the term in a popular 1998 PBS television series on addiction.

changes to the reward system from the continuous and heavy use of drugs like methamphetamines and cocaine do result in features like mental obsession, extremely heightened motivation, and impulsivity, but these are the result of the reward system working normally, even if chronic and heavy drug use can lead to especially strong motivation to continue their use due to their initial heightened pleasurability (2011, 152). As he explains, "The greater the hedonic impact of a stimulus, and the more it is repeated, the more rapidly and intensely dopamine will amplify the salience of its cues" (152). At least initially, drugs are "enormously exciting and pleasurable." Because of this, with sustained use, they have an extreme impact on the reward system and can result in extremely potent dopamine-induced cravings.[147]

The point here is that the unusual strength of addictive motivation can be explained without appealing to hijacking.[148] This position would explain why other rewarding activities can sometimes share the intense motivational profile commonly associated with addictive drugs. The heightened motivation involved in love, for instance, can sometimes persist strongly despite an agent's own opposing evaluations.[149] If Lewis's account is right, addictive drugs should not be likened to an outside force that "takes over" an agent's motivation. Drugs are, at least initially, highly pleasurable and their chronic, heavy use can result in heightened motivation to continue use. However, there is nothing normatively distinctive about them that would mark them as not being an agent's "own" in terms of their fundamental cares. This would mean that

---

[147] Lewis (2011, 2017) explains many other effects that addictive drugs have on the brain with sustained use that are all essential for understanding the natural and progression of addiction. However, since the dopaminergic system is thought to be the key modulator of motivation, I will focus on that here.
[148] See Racine & Barned (2019) for another critique of the hijack model in the clinical setting.
[149] Wonderly (2021); Sinnott-Armstrong and Pickard (2013); Lewis (2011, 2017).

addictive desires cannot be set apart from typical desires. Their motivational strength is no less an agent's own than that of any other desire that would be highly rewarding to satisfy.

Still, even if we accept the hijacking metaphor as aptly capturing how addictive drugs act on the brain and motivation, there remains another major concern with its use in conative QW theories. Mainly, I would contend that this metaphor is essentially a worry about loss of agential control. The core of the hijacking metaphor is that addictive desires can motivate independently of or even in spite of an agent's evaluations regarding the goodness of the motivated action. Even if an agent does not want to use anymore and cares deeply about not using, addictive desires can still be mentally consuming and highly motivating.[150] When hijacked, you are less able to do what you want to do or what you recognize is right. Yet, this is exactly what capacitarian RR theorists of moral responsibility have in mind with the notion of self-control or self-regulation.[151] The conflict between evaluations and strong motivating desires that are unresponsive to these evaluations is the conflict at the heart of self-control. The motivational hijacking is so concerning because one cannot, or is significantly less able to, align their behavior with what they deem best to do all-things-considered.

If I am right, then Arpaly and Schroeder's argument in defense of the claim that addictive desires are distinctive is essentially a veiled claim about how addictive desires make it more difficult for an agent to act on their evaluations against substance use. But this would mean that addiction is only mitigating because it impairs self-control, which would make a consideration of control a fundamental aspect of assessments of responsibility, an idea they explicitly reject.

---

[150] Dill and Holton (2014) use this metaphor to describe the same conflict.
[151] RR theories are not alone in conceiving of self-control in this way. See also Amaya (2020); Mele (2018), Sripada, (2014).

2.3 What These Arguments Show

Again, the challenge that addiction poses for conative QW views is to explain how addiction can mitigate blameworthiness for an addiction-explained wrongdoing even though addictive actions reflect an agent's fundamental cares. A primary strategy for providing this explanation is to distinguish addictive desires from typical desires, marking them or their motivational power as unrepresentative of what an addicted agent really fundamentally cares about morally. From what I have argued here, I think we have good reason to think that addictive desires are not functionally distinctive from intrinsic desires or fundamental cares. Moreover, I think that attempts to make them distinctive in terms of their biological story is fundamentally expressing a concern about how addictive desires impair self-control. The normative metaphor of hijacking just is trying to express the concern that addictive desires are strong, persistent, and resilient in the face of counter-evaluations. And, this conflict is so concerning because agent's are less able to do what they deem practically valuable. As I have suggested here, this explanation is exactly what is meant when many theories discuss compromised self-control

Where does this leave conative QW views? I think it suggests that, for conative QW views, explaining how the condition of addiction as experienced by agents like Carr can be mitigating can only be convincingly done by appealing to how the condition makes it especially difficult to act in light of one's own evaluations. I think this shows that these views need a notion of self-control to explain how conditions like addiction can be mitigating. However, I do not think that the the appeal to the importance of self-control is enough to make conative QW views explanatorily adequate with respect to explaining the possibility of mitigation on account of addiction. In the next section, I will suggest that to explain the most chronic and severe cases

of addiction, conative QW views also need to incorporate a notion of control that concerns the sufficient ability to track considerations against substance use as well.

3. The Case of the Diachronically Stable Willingly Addicted Agent

Carr's struggle with addiction illustrates how the condition can make it difficult to act on one's own evaluations. Carr cares about using cocaine, but he also cares deeply about his children and struggles against his addictive motivation in order to do what is right. But in some especially severe cases of addiction, the agent seems to care little about others, prioritizing substance use almost to the exclusion of other cares. This care exclusion problem seems to be a symptom of severe addiction itself. As some models of addiction suggest, addiction-induced *evaluative blindness* can fully pervade what an agent most fundamentally cares about, impacting how they interpret their situations and obligations or what they notice. If addiction can serve as a mitigating condition, it would seem that such extreme cases deserve the greatest degree of mitigation. But, as I will argue, to explain how addiction can mitigate blameworthiness in these especially severe cases, conative QW views need to incorporate into their theory a concern not just with self-control but with control over the ability to evaluate moral considerations.

3.1 Patrick Griffin

Consider again the case of Griffin described initially in chapter 2. His whole life has been a "consuming, grinding, unending cycle of addiction." Griffin makes the lives of those around him miserable. His mother and father are constantly in a state of emergency preparedness and they await his next drug-related crisis. His father's days are spent on the phone dealing with Griffin's ever accumulating medical bills and legal troubles. He once looked forward to retirement but now says his only hope is that his children do not die before him.

Griffin's mother expresses that because of his addiction, "I lost myself 10 years ago." Griffin's sister, who self-describes as a "former addict," distances herself from him to maintain her own sobriety. The family explains that it is hard whenever Griffin is jailed, but "it [is] nothing compared with the dread of having him out."[152] At his worst, Griffin seems only to care for others to the extent that they support his substance use. His words and behavior reflect an intentional effort to hurt his family out of anger and frustration. Even as his parents sit next to him on the bathroom floor after reviving him from an overdose, Griffin only reacts in anger. On a day when Griffin overdoses 4 times, saved each time by the quick actions of his parents, he responds to their pleas to get help by slicing their furniture with a kitchen knife while accusing them of not trying to understand or care about him.

Griffin's addiction influences his practical evaluations by shaping how he interprets his obligations to others. His cycles of cravings, use, and, withdrawal are the lens through which he interprets the actions and reactions of even those who have shaped their whole lives around the endeavor of trying to save his life. He interprets his parents' attempt to help as mean-spirited attempts to hurt him. Their complaints about how his addiction has impacted their lives is viewed as an intentional failure to try to understand him. Even the motives of complete strangers get interpreted through his addiction lens. When other drivers honk at him as he shoots fentanyl while driving down the interstate, it does not occur to him that what he is doing is wrong or dangerous to others. Rather, his only thought is of himself: "They didn't know that I needed to get this in me so I wouldn't be sick anymore."

---

[152] This is a sentiment shared by the families of other severe addicted agents. See Allie Armbruster's story in Hodges (2018).

3.2 Mitigation and Addiction Myopia

Some might find Griffin's case intuitively unsympathetic. He seems the paradigm of the willingly "selfish addict" who thinks only of himself and completely disregards the needs or interests of others. Yet, here I want to suggest that if any addicted agent is deserving of some degree of excuse for addiction-explained wrongdoing, it is someone like Griffin whose evaluations are wholly influenced by and have been wholly reduced to their addiction. Again, addictive desires are thought to influence evaluations and reasoning in ways that bias an addicted agent toward using. This attentional bias influences thoughts and promotes bias in deliberation, sometimes referred to as "deliberative narrowing" (Dill and Holton, 2014). In deliberative narrowing, considerations in favor of drug use are given excessive weight and considerations against drug use, such as future costs of use, are minimized.[153] An addicted individual might more clearly see reasons against drug use when not craving and in a motivationally "cool" moment, but then less clearly appreciate these reasons when craving and motivationally "hot," only to regret using once cravings are satisfied and heightened motivation to consume has returned to a cool state.

While agents in general experience attentional bias and deliberative narrowing in response to desires, severely addicted individuals are thought to experience it both more strongly and more frequently than typical. For addicted individuals like Griffin whose days and years are a cycle of cravings, use, and withdrawal, there can be little space for deliberative clarity that addicted individuals with milder addictions or typical desirers might experience. Rather, chronic, heavy use results in extremely biased attention and emotions directed toward drug

---

[153] Lowenstein (1999); Berridge and Holton (2013); Ainslie (2001); Bechara (2005).

use.[154] The addicted agent's attention is hyper-focused on the immediate reward of drug use to the exclusion of other concerns, ranging from larger, later goals like being able to maintain employment that one enjoys and maintaining relationships to even smaller, sooner goals that might compete with drug use, such as eating regular meals.

Building on this account, some theorists have suggested that in some cases of severe, chronic addiction, the ability for *evaluations* can get undermined by addiction. This attentional bias and deliberative narrowing is thought to impinge on an addicted individual's ability to track and appropriately weight moral considerations, resulting in an evaluative narrowing as well. Considerations that should be weighted heavily, like the welfare and safety of one's children, drop in priority or are not considered at all – victims of "addiction myopia." Addiction myopia is the idea that the condition "crowds out" engagement with other valuable things and activities worth valuing.[155] As noted in chapter 2, Wonderly (2021) suggests that the evaluative myopia of addiction is distinctive for the extent to which it crowds out other things of value or things that might potentially be of value to an agent and for the fact that it lacks the typically checks and balances that attend other attachments that might involve an evaluative crowding out, such as romantic love. The agent is effectively *blinded* to anything that might compete with the value substance use (24). In cases of especially severe, chronic addiction, evaluation blindness seems to be consistently experienced, though it is most severe during cravings.[156] Some severely and chronically addicted agents testify to being so blinded to the overwhelmingly sufficient reasons

---

[154] Lewis (2011, 2017); Hyman (2007); Heyman (2009). For studies on shifts in neural systems from systems that support long-term reward thinking to short-term reward thinking, see McClure, S.M., D.I. Laibson, G.F. Loewenstein, and J.D. Cohen (2004); Bjork, James M., Reza Momenan, and Daniel W. Hommer (2009).
[155] Wonderly (2021); Elster (1999); Watson (1999a).
[156] Wonderly (2021); Yafee (2011, 128).

against drug use that they could only begin to appreciate these considerations when they were forcibly separated from access to addicted substances for an extended time.[157]

3.3 Explaining the Importance of Evaluative Narrowing

If this model of addiction is right, at least in the case of severely addicted agents such as Griffin, then severely and chronically addicted individuals like him experience a "double whammy" from addiction: unusually heightened motivation and extreme evaluative narrowing. For this reason, it is plausible that addicted individuals like Griffin are deserving of some degree of mitigation. But I think the challenge to explain this mitigation is different and perhaps harder for conative QW views than explaining cases of addiction in which heightened motivation for drug use is the primary explanatory reason for the wrongdoing. In Carr's case, he both wants to care for his children and pursue and consume drugs, and his unusually heightened addictive desires make it hard for him to act on his evaluations to prioritize his children's well-being. But he still is able to evaluate his obligation to care for his children as important, and often he holds their well-being as more importance than substance use. The fact that Carr still holds this evaluation and recognizes this as a sufficient consideration against substance use opens the possibility for conative QW theorists to point to *this* evaluation as reflecting Carr's intrinsic desires. It is *this* evaluation that is being masked by his nefariously heightened motivation for drug use. Carr may intrinsically desire substance use, but his addictive desires impair his self-control in a way that makes it hard for him to act on what he values most.

---

[157] Hodges (2018, 26). Importantly, the testimony here is not advocating for the idea that forced treatment is effective for long-term cessation from substance use. Rather, it is the observation that for some severely and chronically addicted individuals only the complete inability to access and use addictive substances allowed for evaluative clarity.

In Griffin's case, however, this strategy does not straightforwardly work. Yes, he is highly motivated by his addictive desires much like Carr. But Griffin also seems to actually care about drug use much more than anything else. The Carr strategy addresses the motivational strength of addictive desires, but misses the importance of the myopic nature of his evaluations that favor drug use to the almost total exclusion of anything that might compete with it, a phenomenon that itself seems to be a product of his addiction.

In response to my concern, conative QW views might just bite the bullet and say that how evaluations are formed or held does not matter in assessments. The only question of concern in assessments of blameworthiness is whether one's actions reflect one's ill will or moral indifference. As in Carr's case, Griffin does express moral indifference and perhaps even ill will but not as much as it might seem given the strength of his motivation. For this reason, he might be deserving of some degree of mitigation. But Griffin is more blameworthy than someone like Carr because he fundamentally cares more for substance use and less for the well-being or rights of others.

However, I think this response ignores the distinctive way in which the most severe cases of addiction can implicate both the strength of desires and evaluations favoring drug use. The ability for addiction to shape the capacity for evaluations seems to be a very real feature of addiction.[158] Not recognizing how the ability for forming evaluations itself can be influenced by addiction results in a failure to explain the mitigating nature of addiction in the very worst manifestations of the condition.

---

[158] e.g.Watson (1999a); Holton (2009); and Summers (2015).

If conative QW views take up my challenge to explain the relevance of "evaluation myopia" for considerations of mitigation of blameworthiness, one approach might be to extend the scope of the hijacking theory to show that pro-substance evaluations in addiction are not truly reflective of fundamental cares because of their nefarious source in hijacked desires. As such an argument might go, in the same way that addiction can hijack the motivational strength of addictive desires, it can also hijack the ability for forming and holding evaluations in favor of or against drug use. The hijacking does this by producing unusually heightened addictive motivation that gives an all-consuming urgency to focus on pursuing and using drugs. This urgency, in turn, shapes evaluations in favor of satisfying urgent addictive need.

In response, I question whether the hijacking theory can be extended in scope so easily to explain the possibility that addiction might shape the content of evaluations. Again, the hijacking theory is offered to capture what exactly is normatively wrong with addictive desires. It does this by claiming that addictive desires are uniquely resistant to an agent's evaluations against them. But note that this explanation depends on the possibility of conflict between addictive desires and evaluations. The possibility of this conflict is missing from the problem of evaluative myopia. In Griffin's case, there is no conflict between evaluations and addictive desires or between evaluations about substance use, and this is precisely the problem. By merely extending the scope of the hijacking theory to cover the suspect nature of addictive evaluations, we still need to identify what exactly is so problematic about how addiction impacts the ability for evaluation. The content of evaluations are commonly influenced by strong desires. What is distinctively problematic about evaluations related to addiction?

Alas, I would contend that the most straightforward way of providing this explanation is to appeal to how addiction can make it more difficult for someone like Griffin to appreciate

considerations that might oppose the evaluations in favor of drug use He is blinded by his immediate need to the moral considerations in favor of not using drugs to stop hurting his family or others. His ability to step back from his immediate need to get perspective on his family's needs are so substantially impaired that even the most obvious practical considerations are missed. If I am right that this is the best explanation as to why Griffin's evaluation myopia is relevant for assessments of mitigation for blameworthiness, then considerations of control are again shown to be fundamental in our assessments. In the same way that explaining the normative importance of addictive desires for assessments of mitigation turns on how they make self-control harder, explaining the importance of myopic evaluations for these assessments turns on how they can make it more difficult for an agent to appreciate the moral considerations that sufficiently speak against drug use.

3.4 In Favor of Capacitarian Reasons Responsiveness

I have argued that conative QW views need to incorporate considerations of two notions of control in assessments of blameworthiness in order to explain how two manifestations of severe addiction can be mitigating. If I am right what implications does my argument have for thinking about responsible agency in the context of blameworthiness for wrongdoing? Do assessments of fundamental cares still have a role to play?

I think that my arguments above demonstrate that the considerations of control are more fundamental than considerations of fundamental cares. The reason that wrongful actions explained by addiction might be worthy of some degree of excuse is that it is hard for an agent to act otherwise in light of their evaluations. This is primarily a concern about self-control. Similarly, the reason that evaluation myopia matters is that it makes it harder for the agent to act

otherwise by impacting their ability to even appreciate the moral considerations that speak toward acting otherwise. These considerations of control do the fundamental work in the explanation of why addiction is mitigating. Fundamental cares may still play a role in explaining why Carr or Griffin act as they do or in describing the nature of their wrongdoing, but in terms of assessing whether they deserve mitigation for their addiction-explained wrongdoing, considerations of their abilities to recognize their sufficient practical moral reasons and to act in light of these reasons are what matter most. While we might try to salvage a hybrid view to make room for intrinsic desires or fundamental cares, I think we already have a good alternative in the form of capacitarian RR views of moral responsibility that already take these two kinds of control to be central in assessments of blameworthiness. Capacitarian RR views offer a viable theory that can straightforwardly explain why addicted individuals like Carr and Griffin might be deserving of some degree of mitigation.

I should clarify that my arguments above apply to Arpaly and Schroeder's assumed notion of blameworthiness. As mentioned in the introduction, Arpaly and Schroeder aim to provide the necessary and sufficient conditions for an attributability conception of blame rather than accountability. Yet, even if Arpaly and Schroeder want to explain how addiction can be a mitigating condition only for this more limited notion of blameworthiness, I have shown that they already appeal to considerations of self-control in the form of the hijacking theory and they they need to appeal to evaluative control if they want to explain mitigation in cases of addiction like Griffin's. They could avoid these appeals to control by just saying that addiction cannot mitigate blameworthiness in the attributability sense but it can mitigate in the accountability sense where considerations of control are necessary. But this would essentially be agreeing with

me that control should be a fundamental consideration in assessments of blameworthiness in the wide sense of accountability that I assume.

The discussion in this chapter shows that considerations of control are fundamental to explaining why and how addiction can be a mitigating condition of blameworthiness. What I have shown here suggests that views that make the two types of control central in their theories can better track and explain the features of concern in assessments of blameworthiness and excuse. For this reason, I think we have good overall reason to prefer capacitarian RR theories of moral responsibility over theories that do not take these two types of control to be central, including conative non-structural QW theories of moral responsibility.

Chapter 6: Culpability for Addiction and Capacitarian Reasons Responsiveness

1. Introduction

Capacitarian RR accounts of moral responsibility offer a satisfying explanation for why addiction can sometimes be a mitigating condition in assessments of blameworthiness for addiction-explained wrongdoing. As addressed in previous chapters, current models of addiction suggest that the condition features strong, persistent, and resilient motivating desires. In severe cases of addiction, these desires can partially undermine the ability for acting in light of evaluations against substance use or even the ability to evaluate the worthwhileness of drug use. The result is that addiction can make it overall more difficult for some addicted individuals to recognize or act on the sufficient moral reasons that speak against their continued substance use. As such, when an agent's addiction directly explains why they act wrongly, they might be worthy of some degree of mitigation due to their addicted condition.

However, this explanation of mitigation comes with a caveat. Even if an addicted agent's ability to resist substance use is sufficiently impaired at the time of the wrongdoing on account of their addiction, many think that this incapacity is *not* excusing if that agent is culpable for their addicted condition. Further, many also think that most addicted agents are indeed culpable for their addicted condition. As the thought goes, it is common knowledge that substances like opioids, alcohol, or cocaine are addictive. And it is also commonly known that addiction can be self-destructive and harmful to others. Further, the incapacity associated with addiction is typically episodic, allowing for the possibility of recovery. As such, most addicted agents can be culpable for becoming or remaining addicted. This concern about culpability for addiction has led some theorists to argue that overall addiction is typically not a mitigating condition even if it can sufficiently compromise responsible agency at the time of the wrongdoing (Morse, 2007; Glannon, 2009; Moore, 2020; Brink, 2021).

How might this intuition about culpability be explained as relevant for assessments of blameworthiness? Even if it is intuitive to many that addicted agents who act wrongly on account of their condition are fully blameworthy for that action on account of culpability for their condition, how might a capacitarian RR view give a plausible explanation as to whether and why such agents are indeed culpable for their condition and as to how this culpability matters in assessments. In this chapter, I will consider two general strategies that RR theorists have used to explain blameworthiness in the case of addiction given the intuition that addicted agents are culpable for their condition. One such strategy involves "tracing" responsibility back to an earlier point in time when the agent had sufficient control and could foresee the possibility of the incapacity and its related wrongdoing (Fischer and Ravizza, 1998; Fischer and Tognazzini, 2009). A second strategy is to hold the agent responsible for unjustifiably risking the possibility of incapacity and its related consequences (King, 2014; Agule, 2016; Brink, 2021). I will suggest that, given that addiction tends only to partially undermine normative competence, both of these strategies have problems in their attempt to assess the blameworthiness of addicted agents for their addiction-explained wrongdoing. In response to these concerns, I will suggest a third strategy that seeks to assess agents in terms of their capacities at the time of the wrongdoing in question.

To make my case, I will first motivate the idea that addicted agents are generally culpable for their addiction. In part 2, I will then consider what I call the "tracing strategy," which seeks to derive responsibility for a wrongdoing done while incapacitated at prior point in time when the agent is sufficiently normatively competent and can be reasonably expected to foresee the risk of wrongdoing. I will show how addiction-explained wrongdoing presents its own form of the worry that the view double counts blameworthiness for a single wrongful outcome. In

response to this view, I will propose what I call the "circumvention strategy" that explains

addiction-explained wrongdoing by reference to the capacities that the agent could exercise to

avoid the wrongdoing. In part 3, I will consider the "negligence/recklessness strategy" that seeks

to hold addicted agents culpable for the act of unjustifiably risking an incapacity and its

consequences. I will contend that this view also faces a worry about double-counting

blameworthiness. Finally, in part 4 I will reconsider the circumvention strategy and attempt to

show how the negligence/recklessness strategy can be called upon to explain culpability for

addiction and the outcome of period partial normative incapacity even if the circumvention

strategy best explains responsibility for addiction-explained wrongdoing.

1.1 Culpability for Addiction

Many philosophers accept that addiction can reduce reasons responsiveness with respect

to resisting substance use.[159] In some cases, the reduction in the capacity for responsiveness can

be so severe that it might mitigate blameworthiness for wrongdoing directly explained by the

addiction. However, some contend that even if addiction can sufficiently undermine normative

competence, addicted agents are still fully accountable for their addiction-explained wrongdoing

because they are culpable for their addicted condition. Walter Glannon (2009), for instance,

argues that agents addicted to alcohol who develop alcohol-related end-stage liver disease as a

consequence of their addictive alcohol use can be held accountable for the wrongdoing of

unjustifiably adding competition for a scarce medical resource even if their addiction sufficiently

comprises their responsiveness with respect to their consumption of alcohol once they become

addicted. He calls upon a "transfer principle" of responsibility to argue that if one has control

---

[159] Brink (2021); Glannon (2009); Moore (2020); Kennett (2014); Flanagan (2013b)

over when they initially started drinking and know the consequences of their drinking, they can

be held accountable for the wrongdoing by way of blaming responses in the form of a lower

priority for a liver transplant. As he explains:

> If one is not compelled to begin drinking by genetic or environmental factors
> beyond one's control, and if one is not so cognitively disabled or immature as not
> to be capable of knowing the probable addictive consequences of chronic
> drinking, then one can be responsible for the addiction, as well as for the liver
> failure that may result from it. The history of how this condition came about may
> involve enough control for a person to be causally responsible for the addiction
> and subsequent liver failure. [Glannon, 2009, 27]

Brink (2021) likewise suggests that addicted agents can forego their claim to an excuse

for a wrongdoing due to incapacity if the agent is "substantially responsible for bringing about

those conditions" (355). Brink helpfully distinguishes two ways in which an addicted agent

might be culpable for their addiction. First, they can be culpable for *becoming* addicted if the

condition is "acquired voluntarily and responsibly" (355). Second, an addicted agent can be

culpable for "*remaining* addicted over time and not seeking effective treatment" (355, emphasis

mine). While he allows that there may be cases of non-culpable addiction, this is usually limited

to *acquiring* an addiction and includes situations such as becoming addicted under legitimate

medical direction or acquiring an addiction during childhood or adolescence when one is

presumably sufficiently less responsive to the reasons against substance use. Still, Brink allows

the possibility that someone might become addicted non-culpably and also struggle with resisting

substance use despite diligent effort, making the addiction non-culpable overall, though he

implies that such situation is rare.

Glannon and Brink are not alone in thinking that addicted agents are somehow culpable

for their condition. Consider that many social welfare policies treat addiction as distinctive from

supposedly non-culpably acquired incapacitating conditions, such as head injuries, even if

addicted agents are sufficiently incapacitated by their condition once addicted. For instance, addicted agents are not eligible for disability benefits though the U.S. Social Security Administration (SSA) and only recently has addiction rehabilitation been covered by many medical insurance plans.[160]

So, what are some considerations in support of the idea that most addicted agents are culpable for their addiction? In terms of initially becoming addicted, it is common knowledge that certain substances can lead to an addiction. In an era of warning labels, highly publicized lawsuits against pharmaceutical companies, and a proliferation of the substance rehabilitation industry accompanied by aggressive advertising, it can be reasonably expected that any typical adult in a country like the United States is well aware of the potential addictiveness of many common targets of addiction, such as stimulants like cocaine or methamphetamine and depressants like opioids and alcohol. Some individuals even have direct knowledge of their own susceptibility to addiction given their family history. In addition, it is common knowledge that addiction is associated with wrongdoing. Narratives of the harms of alcoholism are common in popular media, police agencies release images of scenes of child neglect resulting from addicted parents, addiction memoirs abound, and, again especially in the case of alcohol, some have firsthand familiarity with the wrongs that follow on the heels of a severe addiction.

Moreover, it seems that addiction is not constantly or permanently incapacitating and is a condition that most agents can overcome. Addictive desires, even if persistent over time, are not constantly present and are not always equally cognitively nor volitionally overwhelming. This

---

[160] This information is accurate as of 2022. Some change in policies is underway, though, due to the campaign in the scientific and medical community to categorize addiction as a brain disease. Even then, there is popular resistance to this approach.

oscillating nature allows for both motivational reprieve and evaluative clarity.[161] Also, given

recent advances in addiction treatment, there seem to be many viable options for addictive

recovery, including pharmaceuticals that lessen the intensity of cravings, individual and group

counseling, motivational programs, and even support by some criminal courts to aid in treatment

as an alternative to criminal penalties.[162]

These considerations together lend support for the idea that, in general, agents can

reasonably be expected both to *foresee* the possibility of addiction and its consequences and to

control their substance use before becoming addicted. In addition, even once addicted, these

considerations support the idea that addicted agents can, as least at certain points, still appreciate

the sufficient reasons against substance use and successfully act on these reasons to successfully

overcome their addicted condition, even if doing so takes considerable time and effort.

2. Tracing Strategy

How might a capacitarian RR best account for the importance of these considerations,

especially given that acquiring an addiction or experiencing a moment of motivational and

evaluative reprieve can happen before – sometimes long before – the actual wrongdoings that

addicted agents commit because of their incapacity?

Fischer and Ravizza (1998) and Fischer and Neal Tognazzini (2009, 2011) defend the

idea that responsibility for a wrongdoing performed while incapacitated at a particular time (call

this "T2") can be *traced* back to prior point in time (call this "T1") at which an agent "has

sufficient control over the action" (i.e. they are reasons responsive with respect to that action)

---

[161] As I discuss in chapter 2 and 5, some models of addiction challenge this assumption with regard to some severe cases of addiction.
[162] Westervelt (2017); Yoon et al. (2021); Gupta, S. (2019); La Gorce, T. (2021).

and they "can be reasonably expected to know the likely results of that action" (Fischer and

Tognazzini, 2009, 532).[163] Fischer and Tognazzini do not require that an agent make an actual

conscious choice or decision to engage in the action that risks the incapacity and outcome at T1.

Rather, they only require an "act of agency," which may be a choice or an omission to make a

choice. Other tracing views, such as that proposed by Nelkin and Sam Rickless (2017), identify

this early point at T1 as an instance of a *witting failure* to take measures to attempt to avoid the

later outcome (124).[164]

This account works especially well for explaining why a severely inebriated non-addicted

driver can be blameworthy for harms inflicted as a result of intoxicated driving. Such a case is

illustrated in the hypothetical story of Dan who is considered blameworthy for running over a

pedestrian even though he was sufficiently incapacitated by inebriation at the time of the

wrongdoing.[165]   There is point in time (T1) at which Dan, with sufficient control and foreseeing

the results of his actions, makes the choice to drink to the point of inebriation even though he

knows or should foresee that he would likely drive, which in turn significantly risks harm to

himself or others.[166]

Note that the idea here is that an agent's responsible agency in regard to the wrongdoing

is assessed as if it is not compromised at the point of the wrongdoing because the exercise of

---

[163] Again, Fischer is committed to the RR Mechanism view. The tracing strategy expressed in terms of a mechanism is: "When one acts from a reasons-responsive mechanism at time T1, and one can reasonably be expected to know that so acting will (or may) lead to acting from an unresponsive mechanism at some later time T2, one can be held responsible for so acting at T2" (Fischer and Ravizza, 1998, 50). However, the idea can apply to the reasons responsiveness of an agent as well.
[164] Nelkin and Rickless offer their tracing view as an account to explain unwitting omissions specifically, but I think the account works for explaining responsibility for actions as well.
[165] Fischer and Tognazzini (2009) mention this case. But a case of this sort is discussed by many others as well.
[166] Some RR views treat the awareness condition, also commonly called the epistemic condition, as an aspect of the control condition. On such a view, being sufficiently reasons responsive requires meeting the awareness condition. See Nelkin and Rickless (2017) for an example of such a view. While I am sympathetic to this approach, I will speak of both conditions here since, in this discussion, foreseeability is of particular concern.

responsible agency with regard to the wrongdoing at T2 is derived from the exercise of agency at T1. In regard to the wrongdoing, the addicted agent is assessed *as though* he did the wrongdoing while sufficiently meeting the standard for full reasons responsiveness. Consider again Dan. Given his control over his action of drinking at T1, he can be assessed as fully responsible at T2 when he hits the pedestrian with his vehicle. As stipulated, even though Dan is completely incapacitated at the time of hitting the pedestrian, he has control over the decision to get drunk prior to his first drink and can be reasonably expected to have foreseen that he would drive drunk. It is also reasonable to expect that he should know that driving intoxicated, especially extremely intoxicated, significantly risks harm.

The Dan case nicely illustrates many features that seem relevant to considerations of culpable incapacity. The risk of impaired driving is extremely likely. The causal link between the risked behavior and wrongdoing is also tight. It is easy to identify a specific time at which the choice to engage in risky behavior is made. Plus, it is reasonable to expect that he should have foreseen the likely results of his action both in terms of the incapacity due to inebriation and the significant likelihood of wrongdoing. Given that the choice to engage in the action that leads to the pedestrian's death can be clearly identified at a point where Dan was sufficiently responsive and aware, on the tracing strategy he is blameworthy for the wrongdoing as though he hit a pedestrian with his car while fully reasons responsive.

2.1 Applying the Tracing Strategy to Addiction

Many find the straightforward application of the tracing strategy as illustrated by the Dan case to be highly compelling. However, the application to cases of wrongdoing explained by addiction is more complicated. The tracing strategy might explain culpability for addiction in

two ways. First, addicted agents might be considered responsible for wrongdoing at T2 resulting from an addiction because there is an earlier point T1 at which the agent had sufficient control over *acquiring* an addiction and could or should have foreseen the possibility of becoming addicted and the wrongdoing. Or, second, addicted agents might be considered responsible for wrongdoing at T2 resulting from an addiction because there was an earlier point T1 at which the agent had sufficient control over *staying* addicted and they could or should have foreseen the possibility of remaining addicted and the wrongdoing that would likely follow.

In assessing the application of the tracing strategy to addiction, first consider the plausibility that addicted agents meet the control or awareness conditions at T1. In terms of *acquiring* an addiction or *becoming* addicted, is it fair to assume that most or even many addicted agents could have met the control and foreseeability conditions? In Dan's case, there is a definitive point in time at which the action to risk the incapacity and wrongdoing takes place. In contrast, however, there is no such definitive point for acquiring an addiction. Rather, addictions are typically developed through a series of actions that accumulate over time. At the point of any one of those actions, acquiring an addiction on account of *that* action is quite small, making any one point an unlikely candidate for serving as *the* T1 from which responsibility can be derived in terms of initially acquiring an addiction.

However, assuming that there is such a point, I think we should resist the idea that many or maybe even most non-addicted agents can meet the foreseeability condition. Namely, I do not think it is reasonable for most non-addicted agents to anticipate that they risk an addiction, let alone addiction-explained wrongdoing. To illustrate my concern, compare the case of addiction to a seemingly analogous case offered by Manuel Vargas (2005) of Jerky Jeff, who does seem to meet the foreseeability condition for acquiring a condition associated with wrongdoing. Jeff is a

165

mean and nasty man who fires his employees in a very inconsiderate way and is intuitively blameworthy for this wrongdoing. At the time of the firing, Jerky Jeff is presumably not sufficiently normatively competent in regard to his wrongdoing. But Fischer and Tognazinni (2009) argue that it is reasonable to expect that that Jerky Jeff should have foreseen both that he would become a jerk on account of his individual actions and that he could foresee the general wrongdoings that would likely occur as a result, namely that he would treat "some people poorly at some point in the future as a result of his jerky character" (537).

While the conditions of acquiring an addiction and a acquiring a bad character in this case share many features in common, e.g. both significantly risk wrongdoing at some point, both take time to develop, and both develop from a series of actions, there is one important relevant difference. Namely, while it is reasonable to expect someone to foresee that they will become a jerk by repeatedly acting like a jerk, it is often not reasonable for someone who consumes an addictive substance regularly to expect that *they* will acquire an addiction. There is substantial empirical evidence that most people who use an addictive substance regularly will not go on to develop an addiction.[167] Of course, there are exceptions. In some instances, it is known both that addiction to a particular substance has a genetic component and an agent can be aware of their own genetic susceptibility to addiction given their family history. Both of these conditions can often be met in the case of alcohol addiction.[168] For most others, though, knowing one's own risk for developing an addiction to a particular drug is not readily available. Research about genetic susceptibly to non-alcohol addictions is nascent, let alone accessible at the individual

---

[167] Anthony et al. (1994).
[168] Blum et al. (1990).

level.[169] For those who do not know their own susceptibly to addiction, chances are low that they will acquire one. For instance, even among those who regularly use alcohol, only 1 out of 10 are thought to even have a minor addiction.[170] It is reasonable to expect that most non-addicted agents should know the connection between using addictive substances and addiction and between addiction and harm, but it is not reasonable to expect that most should foresee *their* own significant likelihood of acquiring an addiction.

Even if I am right, however, this concern about foreseeability is not a challenge to the tracing strategy per se. Rather, for any strategy that attempts to hold addicted agents accountable for becoming addicted, this concern is relevant if the strategy also includes a foreseeability condition. Still, we might ask whether an addicted agent can be culpable for remaining addicted even if most are not culpable for acquiring an addiction. Some evidence suggests that it might be reasonable to expect that most addicted agents can discontinue addictive substance use since most addicted agents do.[171] One consideration to support this idea is that addictive motivation tends to exert itself in waves of intensity, allowing for periods of clarity in which addicted agents are responsive to the considerations that speak against their continued substance use. These periods of reprieve can be considerable for those with mild or moderate addictions, but there are likely such periods even for individuals with severe addictions. Still, in the case of severe addiction, the periods may only be long enough for tracking the considerations against substance use but not for actually effectively acting on this recognition. Getting a handle on a severe addiction is effortful and typically requires an extended period of time in which one can find

---

[169] Hiroi & Agatsuma (2005).
[170] (CDC report); Vowles et al. (2015).
[171] Winick (1962). Many people cite this study in particular, though it has recently been challenged. See Vergés et al. (2013).

resources, discover which resources are actually effective, and implement a plan to make continual use of these resources. There can be immense challenges to each of these steps even for those who can retain evaluative clarity regarding their addiction for an extended period of time. It can be overestimated how accessible or effective addiction recovery resources are for cases of severe addiction. For some addicted individuals, resources are simply not available. But others have to deal with predatory rehabilitation programs, increasingly cumbersome access to pharmaceuticals like Suboxone that are proven to help with addiction, political resistance to effective treatment programs, etc.[172]   For those who struggle in maintaining evaluative clarity over time, making effective use of these resources can be exceedingly difficult.

Still, given what can be at stake morally once an addiction is acquired, it seems reasonable to expect that considerable effort be expended to resist addictive substance use, which supports the idea that there may be points at which many severely addicted agents meet both the control and foreseeability conditions with regard to remaining addicted.

2.2 Assessing the Tracing Strategy

The tracing strategy can give an account of why some addicted agents are culpable for acquiring an addiction and many more are culpable for remaining addicted. Still, though, there are reasons to think that the tracing strategy is flawed in assessing blameworthiness in the case of addiction. Take two recent arguments against the tracing strategy. First, Michael Moore (2020) raises the concern that this strategy equates the agent's blameworthiness at T1 with how blameworthy the agent would be *if* he had done the action at T2 while sufficiently reasons

---

[172] Goodnough, Abby (2020); Mann, B. (2021); Pattani, A. (2021). Take the case of contingency management programs. Despite success rates, contingency management programs are frequently not publicly funded due to concerns about "rewarding" culpable addicted agents. California recently rejected supporting these programs via a veto by the governor.

responsive (414). The worry here is that this account assesses the agent as more blameworthy than he actually is for the wrongdoing. Yet, the agent is blameworthy for the harm done to the pedestrian, but the blameworthiness should not be the same as doing the act while fully reasons responsive because the agent is *not* fully responsive.

A second related concern raised by both Craig Agule (2016) and Brink (2021, 108) is that the tracing strategy double counts blameworthiness such that the agent is actually held accountable twice for one outcome. The concern here, as applied to a case like Dan, is that when the agent initially freely acts to get drunk knowing that he significantly risks driving while intoxicated and that doing so significantly risks harm to others at T1, he is actually blameworthy for either negligence or recklessness at T1 and its consequences. The wrongdoing here is knowingly taking an unjustified risk that has a significant likelihood of causing harm and the consequence of the pedestrian being harmed. But, the tracing strategy *also* considers the agent fully blameworthy for the wrongdoing of harming the pedestrian at T2 as derived from the agent's sufficient normative competence at T1. Thus, the addicted agent is accountable for *two* actions related to the same outcome. Yet, the concern is that the result in cases of culpable incompetence should not be that the agent in question is even *more* blameworthy than someone who brings about the same consequences while acting with full normative competence.

If worries about assessing the addicted agents as *more* blameworthy than they are is a legitimate concern, these two concerns might be enough to reject the tracing strategy. Here, though, I want to raise a third concern that turns on the fact that at the point of addiction-explained wrongdoing addicted agents are usually only *partially* incapacitated and not fully incapacitated. As discussed in chapter 2, even in severe cases addiction usually only partially undermines control over substance use. This suggests that in cases of addiction-explained

wrongdoing, addicted agents can be partially responsible even at the point of the action in question. On the tracing strategy, this would mean that at T2 the agent is likely partially responsive in respect to his wrongdoing. This complicates the tracing strategy because there is not a straightforward reason for why we would consider an agent who is already responsible to some extent for the wrongdoing at T2 fully responsible for the wrongdoing by appealing to their *fully* responsive state at T1.

Given this consideration, why not just say that severely addicted agents are partially responsible for their act at T2 and then set aside the consideration about culpability for addiction? One possible reason is that, to many, an assessment of partial responsibility seems too lenient in the case of addiction-explained wrongdoing. Applying the tracing strategy in the case of partial normative competence would mean holding the addicted agent *fully* responsible for the wrongdoing at t2 and this assessment *just is* the right verdict. In other words, addicted agents should be considered *fully* responsible for their wrongdoing and not just *partially* responsible because they really are fully responsible.

Perhaps it is right that addicted agents should be considered fully responsible for their wrongdoing. But I think there are reasons to resist using the tracing strategy to affirm this verdict. Beside the still present worry about double-counting blameworthiness, the view must also explain why it is appropriate to assess cases involving partial incapacity as if the agent were fully responsible. In assessing reasons responsiveness at the point of the wrongdoing, capacitarian RR views already have a means to explain responsibility and blameworthiness. What would justify looking backward to a point well beyond the proximate history of the agent to find a different point in time to anchor responsible agency? My primary concern here is that, depending on the exact standard of what is required in terms of a controlled action and

foreseeability at T1, this view will consider most agents who are partially responsible at T2 as fully responsible if there is a suitable point at which they could have or should have been aware of the likely results of an action that led to their partial incapacity. I think this strategy loses its appeal when applied to cases of non-addicted agents that are structural similar to the case of addicted agents.

2.3 The Option of Circumvention

There are, then, good reasons to reject the tracing strategy for assessing blameworthiness in cases of addiction-explained wrongdoing. This may not be a problem for capacitarian RR views since they already have the theoretical apparatus in place to account for assessments of responsible agency in cases of partial incapacity. As many capacitarian RR views conceive of volitional control, acting on one's sufficient moral considerations can require synchronic or diachronic tactics in the form of either *conquering* errant desires using sheer mental exertion in the moment to do the right thing or *circumventing* these desires to either avoid acting on them or avoid the desire-related wrongdoing. Given these two recognized aspects of volitional control, when assessing an agent who commits a wrongdoing capacitarian RR views consider if there any point in the *proximate* history of the agent such that it is reasonable to expect that the agent could have done something to resist acting on the errant desire or to avoid the wrongdoing in some other way. On this approach, an addicted agent can be assessed as *fully responsible* for an addiction-explained wrongdoing even if they are partially compromised in respect to being able to conquer the addictive desire at the exact point of the outcome if they were sufficiently able to take steps to avoid substance use or the wrongdoing. To illustrate, Carr would be assessed as fully responsible for his addiction-explained wrongdoing if he could have without substantial difficulty called a neighbor to come stay with his twins at his apartment that evening.

Presumably, though, many severely addicted agents will only be partially responsible even as assessed by the circumvention strategy.

As tool for determining responsibility for acts done while partially or fully incapacitated, this strategy does depend on a more precise understanding of the scope of "proximate" history of an agent. Capacitarian RR views tend to limit assessments of agents to the capacities of the agent at the time of wrongdoing. But the notion of proximate is meant to include "proximate aspects of her situation, psychology, and deliberation that have a direct bearing" on an agent's action (Brink, 2021, 103). On such a view, it would not be a stretch to say that the psychological state of the agent in the days or possibly weeks prior to the action in question can be included in the scope of the proximate history of the agent. It is less plausible that scope would extend to include the situation or psychological states of the agent months or years prior to the action in question. As such, this strategy will not likely account for the culpability of *acquiring* an addiction in many cases and might not even account for culpability for remaining addicted. Still, the circumvention strategy provides a straightforward way to explain blameworthiness for addiction-explained wrongdoing.

3. The Negligence/Recklessness Strategy

The circumvention strategy seeks to assess addiction-explained wrongdoing by focusing on the proximate history of the agent and whether they could have exercised any abilities to resist acting on their addictive desires or to avoid the wrongdoing. Still, some might resist this strategy. One worry is that it wrongly assesses the severely addicted agent as only partially responsible because it neglects incorporating considerations about culpability for acquiring or maintaining an addiction. The right assessment is that the addicted agent is fully responsible for

their addiction-explained wrongdoing. As such we need a better strategy to account for this intuition and explain the role of culpability for addiction.

Another possible strategy that might check all of these explanatory boxes is the negligence/recklessness strategy (N/R strategy).[173] This strategy can explain the relevance of earlier acts that contribute to acquiring or remaining addicted. Basically, the strategy assesses responsibility by asking whether an agent unjustifiably risks a reasonably foreseeable outcome and its harmful consequences. The difference between recklessness and negligence turns on the agent's state of mind with regard to the link between the action and the risked harm. In the case of recklessness, the agent is aware of but disregards an unjustifiable risk of significant harm. In the case of negligence, the agent fails to be aware of the unjustifiable risk of significant harm though it is reasonable to expect that they should have been aware of it. Notably, on the negligence/reckless strategy, an agent who engages in the action that risks significant harm is specifically accountable for the negligent/reckless act and its outcome. If the outcome in question is at T2, the strategy holds the agent accountable for what they did or failed to do at T1 that unjustifiably risked this outcome. This means that responsible agency for the act at T2 is not anchored in T1. Rather, the agent is being assessed only for what they do at T1 and its consequences.

As applied to the case of Dan, this strategy would say that his wrongdoing is one of recklessness, assuming that he did actually consider the possibility of the significantly risked harm as he started to drink. Given that he engaged in this act while fully reasons responsive, he is fully responsible for acting recklessly even if no one is ultimately harmed. Since he actually

---

[173] A version of this view is defended by King (2014); Agule (2016); and Brink (2021).

hit a pedestrian while incapacitated, he is responsible for his reckless act and the outcome of the act – an overall more severe wrongdoing than merely significantly risking the foreseen harm.

3.1 Applying the Negligence/Recklessness Strategy to Addiction

How might the N/R strategy apply to a case of addiction? Can addicted agents be considered negligent or reckless for their addicted condition and its consequences? In terms of accounting for *acquiring* an addiction, I will reiterate my point from section 2.1. I think we should resist the idea that it is reasonable to expect that many or maybe even most agents who use addictive substances meet the foreseeability condition with respect to becoming severely addicted. There are exceptions, but any view that requires reasonable foreseeability for a condition like addiction will have a hard time making the case that individuals without a family history of addiction could have or should have foreseen that they significantly risked a severe addiction. For the remaining agents who could reasonably be expected to foresee the significant risk of their addiction, the N/R strategy does provide an explanation for how they might be assessed as blameworthy for acquiring their condition.

Again, though, it is more plausible that many agents, even those suffering from severe addictions, might be negligent or reckless for *remaining* addicted because they experience points at which they are both able to appreciate the considerations against their substance use and take advantage of resources to help reduce or stop their addictive use and they are able to foresee the harm of continuing use, especially early on in their addiction. Many addicted agents may be assessed as responsible for negligently or recklessly remaining addicted by failing to exert effort to overcome their addiction and prevent harmful addiction-related consequences.

3.2 Assessing the Negligence/Recklessness Strategy

A presumed primary benefit of this strategy is that it avoids double-counting blameworthiness for a single outcome by not locating responsibility for the harm at T2 *derivatively* on an action at T1. There is only one locus of accountability: the competent assumption of an unjustified risk at T1 and the outcome of that risk.

Another possible benefit of this view is that it manages to satisfy two seemingly incompatible intuitions about addicted agents and addiction-explained wrongdoing. On the one hand, many have the intuition that addicted agents are *fully responsible* for their addiction-explained wrongdoing because they are somehow culpable for their addiction. The N/R strategy accounts for this intuition by showing how addicted agents can be fully responsible for acquiring or remaining addicted because they have sufficient control with regard to negligently or recklessly acquiring an addiction or continuing in their substance use once addicted. On this strategy, many addicted agents do possess sufficient responsible agency with regard to their addiction and its harmful consequences.

On the other hand, this strategy can also accommodate the intuition that addicted agents who act wrongly on account of their addiction are not overall as *blameworthy* as non-addicted agents who do a similar harm while fully responsible.[174] The N/R strategy can account for this intuition too because it actually identifies a different wrongdoing as the focus of the assessment than it would in the case of a non-incapacitated agent who acted similarly. To see what I mean, take again the case of Dan. The N/R strategy identifies his act of willingly drinking to the point of inebriation while foreseeing the risked consequences as the moral wrong to be assessed. The moral wrong in question is the reckless act and its consequences. In contrast, if an agent we call

---

[174] As I mentioned in chapter 5, many philosophers have this intuition.

Ron is sufficiently responsive when he hits a pedestrian, this act would be the focus of the assessment. The moral wrong would be hitting the pedestrian.

Again, this difference is not just a difference in which act is identified but is also a difference in the nature of the wrongdoing that is identified as the focus of assessment. Of course, whether there is an actual difference in the severity of the wrongdoings being compared turns on the severity of Ron's wrongdoing in comparison to Dan's. If Ron intentionally hits the pedestrian then it is plausible that, even though Dan and Ron are both sufficiently responsive with regard to their respective wrongdoings, Ron is more blameworthy because his wrongdoing is more severe.[175] But if even both acts are assessed as instances of recklessness, the acts themselves may still have morally different weights in terms of severity. While Dan is risking inebriation that in turn risks harm, the non-incapacitated driver more directly risks a significant harm.

Given the possibility of differences in the severity of the wrongdoing, this opens the possibility that addicted agents may be less blameworthy for their addiction-explained wrongdoing than non-incapacitated agents who act similarly.[176] If this is right, then the N/R strategy can preserve the intuition that the addicted agent is fully responsible with respect to remaining addicted and harmful outcome, while also preserving the intuition that addicted agents should be assessed as less blameworthy than their non-addicted counterparts who act to bring about a comparable harmful outcome.

---

[175] This move to differentiate the severity of wrongdoing by the agent's mental state with regard to the action turns on the idea that *intending* to act wrongfully is morally worse than acting negligently or recklessly. See Brink (2021, 160) for an explanation of this hierarchy of moral gradation in the context of the criminal law notion of *mens rea.*
[176] This of course rests on the further assumption that blame should be proportional to the agent's *desert*, understood as the product of their wrongdoing and their responsibility for it – an idea that I endorsed in the introduction.

However, despite these virtues, the N/R strategy, may face a problem in light of the fact that at the time of the harmful outcome, many addicted agents are only partially incapacitated by their condition. The concern here is that even if an agent's reckless or negligent act and its outcome is the target of assessment, there is an additional plausible target of assessment at the point of the outcome. Thus, it looks like the N/R strategy may also fall prey to the double-counting worry. If this is the case, then in instances of partial responsibility, it looks like the N/R strategy encounters similar problem faced by the tracing strategy. The addicted agent is blameworthy both for risking their incapacity and its outcome and for the act that directly brings about the outcome. If double-counting is a mark against the tracing strategy, then in the case of partial incapacity it is a mark against the N/R strategy as well.

4. Avoiding Double Counting and the Role of the Circumvention Strategy

I think the double-counting worry with both strategies lends support for the idea that the circumvention strategy provides the best framework for assessing an addicted agent for addiction-explained wrongdoing. Again, this strategy takes into account an agent's capacities at the point of the act of harm as well as the relevant proximate situational and psychological aspects of the agent that directly bear on whether the agent is able to act on their sufficient moral considerations in assessments of responsible agency. In the case of agents whose normative capacities are sufficiently but partially undermined with respect to the wrongful act because of their severe addiction, this would mean assessing them as partially responsible for their wrongdoing. Importantly, this assessment is focused entirely on the agent's capacities with the regard to the wrongdoing and is not concerned with culpability for acquiring or maintaining their addiction, unless either is part of the proximate history of the agent.

Still, the worry that this strategy is too lenient in its assessment of blameworthiness may lurk. It might still somehow seem relevant that some addicted agents are partially or fully responsible for acquiring or maintaining their addicted condition. If this intuition is right, the circumvention strategy seems to be missing something important in assessing cases of addiction-explained wrongdoing. So, even if the tracing or N/R strategies do not provide the best explanation of this culpability, why think that we should settle for a view that entirely neglects addressing culpability for addiction.

There may be some good reasons to stand firm and reject the idea that culpability for an addiction is something that we should seek to tie directly to assessments of wrongful acts that are explained by addiction. Mainly, there is not a clear reason why we should seek to find addicted agents culpable for their partially incapacitated conditions when there are numerous "conditions" that agents voluntarily acquire or maintain that can episodically undermine responsible agency to a sufficient degree but that we do not attempt to assess culpability for.

The urge to hold addicted agents culpable for their condition invites assessing agents responsible for all sorts of conditions that can feature episodic partial incapacity and that risk harm as a result. Consider taking on a demanding, exhausting job. Being chronically tired and mentally taxed can make it considerably more difficult to recognize sufficient moral reasons or act in light of these considerations. Someone who is consumed by a demanding job may be so focused or busy that on occasion they lack the sufficient volitional control required to act on their recognition of the sufficient reasons to prioritize the emotional needs of their children, assuming that they have the mental space to notice these needs. Assuming that this parent could exercise sufficient control with accepting or remaining at this job, they would be assessed as culpable for this choice as well.

One might respond that what will distinguish addiction from these other cases is that acquiring or maintaining an addiction *unjustifiably* risks harm. Voluntarily taking on or remaining in a demanding job when one foresees the risk of significant harms as a consequence is not a culpable act because it is justified, whereas voluntarily taking on or remaining in an addiction is not. This appeal to justifiability as the distinguisher may have promise. But I think we should be cautious about how effective it will be in sorting addiction from other cases that risk episodic partial incapacity. Namely, I think that the reasons in favor of substance use - even addictive substance use - can be underrated. Consider cases of severe, chronic physical or emotional pain that can only be managed with heavy opioid use or cases of those who find a good and meaningful life in their drug addiction.[177] In contrast, the reasons in favor of other "conditions" such as taking on a demanding job can be overrated.

Still, if someone were to make a compelling case that acquiring or maintaining an addiction is usually unjustified, there may be a way for the circumvention strategy to accommodate this as well. Namely, we might hold an agent negligent or reckless specifically for the wrongdoing of acquiring or maintaining an addiction and significantly risking the outcome of episodic partial normative incapacity. While engaging in acts that bring about episodic partial normative incapacity may not be wrong in every instance or even many instances, the sort of episodic partial incapacity that characterizes addiction may be distinctive such that someone might be accountable for risking it. For instance, addictive partial incapacity can be especially hard to overcome even when an agent recognizes sufficient reasons against substance use. It can also be exceedingly difficult to control the occurrence, timing, or frequency of periods of

---

[177] Pickard (2020).

addictive partial capacity. Further, there is an especially tight causal link between addictive partial incapacity and wrongdoing. Finally, if Wonderly (2021) is right, addictive partial incapacity can involve a compromised capacity for evaluation that resists typical checks and balances.

On such a hybrid account, an addicted agent would be deemed negligent or reckless for voluntarily but unjustifiably acquiring or maintaining an addiction and for the outcome of periodic partial normative incapacity. An agent would be considered negligent or reckless with regard to acquiring or maintaining their addicted condition whether or not their partial incapacity actually results in a wrongdoing. But when that addicted agent engages in a wrongdoing because of their partial incapacity, they would be assessed as partially responsible for that wrongdoing due to considerations about their responsible agency at the time of the wrongdoing. In other words, this hybrid strategy would hold the agent accountable for *two* different wrongdoings related to *two* different outcomes, avoiding the worry about double-counting.

4.1 Conclusion

In this chapter, I have considered three different strategies that attempt to explain responsibility for addiction-explained wrongdoing given the intuition that addiction is not overall a mitigating condition because addicted agents are culpable for acquiring or maintaining their addiction. I have argued that because of the fact that addiction usually only partially incapacitates agents at the time of an addiction-explained wrongdoing, two common strategies for explaining how culpability for an addiction is relevant to assessments of the agent's blameworthiness end up double-counting blameworthiness for a single outcome. I have suggested that the best way to avoid this is to focus assessments on the capacities of the addicted

agent at the time of the addiction-explained wrongdoing. I conclude by allowing the possibility

that the notions of recklessness or negligence can play a role in assessing responsibility for

addiction given that it risks an especially worrisome form of partial incapacity, but I suggest that

this assessment should be distinct from assessing addiction-explained wrongdoing.

Bibliography

Agule, C. K. (2016). "Resisting tracing's siren song," *Journal of Ethics & Social Philosophy* 10 (1): 1-24.

Ahmed, S. H. (2010). "Validation Crisis in Animal Models of Drug Addiction - Beyond Non-Disordered Drug Use Toward Drug Addiction," *Neuroscience & Biobehavioral Reviews* 35(2): 172-184.

Ahmed, S. H. (2012). "The Science of Making Drug-Addicted Animals," *Neuroscience* 211: 107-125.

Ainslie, G. (2000) "A Research-Based Theory of Addictive Motivation," *Law and Philosophy* 19 (1): 77-115.

Ainslie, G. (2001). *Breakdown of Will*. Cambridge University Press.

Ainslie, G. (2013a) "Grasping the Impalpable: The role of Endogenous Reward in Choices, Including Process Addictions," in *Inquiry* 56: 446–69.

Ainslie, G. (2013b) "Money as MacGuffin: a Factor in Gambling and Other Process Addictions," in N. Levy (ed.), *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*. Oxford University Press: 16–37

Ainslie, G. (2017) "Palpating the Elephant: Current Theories of Addiction in the Light of Hyperbolic Delay Discounting" in N. Heather and G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship.* Oxford University Press: 227-244.

Alexander, B. K., Coambs, R. B., & Hadaway, P. F. (1978). "The Effect of Housing and Gender on Morphine Self-Administration in Rats," *Psychopharmacology* 58 (2): 175-179.

Alexander, B. (2018)." Addiction: A Structural Problem of Modern Global Society" in H. Pickard, & S. H. Ahmed, S. H. (eds.), T*he Routledge Handbook of Philosophy and Science of Addiction*. Routledge: 501-510.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.).

Anthony, J.C., Warner, L.A., Kessler, R.C. (1994). "Comparative Epidemiology of Dependence on Tobacco, Alcohol, Controlled Substances, and Inhalants: Basic Findings from the National Comorbidity Survey," *Experimental and Clinical Psychopharmacology* 2: 244–268.

Arpaly, N. (2006). *Merit, Meaning, and Human Bondage: An Essay on Free Will.* Princeton University Press.

Arpaly, N., & Schroeder, T. (2014). *In Praise of Desire*. Oxford University Press.

Bechara, A. (2005). "Decision Making, Impulse Control and Loss of Willpower to Resist Drugs: A Neurocognitive Perspective," *Nature Neuroscience* 8 (11): 1458-1463.

Bermúdez, J. P. (2021). "The Skill of Self-Control," *Synthese:* p1-23.

Berridge, K. C., & Robinson T. E. (1998). "What is the Role of Dopamine in Reward: Hedonic Impact, Reward Learning, or Incentive Salience?" *Brain Research Reviews* 28 (3): 309-369.

Berridge, K. C., & Robinson, T. E. (2011), "Drug Addiction as Incentive Sensitization" in J. Poland and G. Graham (eds.) *Addiction and Responsibility.* MIT Press*: p21-54.

Berridge, K. C., & Robinson, T. E. (2016). "Liking, Wanting and the Incentive-Sensitization Theory of Addiction," *The American Psychologist* 71 (8): p670-679.

Berridge, K. C., Robinson, T. E. & Aldridge, J.W. (2009). "Dissecting Components of Reward: 'Liking', 'Wanting', and Learning," *Current Opinion in Pharmacology* 9: 65–73.

Bjork, J. M., Momenan, R. & Hommer, D. W. (2009). "Delay Discounting Correlates with Proportional Lateral Frontal Cortex Volumes," *Biological Psychiatry* 65: 710–713.

Blum K., Noble, E.P., Sheridan, P.J., Montgomery, A., Ritchie, T., Jagadeeswaran P., Nogami, H., Briggs, A. H., Cohn, J. B. (1990). "Allelic Association of Human Dopamine D2 Receptor Gene in Alcoholism," *Journal of the American Medical Association* 263: 2055–60.

Bratman, M. E. (1997). "Responsibility and Planning" *The Journal of Ethics* 1(1): 27-43.

Bratman, M. E. (2004). "Three Theories of Self-governance" in *Philosophical Topics* 32 (½): 21-46.

Brink, D. O. (2021). *Fair Opportunity and Responsibility*. Oxford University Press.

Brink, D. O., & Nelkin, D. (2013). "Fairness and the Architecture of Responsibility" in D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility Vol 1*. Oxford University Press: 13-132.

Burgess, B. (2016). "How a Stigmatic Structure Enslaves Addicts" in N. Heather and G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship.* Oxford University Press**:** 409-425.

Butlin, P. and Papineau, D. (2016). "Normal and Addictive Desires" in N. Heather and G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship.* Oxford University Press: 99-115.

Carr, D. (2008). *The Night of the Gun.* Simon and Schuster.

Carter, A., & Hall, W. (2013). "Ethical Implications of Research on Craving" in *Addictive Behaviors* 38(2): 1593-1599.

Clarke, R. (2009). "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism," *Mind* 118 (470): 323-351.

Clarke, R. (2014). *Omissions: Agency, Metaphysics, and Responsibility.* Oxford University Press.

Clarke, R. (2015). "Abilities to Act," *Philosophy Compass* 10(12): 893-904.

Coates, D. J., & Swenson, P. (2013). "Reasons-Responsiveness and Degrees of Responsibility, *Philosophical Studies* 165(2): 629-645.

Coates, D. J., & Tognazzini, N. A. (2013). "The Contours of Blame" in D. J. Coates & N.A. Tognazzini (eds.) *Blame: Its Nature and Norms*. Oxford University Press: 3-26.

Cross, T. (2012). "Recent Work on Dispositions" *Analysis* 72 (1): 115–124.

Crowley, D. M., Connell, C. M., Jones, D., & Donovan, M. W. (2019). "Considering the Child Welfare System Burden from Opioid Misuse: Research Priorities for Estimating Public Costs," *American Journal of Managed Care* 25(13): S256-S263.

Davies, J. B. (2013). *The Myth of Addiction: An Application of the Psychological Theory of Attribution to Illicit Drug Use.* Routledge.

Dill, B., & Holton, R. (2014). "The Addict in Us All," *Frontiers in Psychiatry* 5: 39.

Elster, J. (ed.) (1999). *Addiction: Entries and Exits.* Russell Sage Foundation.

Field, M., & Cox, W. M. (2008). "Attentional Bias in Addictive Behaviors: A Review of Its Development, Causes, and Consequences," *Drug and Alcohol Dependence* 97(1): 1-20.

Field, M., Mogg, K., & Bradley, B. P. (2006). "Attention to Drug-Related Cues in Drug Abuse and Addiction: Component Processes" in R. W. Wiers & A. W. Stacy (eds.) *Handbook of Implicit Cognition and Addiction:* 151-163.

Field, M., Munafò, M. R., & Franken, I. H. (2009). "A Meta-analytic Investigation of the Relationship Between Attentional Bias and Subjective Craving in Substance Abuse, *Psychological Bulletin* 135(4): 589.

Fischer, J. M. (1987)." Responsiveness and Moral Responsibility" in F. Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology.* Cambridge University Press: 81-106.

Fischer, J. M. (2012). "Semicompatibilism and Its Rivals," *The Journal of Ethics*, 16(2), 117-143.

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility.* Cambridge University Press.

Fischer, J. M., & Tognazzini, N. A. (2011). "The Physiognomy of Responsibility," *Philosophy and Phenomenological Research* 82(2): 381-417.

Flanagan, O. (2011). "What Is It Like to Be an Addict," in J. Poland & G. Graham (eds.), *Addiction and Responsibility*. MIT Press: 269-292.

Flanagan, O. (2013a). "Identity and Addiction: What Alcoholic Memoirs Teach" in K. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, & T. Thornton (eds.), *Oxford Handbook of Philosophy of Psychiatry.* Oxford University Press: p. 865-888.

Flanagan, O. (2013b). "The Shame of Addiction," *Frontiers in Psychiatry* 4 120: 142-152.

Flanagan, O. (2017). "Willing Addicts? Drinkers, Dandies, Druggies, and Other Dionysians" in N. Heather and G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship.* Oxford University Press: 66-81.

Flanagan, O. (2018) "Identity and Addiction" in H. Pickard, & S. H. Ahmed, S. H. (eds.), T*he Routledge Handbook of Philosophy and Science of Addiction*. Routledge: 77-89.

Franken, I. H. (2003). "Drug Craving and Addiction: Integrating Psychological and Neuropsychopharmacological Approaches," *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 27 (4): 563-579.

Frankfurt, H. (1971). "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68 (1): 5-20.

Germano, B. (2016, 23 Sept). *Mom Who Overdosed In Front Of Daughter Admits She Hit Rock Bottom.* CBS WBZ Boston*.* http://boston.cbslocal.com/2016/09/23/mother-overdose-fentanyl-lawrence-caught-on-camera-recovery/

Ginley, M. K., Pfund, R. A., Rash, C. J., & Zajac, K. (2021). "Long-term Efficacy of Contingency Management Treatment Based on Objective Indicators of Abstinence from Illicit Substance Use Up to 1 Year Following Treatment: A Meta-analysis," *Journal of Consulting and Clinical Psychology* 89(1): 58-71.

Glannon, W. (2009). "Responsibility and Priority in Liver Transplantation," *Cambridge Quarterly of Healthcare Ethics* 18(1): 23-35.

Glover, J. (1970). *Responsibility*. Routledge.

Goldstein, R.Z., & Volkow, N.D. (2002). "Drug Addiction and its Underlying Neurobiological Basis: Neuroimaging Evidence for the Involvement of the Frontal Cortex," *American Journal of Psychiatry* 159: 1642-1652.

Goodnough, A. (2020, October 27). *This Addiction Treatment Works. Why Is It So Underused.* The New York Times. www.nytimes.com/2020/10/27/health/meth-addiction-treatment.html.

Graham, M. D., Young, R. A., Valach, L., & Wood, A. R. (2008). "Addiction as a Complex Social Process: An Action Theoretical Perspective," *Addiction Research & Theory* 16(2): 121-133.

Gupta, S. (2019, March 12). *Earning Prizes for Fighting an Addiction*. The New York Times. www.nytimes.com/2019/03/12/opinion/earning-prizes-for-fighting-an-addiction.html

Haden, P. (2017, August 15). *People With Addictions Lured To Bogus Rehab Treatment Via 'Body Brokers.'* National Public Radio. www.npr.org/sections/health-shots/2017/08/15/542630442/body-brokers-get-kickbacks-to-lure-people-with-addictions-to-bad-rehab

Haji, I. (1998). *Moral Appraisability: Puzzles, Proposals, and Perplexities*. Oxford University Press.

Hart, H.L.A. (1968). *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford University Press.

Harvard Mental Health Letter (2011) "How Addiction Hijacks the Brain. Desire Initiates the Process, but Learning Sustains It," *The Harvard Mental Health Letter.* 28(1):1–3.

Henden, E. (2018)." Addiction as a Disorder of Self-Control" in Pickard, H., & Ahmed, S. H. (eds). *The Routledge Handbook of Philosophy and Science of Addiction.* Routledge: 45-53.

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

Hiroi, N. & Agatsuma, S. (2005). "Genetic Susceptibility to Substance Dependence," Molecular Psychiatry 10: 336–344.

Hodges, Benjamin. (2018)." A Life Worth Fighting For," *Georgia State University Magazine* Quarter 2: 22-27.

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

Holton, R (2016). "Addiction, Self-Signalling and the Deep Self," *Mind & Language*. 31(3): 300-313.

Holton, R., & Berridge, K. (2013). "Addiction Between Compulsion and Choice" in N. Levy (ed.) *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience.* Oxford University Press: 239-268.

Holton, R., & Schute (2007). "Self-Control In the Modern Provocation Defense," *Oxford Journal of Legal Studies* 27(1): 49-73.

Husak, D. (1999). "Addiction and Criminal Liability," *Law and Philosophy* 18(6): 655-684.

Husak, D. (2000). "Liberal Neutrality, Autonomy, and Drug Prohibitions," *Philosophy & Public Affairs* 29(1): 43-80

Husak, D. (2004). "The Moral Relevance of Addiction," *Substance Use & Misuse* 39(3): 399-436.

Hyman, S. E. (2007). "The Neurobiology of Addiction: Implications for Voluntary Control of Behavior," *The American Journal of Bioethics* 7(1): 8-11.

Jaworska, A. (2017). "Identificationist Views" in K. Timpe, M. Griffith, N. Levy (eds.) *The Routledge Companion to Free Will.* Routledge: 15-26.

Kalant, H. (2010). "What Neurobiology Cannot Tell Us about Addiction,"*Addiction* 105(5): 780-789.

Kalivas, P. W., & Volkow, N. D. (2005). "The Neural Basis of Addiction: a Pathology of Motivation and Choice," *American Journal of Psychiatry* 162(8): 1403-1413.

Kennett J. (2013). "Addiction, Choice, and Disease: How Voluntary is Voluntary Action in Addiction?" in N. Vincent (ed.) *Neuroscience and Legal Responsibility.* Oxford University Press: 257–78.

Kennett, J. (2014). "Just Say No? Addiction and the Elements of Self-Control" in N. Levy (ed.) *Addiction and Self-Control: Perspectives From Philosophy, Psychology, and Neuroscience.* Oxford University Press: 144-164.

Kennett, J., & McConnell, D. (2013). "Explaining Addiction: How Far Does the Reward Account of Motivation Take Us?" *Inquiry* 56(5): 470-489.

Kennett, J., & Smith, M. (1996). "Frog and Toad Lose Control," *Analysis*, 56(2): 63-73.

King, M. (2014). "Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility," *European Journal of Philosophy* 22(3): 463-482.

Koob, G. F. (2013). "Negative Reinforcement in Drug Addiction: The Darkness Within," *Current Opinion in Neurobiology* 23(4): 559-563.

Koob, G. F. (2017). "Antireward, Compulsivity, and Addiction: Seminal Contributions of Dr. Athina Markou to Motivational Dysregulation in Addiction," *Psychopharmacology:* 1-18.

Koob, G. F., & Le Moal, M. (2005). "Plasticity of Reward Neurocircuitry and the 'Dark Side' of Drug Addiction," *Nature Neuroscience* 8(11): 1442-1444.

La Gorce, T. (2021, May 14). *On Parole, Staying Free Means Staying Clean and Sober.* New York Times. www.nytimes.com/2021/05/14/opinion/parole-support-program.html

Leon, M. (2001)." The Willing Addict–Actor or (Helpless) Bystander?" *Philosophia*, 28(1-4): 437-443.

Leshner, A. I. (1997). "Addiction is a Brain Disease, and It Matters," *Science* 278(5335): 45-47.

Levy, N. (2007). "The Social: A Missing Term in the Debate Over Addiction and Voluntary Control," *The American Journal of Bioethics* 7(1): 35-36.

Levy, N. (2012). "Autonomy, Responsibility and the Oscillation of Preference" in A. Carter, W. Hall, and J. Illes (eds.) *Addiction Neuroethics: The Ethics of Addiction Neuroscience Research and Treatment.* Academic Press: 139-151.

Lewis, M. (2011). "Dopamine and the Neural 'Now': Essay and Review of 'Addiction: A Disorder of Choice,'" *Perspectives on Psychological Science* 6(2): 150-155.

Lewis, M. (2017). "Addiction and the Brain: Development, Not Disease," *Neuroethics* 10(1): 7-18.

Loewenstein, G. (1999). "A Visceral Theory of Addiction" in Elster, J., & Skog, O. J. (Eds) *Getting Hooked: Rationality and Addiction.* Cambridge University Press: 235-264.

Lussier, J. P., Heil, S. H., Mongeon, J. A., Badger, G. J., Higgins, S. T. (2006). "A Meta-Analysis of Voucher-Based Reinforcement Therapy for Substance Use Disorders" *Addiction* 101(2): 192-203.

Mann, B. (2021, February 15). *As Addiction Death Surge, Profit-Driven Rehab Industry Faces 'Severe Ethical Crisis.'* National Public Radio. www.npr.org/2021/02/15/963700736/as-addiction-deaths-surge-profit-driven-rehab-industry-faces-severe-ethical-cris

Matthews, S., Dwyer, R., & Snoek, A. (2017). "Stigma and Self-stigma in Addiction," *Journal of Bioethical Inquiry* 14(2): 275-286.

McClure, S.M., Laibson, D.I., Loewenstein, G.F. and Cohen, J.D. (2004). "Separate Neural Systems Value iImmediate and Delayed Monetary Rewards," *Science* 306: 503–507

McConnell, D. (2016). "Narrative Self-constitution and Recovery from Addiction," *American Philosophical Quarterly* 53(3): p307-322.

McKenna, M. (2001). "Responsibility and Control: A Theory of Moral Responsibility," *The Journal of Philosophy* 98(2): p93-100.

McKenna, M. (2013). "Reasons-responsiveness, Agents, and Mechanisms" in D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility Vol 1.* Oxford University Press: p151-83.

McKenna, M., & Pereboom, D. (2016). *Free will: A Contemporary Introduction.* Routledge.

McKenna, M., & Van Schoelandt, C. (2015) "Crossing a Mesh Theory with a Reasons-Responsive Theory: Unholy Spawn of an Impending Apocalypse or Love Child of a New Dawn?" in eds. A. Buckareff, C. Moya, S. Rosell (eds.). *Agency, Freedom, and Moral Responsibility.* Palgrave Macmillan: 44-64.

Mele, A. R. (1990). "Irresistible Desires," *Noûs.* 24(3): 455-472.

Mele, A. R. (1996). "Addiction and Self-Control," *Behavior and Philosophy* 24(2): 99-117.

Mele, A. R. (2002). "Akratics and Addicts," *American Philosophical Quarterly* 39(2): 153-167.

Mele, A. R. (2004). "Action, Volitional Disorder and Addiction" in J. Radden (ed.) *The Philosophy of Psychiatry: A Companion*. Oxford University Press.

Mele, A. (2018). "Exercising Self-Control: An Apparent Problem Resolved" in J. L. Bermúdez (Ed.), *Self-control, Decision Theory, and Rationality: New essays.* Cambridge University Press: 204–217.

Moore, M. S. (1990). "Choice, Character, and Excuse," *Social Philosophy and Policy* 7(02): p29-58.

Moore, M. S. (2010). *Placing Blame: A Theory of the Criminal Law*. Oxford University Press.

Moore, M. S. (2020) "Addiction, Responsibility, and Neuroscience," *University of Illinois Law Review* 2: 375-470.

Nelkin, D. K. (2011). *Making Sense of Freedom and Responsibility*. Oxford University Press.

Nelkin, D. K. (2013). "Desert, Fairness, and Resentment," *Philosophical Explorations*, 16(2): 117-132.

Nelkin, D. K. (2015). "Psychopaths, Incorrigible Racists, and the Faces of Responsibility," *Ethics*, 125(2): 357-390.

Nelkin, D. K. (2016). "Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness," *Noûs* 50(2): 356-378.

Nelkin, D. K. (2016). "Blame" in K. Timpe, M. Griffith, N. Levy (eds.) *The Routledge Companion to Free Will.* Routledge: 623-634.

Nelkin, D. K., & Rickless, S. C. (2017). "Moral Responsibility for Unwitting Omissions: A New Tracing View" in D. K. Nelkin and S. C. Rickless (eds.) *The Ethics and Law of Omissions:* p106-129.

Nestler, E.J. (2005) "Is There a Common Molecular Pathway for Addiction?" in *Nature Neuroscience* 8: 1445-1449.

Noggle, R. (2016). "Addiction, Compulsion, and Persistent Temptation," *Neuroethics* 9(3): 213-223.

Pattani, A. (2021, November 8). *DEA Takes Aggressive Stance Toward Pharmacies Trying to Dispense Addiction Medicine*. National Public Radio. https://www.npr.org/sections/health-shots/2021/11/08/1053579556/dea-suboxone-subutex-pharmacies-addiction

Percy, J. (2018, 10 Oct 2018) *Trapped by the 'Walmart of Heroin.'* The New York Times Magazine. www.nytimes.com/2018/10/10/magazine/kensington-heroin-opioid-philadelphia.html.

Pickard, H. (2012). "The Purpose in Chronic Addiction," *The American Journal of Bioethics-Neuroscience* 3(2): 40-49.

Pickard, H. (2015). "Psychopathology and the Ability to Do Otherwise," *Philosophy and Phenomenological Research* 90(1): 135-163.

Pickard, H. (2017). "Responsibility Without Blame for Addiction," *Neuroethics:* 1-12.

Pickard, Hanna (2018)." The Puzzle of Addiction" in H. Pickard & S. H. Ahmed (eds.) *The Routledge Handbook of Philosophy and Science of Addiction*. Routledge: 501-510.

Pickard, H. (2020). "Addiction and the Self," *Noûs*: 1-25.

Pober, J. M. (2012). "Is Addiction a Heterogeneous Condition? Reflections on Pickard's 'The Purpose in Chronic Addiction,'" *American Journal of Bioethics-Neuroscience* 3(2): 52-54.

Pober, J. M. (2013). "Addiction is Not a Natural Kind," *Frontiers in Psychiatry* 4 123: 1-11.

Racine, E., & Barned, C. (2019). "Addiction and Voluntariness: Five "Challenges" to Address in Moving the Discussion Forward," *Cambridge Quarterly of Healthcare Ethics*, 28(4): 677-694.

Robinson, M. J., Robinson, T. E., & Berridge, K. C. (2018). "The Current Status of the Incentive Sensitization Theory of Addiction" in H. Pickard & S. H. Ahmed (eds.) *The Routledge Handbook of Philosophy and Science of Addiction*. Routledge: p351-361.

Robinson, T. E., & Berridge, K. C. (1993) "The Neural Basis of Drug Craving: an Incentive-Sensitization Theory of Addiction," *Brain Research Reviews* 18(3) p247-291.

Robinson, T. E., & Berridge, K. C. (2000). "The Psychology and Neurobiology of Addiction: An Incentive-Sensitization View," *Addiction* 95: 91–117.

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

Robinson, T.E. and Berridge, K.C. (2008). "The Incentive Sensitization Theory of Addiction: Some Current Issues," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1507): 3137-3146.

Scanlon, T. (2008). *Moral Dimensions: Permissibility, Meaning, Blame.* Harvard University Press.

Schroeder, T. (2010) "Irrational Action and Addiction" in *What is Addiction* eds. D. Ross, H. Kincaid, D. Spurrett, and P. Collins. MIT Press: p391-407.

Schroeder, T. & Arpaly, A. (2013) "Addiction and Blameworthiness" in N. Levy (ed.) *Addiction and Self-Control: Perspectives From Philosophy, Psychology, and Neuroscience*. Oxford: 214-238.

Searby, A., Maude, P., & McGrath, I. (2015). "Maturing Out, Natural Recovery, and Dual Diagnosis: What are the implications for older adult mental health services? *International Journal of Mental Health Nursing* 24(6): 478-484.

Seelye, K.Q. (21 Jan 2018). *1 Son, 4 Overdoses, 6 Hours*. The New York Times. www.nytimes.com/2018/01/21/us/opioid-addiction-treatment-families.html.

Seelye, K.Q. (28 Sept 2016). *Addicted Parents Get Their Fix, Even With Children Watching.* The New York Times. www.nytimes.com/2016/09/28/us/addicted-parents-get-their-fix-even-with-children-watching.html

Seelye, K.Q. Turkewitz, J., Healy, J., & Blinder, A. (2018, December 11). *How Do You Recover After Millions Have Watched You Overdose?* The New York Times. www.nytimes.com/2018/12/11/us/overdoses-youtube-opioids-drugs.html.

Sher, G. (2006). *In Praise of Blame*. Oxford University Press.

Shoemaker, D. (2011). "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics* 121(3): 602-632.

Shoemaker, D. (2015). *Responsibility from the Margins*. Oxford University Press.

Sinnott-Armstrong, W., & Pickard, H. (2013). "What is Addiction" in K. Fulford, M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, & T. Thornton (eds.), *Oxford Handbook of Philosophy of Psychiatry.* Oxford University Press: 851-64.

Sinnott-Armstrong, W. & Summers, J. (2018) "Defining Addiction: A Pragmatic Perspective" in H. Pickard, & S. H. Ahmed, S. H. (eds.), T*he Routledge Handbook of Philosophy and Science of Addiction*. Routledge: p123-131.

Skinner, M. D., & Aubin, H. J. (2010). "Craving's Place in Addiction Theory: Contributions of the Major Models," *Neuroscience & Biobehavioral Reviews* 34(4): 606-623.

Smith, A. M. (2000). "Identification and Responsibility," in van den Beld, A. (ed.) *Moral Responsibility and Ontology.* Kluwer Academic Publishers: 233-246.

Smith, A. M. (2004). "Conflicting Attitudes, Moral Agency, and Conceptions of the Self," *Philosophical Topics* 32(1/2): 331-352.

Smith, A. M. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115(2): 236-271.

Smith, A. M. (2007). "On Being Responsible and Holding Responsible," *The Journal of Ethics* 11 (4): 465-484.

Smith, A. M. (2008). "Control, Responsibility, and Moral Assessment," *Philosophical Studies* 138 (3): 367-392.

Smith, A. M. (2012). "Attributability, Answerability, and Accountability: In Defense of a Unified Account,"*Ethics* 122(3): 575-589.

Smith, A. M. (2013)." Moral Blame and Moral Protest" in D.J. Coates and N.A. Tognazzini (eds), *Blame: Its Nature and Norms.* Oxford University Press: 27–47.

Smith, M. (2003) "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion" in S. Stroud & C. Tappolet (eds) *Weakness of Will and Practical Irrationality.* Oxford University Press: 17-38.

Snoek, A., Levy, N. & Kennett, J. (2016). "Strong-willed but Not Successful: The Importance of Strategies in Recovery from Addiction," *Addictive Behaviors Reports* 4: 102-107.

Snoek, A., McGeer, V., Brandenburg, D., & Kennett, J. (2021). "Managing Shame and Guilt in Addiction: A Pathway to Recovery" in *Addictive Behaviors* 120 106954: 1-10.

Snoek, A., & Matthews, S. (2017). "Introduction: testing and refining Marc Lewis's critique of the brain disease model of addiction" in *Neuroethics* Vol. 10 No. 1: p1-6.

Sripada, C. (2014). "The Second Hit in Addiction" in W. Sinnott-Armstrong (ed) *Moral Psychology Volume 4: Free Will and Moral Responsibility.* MIT Press: 295–304.

Sripada, C. (2015)." Moral Responsibility, Reasons, and the Self" in D. Shoemaker (ed). *Oxford Studies in Agency and Responsibility Vol. 3.* Oxford University Press: 242-264

Sripada, C. (2016). "Self-expression: A Deep Self Theory of Moral Responsibility," *Philosophical Studies.* 173(5): 1203-1232.

Sripada, C. (2017). "Frankfurt's Unwilling and Willing Addicts," *Mind* 126(503): 781–815.

Sripada, C. (2018). "Addiction and Fallibility" in *The Journal of Philosophy* 115(11): 569-587.

Sripada, C. (2019) "The Fallibility Paradox," *Social Philosophy and Policy* 36(1): 234-248.

Sripada, C. (2022). "Impaired Control in Addiction Involves Cognitive Distortions and Unreliable Self-control, Not Compulsive Desires and Overwhelmed Self-control. *Behavioural Brain Research* 418(113639): 1-9.

Strawson, P. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48: 187-211.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2003) Results from the 2002 National Survey on Drug Use and Health: National Findings NHSDA Series H-22, DHHS Publication No. SMA 03-3836. Office of the Applied Studies.

Summers, J. S. (2015). "What is Wrong with Addiction," *Philosophy, Psychiatry, & Psychology* 22(1): 25-40.

Talbert, M. (2008). "Implanted Desires, Self-formation and Blame," *Journal of Ethics & Social Philosophy* 3(2): 1-18.

Talbert, M. (2016). *Moral Responsibility: An Introduction*. John Wiley & Sons.

Talbert, Matthew (2019, Winter) "Moral Responsibility" in Zalta, E. N. (ed.) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/

Vargas, M. (2005)." The Trouble with Tracing," Midwest Studies in Philosophy XXIX: 269-91.

Vargas, M. (2013). *Building Better Beings*. Oxford University Press.

Vergés, A., Haeny, A.M., Jackson, K.M., Bucholz, K.K., Grant, J.D., Trull, T.J., Wood, P.K. & Sher, K.J. (2013). "Refining the Notion of Maturing Out: Results from the National Epidemiologic Survey on Alcohol and Related Condition," *American Journal of Public Health*, 103(12): e67-e73.

Volkow & Fowler (2000)." Addiction, a Disease of Compulsion and Drive: Involvement of the Orbitofrontal Cortex," *Cerebral Cortex* 10: 318-325.

Volkow, N.D., Koob, G. F., and McLellan, A. T. (2016). "Neurobiologic Advances from the Brain Disease Model of Addiction," *New England Journal of Medicine* 374: 363–371.

Vowles, K., McEntee, M., Julnes, P., Frohe, T., Ney, J., van der Goes, D. (2015)." Rates of Opioid Misuse, Abuse, and Addiction in Chronic Pain" in *Pain* 156(4): 569-576.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments.* Harvard University Press.

Wallace, R. J. (1999). "Addiction as a Defect of the Will: Some Philosophical Reflections," *Law and Philosophy* 18(6): 621-654.

Wallace, R. J. (2011). "Dispassionate Opprobrium: On Blame and the Reactive Sentiments" in R. Jay Wallace, R. Kumar, & S. Freeman (eds) *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford University Press: 106-128.

Watson, G. (1975). "Free Agency," *The Journal of Philosophy:* 205-220.

Watson, G. (1977). "Skepticism About Weakness of Will," *The Philosophical Review* 86(3): 316-339.

Watson, G. (1987) "Free Action and Free Will" in *Mind* 96(382): 145-172.

Watson, G. (1996). "Two Faces of Responsibility," *Philosophical Topics*, 24(2): 227-248

Watson, G. (1999a). "Disordered Appetites: Addiction, Compulsion, and Dependence" in J. Elster (ed.) *Addiction: Entries and exits.* Russell Sage Foundation: 3-28.

Watson, G. (1999b). "Excusing Addiction," *Law and Philosophy* 18(6): 589-619.

Watson, G. (2001) "Reasons and Responsibility," *Ethics* 111: 374 –394

Watson, G. (2011). "The Trouble With Psychopaths" in R. Jay Wallace, R. Kumar, & S. Freeman (eds.) *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford University Press: 275-300.

Westervelt, E. (2017, October 5). *To Save Opioid Addicts, This Experimental Court Is Ditching The Delays*. NPR. www.npr.org/sections/health-shots/2017/10/05/553830794/to-save-opioid-addicts-this-experimental-court-is-ditching-the-delays.

Whittle, A. (2010). "Dispositional Abilities," *Philosophers' Imprint* 10 (12): 1-23.

Winick, C. (1962). "Maturing Out of Narcotic Addiction," *Bulletin on Narcotics* 14: 1-7.

Wolf, S. (1993). *Freedom Within Reason*. Oxford University Press.

Wonderly, M. (2021). "Attachment, Addiction, and Vices of Valuing" in E. Harcourt (ed.) *Attachment and Character: Attachment Theory, Ethics, and the Developmental Psychology of Vice and Virtue.* Oxford University Press*: 224-245.

Woolsey, A. & Willingham, A.J. (2016, 23 September). *A Mom Apparently Overdoses next to Her Child -- and Why Police Want You to Watch.* CNN. www.cnn.com/2016/09/23/health/heroin-overdose-video-massachusetts-trnd/index.html.

Yaffe, G. (2011). "Lowering the Bar for Addicts" in G. Graham and J. Poland (eds.) *Addiction and Responsibility.* MIT Press*: 113-138.

Yoon, Y.& Chen, C. (2018, April). *Liver Cirrhosis Mortality in the United States: National, State, and Regional Trends, 2000-2015.* National Institute on Alcohol Abuse and Alcoholism, U.S. Department of Health and Human Services. NIH. https://pubs.niaaa.nih.gov/publications/surveillance111/Cirr15.htm.

Yoon, J. H., Suchting, R., de Dios, C., Vincent, J. N., McKay, S. A., Lane, S. D., & Schmitz, J. M. (2021). "Decreased Cocaine Demand Following Contingency Management Treatment," *Drug and Alcohol Dependence* 226(108883): 1-10.

Zimmerman, M. J. (1988). *An Essay on Moral Responsibility*. Rowman and Littlefield.