Narrative, self-governance, and addiction

Doug McConnell MA BSc

Macquarie University

Philosophy Dept.

Submitted for the degree of Doctor of Philosophy

April 11th 2014

Contents

Abstract

Addiction has long inflicted heavy costs on individuals and society yet we still lack an account of agency that can satisfactorily explain addicted behavior. This thesis attempts to develop such an account by drawing on theories of narrative self-constitution. I argue that a narrative account improves on the reward maximisation views promoted by Gene Heyman and George Ainslie and the normative planning views promoted by Richard Holton and Michael Bratman.

Heyman and Ainslie define all action, including addicted action, as reward maximizing where rewards are fixed by extra-agential forces. This account eliminates both the possibility of a synchronic struggle against addictive desire and the possibility of *self*-governance in general. Therefore, it clashes with the experience of addicts and clinicians who see recovery as an extended agential struggle. Normative planning theories improve on these accounts by making room for self-governance. According to these views, self-governance varies according to how much effort the agent puts into conforming to norms of practical reason as they form and enact their plans and policies. However, there remain a variety of agential phenomena that are mysterious on normative planning accounts. The most glaring cases are where addicts recover despite there being no improvement in their planning skills or circumstances and where addicts fail to recover despite disvaluing their lifestyle and having the planning skills to pursue available alternatives.

The narrative account defended in this thesis builds on normative planning accounts by showing how networks of intentions are nested within more holistic self-narratives that include interpretations of one's contingent circumstances. The way agents self-narrate, therefore, affects their self-governance in ways that go beyond mere normative organization of intentions. As a consequence, certain styles of self-narration can entrench addiction or facilitate recovery somewhat independently of one's values, planning skills, and available opportunities.

Statement of Candidate

I certify that the work in this thesis entitled “Narrative, self-governance and addiction” has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Doug McConnell

Student No. 41959132

11/4/2014

Acknowledgements

This work could not have been completed without the guidance and support of many people. Chief among them is Jeanette Kennett, my supervisor, who kept the project on track throughout. Jeanette and Catriona Mackenzie, my co-supervisor, always had time to discuss ideas and they painstakingly reviewed countless drafts without complaint. Thanks also to my adjunct supervisor in Melbourne, Craig Fry, who was always on hand to provide an alternative perspective grounded in the practical concerns of public health.

I have been lucky to be part of the research team working on the project, 'Addiction, moral identity and moral agency.' That team includes Jeanette and Craig along with Steve Matthews, Anke Snoek, Robyn Dwyer, and Merle Spriggs. The group continues to be a great environment to test and generate new ideas. Thanks also to the interviewees who shared their stories with us and the treatment facilities who accommodated our research. The funding for that project, including my PhD, was gratefully received from the Australian Research Council.

I enjoyed a month in Chicago working with Marya Schechtman who made me feel at home and enthusiastically engaged with my project. Thanks also to the University of Illinois at Chicago graduate students, particularly Aleks Zarnitsyn and Zac Harmon who took the time to show me around. The Postgraduate Research Fund from Macquarie University made that trip possible for which I am grateful.

The following have also gone out of their way to provide helpful feedback on my work or discuss related issues: Dave Ripley, John Davenport, Fred Wertz, and Bojun Hu. Thank you.

Free accommodation services around the world were gratefully received from the following: Ean, Val and Jane, Anke and Casper, Adrian, Muthu and Asa, Gabs and Brendan, Robbie "The Coach" Petrie, and Camilla Hansen.

Finally, some more sentimental but no less important acknowledgements. Thanks to Grant Gillett for nurturing my philosophical interests and pointing them in a meaningful direction. Thanks to the Ayers family and my own family for putting up with me and, most of all, thanks to Katie, for everything.

Introduction

The goal of this thesis is to demonstrate that a complete understanding of agency must be informed by the motivational effects of agents' self-narratives which may undermine or support self-governance. My claim is that agents may suffer from poor self-governance, in part, because of their self-narrative and through re-narration their self-governance might improve. In making this claim, I aim to contribute to the philosophy of action. Part of that contribution is to support accounts of narrative self-constitution in general, such as that of Marya Schechtman (1996, 2007) and David Velleman (1989, 2005), however there are two novel aspects to this contribution. First, I develop a detailed account of exactly how narrative theories of agency provide more explanatory power than non-narrative theories, such as that of Michael Bratman (1999, 2007). Second, my argument that self-narratives can undermine self-governance provides a novel line of argument in favour of Christman's (2009) claim that enacting an alienating self-narrative is insufficient for authentic self-governance or what some call autonomy.

The motivation to connect this philosophical project with models of addiction is that, despite a recent surge of philosophical work on addiction, nobody has yet applied a narrative approach to the area. I argue that a narrative account of agency provides a way to understand a range of addicted behaviour that other accounts cannot. Those cases are where the agent's self-narrative entrenches addiction and where changes to self-narrative aid recovery. Addiction provides an especially good testing ground for theories of agency. We can often learn about a system by considering cases where it has gone wrong. If something has gone wrong with agency in cases of addiction, and I argue that it has, then our theory of agency needs to be able to specify what addicted agency lacks that normal agency does not. The narrative account performs better on this testing ground than the other theories of agency I consider.

The costs of addiction

There is, of course, a clear motivation to address the problem of addiction independent of any interest in the philosophy of action. Addiction is extremely costly to society, individuals and their families. To get an idea of the scale of the problem, the National Institute on Drug Abuse (NIDA) estimates that drug abuse and dependence on tobacco, alcohol and illegal drugs costs the United States (US) over $600 billion annually (2012, p. 13). This includes the costs of health care, lost productivity, crime, incarceration and law enforcement. In the 2012 National Survey on Drug Use and Health in the US, 8.5% of the population aged twelve years or older were classified as being substance dependent or abusers of alcohol and/or illicit drugs in the past

year.[1] Overall, 17.7 million people had alcohol dependence or abuse, and 7.3 million people had illicit drug dependence or abuse (Substance Abuse and Mental Health Services Administration, 2013). Tobacco smoking causes 20% of American deaths annually and 16 million Americans suffer from a disease caused by smoking (U.S. Department of Health and Human Services, 2014). So it is not surprising that, of adult smokers, 69% report wanting to quit and 43% have tried to quit in the last year (U.S. Department of Health and Human Services, 2014).

The costs of drug abuse and dependence are clearly not isolated to the United States. In 2009, the UN Office on Drugs and Crime estimated that, worldwide, 15-39 million people are injecting drug-users or problem users of opioids, cocaine, or amphetamines (Degenhardt & Hall, 2012). The World Health Organization estimates that, in 2004, 2.5 million people died from the harmful use of alcohol, i.e. 3.8% of all deaths in the world that year. If we subtract deaths caused by acute intoxication, i.e. intentional and unintentional injuries, that still leaves 58% caused by diseases attributable to alcoholism, such as, cirrhosis and cancer of the liver, oesophageal cancer, hypertension and cerebrovascular disease (World Health Organization, 2011). Whatever the exact costs of addiction there seems no doubt that they are large and widespread. However, one might object that, although the costs are real enough, it is debatable what role 'addiction' plays.

What is addiction?

Some experts would say that most of the above costs are the result of addiction while others would say that none of them are. Experts can take such disparate positions because they disagree on the nature of addiction. The fundamental point of dispute is whether addiction involves a characteristic agential impairment. The dominant position of those who think it does involve an agential impairment is the 'brain disease model.' According to this model addiction is a chronic, often relapsing, brain disease that causes compulsive drug use despite negative consequences (Baler & Volkow, 2006; Hyman, 2005; Leshner, 1997; Volkow & Li, 2005). Drug-related stimuli can "hijack the top down goal-driven cognitive resources needed for the normal operation of the reflective system" (Bechara, 2005, p. 1461). Proponents of the 'brain disease model' consider most of the costs outlined above to be attributable to addiction because it is a disease that causes drug-users to behave as they do.

[1] Including the use of inhalants and the nonmedical use of prescription-type psychotherapeutic drugs.

Other experts, and probably most lay people, claim that chronic drug-use is the result of somewhat unusual preferences but perfectly normal agency. Just because drug-use costs society money, drug-users die prematurely, and clinicians classify many people as having substance use disorders, that does not mean these people are not doing what they most want to do. On this view, 'addiction' has no physiological existence; it is merely a coercive social label designed to encourage people to pursue more mainstream preferences. Those who think that this coercion is justified support the 'moral model' while those who think there is nothing inherently wrong with drug-use preferences support a 'liberal model' (Foddy & Savulescu, 2010). Therefore, this group believe that the costs of drug use are mainly caused by people doing what they most want while others try to prevent them. However, both the brain disease model and the moral/liberal models struggle to accommodate significant evidence.

Problems for the brain disease model

The main problem for the brain disease model is that chronic drug-users engage in extended and sophisticated plans to obtain and use drugs and such patterns of activity are not easily reconciled with the strong form of compulsion the model suggests. "It might be possible that an agent could be compelled—say by an irresistible desire—to take a drug, but it is intuitively implausible that she should be compelled to engage in this long series of actions" (Levy, 2014). Furthermore, addicts carefully consider consequences as they pursue their drug-use, responding to price, avoiding the law, concealing their use, et cetera. In other words, drug-using behaviour involves the consistent integration of a variety of higher-order cognitive abilities that makes it seem like action. Finally, large numbers of addicts eventually give up their drug-use without treatment (Heyman, 2009). That should be impossible on the brain disease model as long as the addict's environment continues to include addictive rewards and cues. None of this is to say that the neuroscientific findings are false but rather that they should be reinterpreted in a way that is compatible with the drug-using action that we see. Whatever neurological impairments addicts have or develop, they do not seem to result in compulsion via irresistible desire. These observations might encourage us to adopt the moral or liberal models but these models have their own difficulties.

Problems for the moral/liberal models

Chronic drug-use appears to count as action, as opposed to physically compelled unintentional behaviour, yet it also often has features that make it seem less than fully controlled. First, many people try to control their drug-use and fail; as we saw above, roughly half of adult smokers

have tried and failed to quit within the previous year. One might argue that all attempts to quit are coerced but many people appear genuinely to want to quit. Second, the latest American Psychiatric Association's criteria (DSM-V) for diagnosing substance use disorder, developed and refined through decades of clinical experience, specifically focus on failures of agency:

1. Using greater amounts or using over a longer time period than intended
2. Persistent desire or unsuccessful efforts to cut down or control substance use
3. Spending a lot of time obtaining, using, or recovering from using a substance
4. Experiencing a craving to use a substance that was so strong they could think of nothing else
5. Repeatedly unable to fulfil major obligations at work, school, or home due to substance use
6. Continued use despite persistent or recurring social or interpersonal problems caused or made worse by substance use
7. Stopping or reducing important social, occupational, or recreational activities due to substance use
8. Recurrent use of a substance in physically hazardous situations
9. Consistent use of a substance despite knowing that a persistent or recurrent physical or psychological problem is likely to be caused or exacerbated by substance use
10. Tolerance as defined by either a need for markedly increased amounts to achieve intoxication or desired effect or markedly diminished effect with continued use of the same amount
11. Withdrawal manifesting as either characteristic syndrome or the substance is used to avoid withdrawal

The manual suggests that substance use disorder be graded for severity according to how many criteria the patient fulfils, 2-3 criteria is mild, 4-5 criteria is moderate, and 6-7 criteria is severe (American Psychiatric Association, 2013, pp. 483-484).[2]

Criteria, 5-9 detect if the agent's drug-use is threatening their job, their financial independence, their physical and mental health, and their relationships with friends and family. These features of life are almost universally valued and so we would expect most people who meet these criteria to be distressed about the damage their drug-use is causing. Of course certain chronic drug-users may just like drug use *that* much. But, I assume, such agents are very rare and cannot

[2] Throughout the thesis I focus on drug addiction but the points I make would be just as applicable to behavioural addictions. At present gambling addiction is the only behavioural addiction officially recognized by the American Psychiatric Association.

account for a large proportion of people diagnosed with substance use disorder. Criteria 1, 2 and 4[3] detect if the agent is failing to meet his own standards of control regarding drug use, which would indicate a distressing failure of agency *independent* of his preferences. That said, most chronic drug-users would meet criteria 3, 10 and 11 because they spend a lot of time in drug-use related activities and inevitably experience tolerance and withdrawal effects. Meeting these criteria would not necessarily impinge on other values so some people diagnosable as having mild substance use disorder may not exhibit any obvious lack of control or distress. However, according to the APA criteria, it appears that those diagnosed with moderate and severe substance use disorder will suffer from some form of agential impairment.

So where does this leave us? There appears to be a clinically diagnosable impairment of agency related to substance use that we can call 'addiction,' even though that agential impairment does not seem to involve the strong form of compulsion suggested by the brain disease model. Furthermore, although *some* chronic drug users control their use, many do not; the moral/liberal models can only explain the action of that subset who do control their use. There may be ways of forcing this evidence into either the brain disease or moral/liberal models but a different model would fit the evidence more comfortably. Addiction appears to involve drug-using behaviour that counts as action, so is not strongly compelled, yet still lacks agential control in some important sense. The theories of agency I consider each attempt to explain what that sense is.

The thesis is divided into two main parts: Part One consists of Chapters 1 through 4; Part Two consists of Chapters 5 and 6. My main focus in Part One is to develop an account of normal agency. An inadequate view of normal agency will incorrectly set the benchmark for gauging impaired agency. With this in mind, the first three chapters are devoted to critiquing existing theories of agency that purport to explain addiction. In the fourth chapter, I develop a narrative account of agency which I argue provides a more accurate view of normal agency. In Part Two, my focus is to explain the phenomena associated with addiction, recovery and relapse. I assess the theories introduced in Part One and argue that the narrative account provides the most explanatory power.

[3] Even for somebody who values drug-use highly, an unanticipated craving that prevents flexible cognition would presumably represent a distressing lack of control in at least some situations.

Chapter 1 – Addiction: A disorder of choice?

I begin with Gene Heyman's account of addiction (2009, 2013). Heyman argues that addiction results not from abnormal brain function but from normal processes of choice. His view is, therefore, relatively closely aligned to the liberal and moral models but, unlike them, he nevertheless considers addiction to be a problem for the addict. For Heyman, normal choice is defined as an attempt to maximise reward. Crucially, he considers pursuit of even the most synchronic rewards to count as normal agency if the agent believes that such choices will maximize reward. Heyman argues that addictive rewards have 'toxic' properties that encourage the agent to choose with a highly synchronic focus. Addicts repeatedly choose drugs to maximize synchronic reward; they do not choose the relatively reward-poor diachronic trajectory of addiction. Of course, not all drug-users become addicts and some addicts recover. Heyman explains these variations in patterns of choice primarily with reference to prudential rule-following where agents learn to follow socially available rules such as, 'all things in moderation.' Less commonly, the agent might exercise the cognitively challenging skill of choosing diachronic sequences of rewards or what Heyman calls, 'global choice.'

I argue that consistent synchronic choice precludes many of the things that people value most. It is, therefore, misleading to call such patterns of choices 'normal.' Healthy agency involves much more diachronic sophistication than Heyman considers. Most notably, agents form and pursue diachronic plans in pursuit of unique goals, e.g. careers, hobbies, holidays. Success with such diachronic plans entails having the capacity for means-ends coherence and ends-ends consistency in planning, considering oneself temporally extended, and being aware of one's current diachronic context. I argue that these wider diachronic skills also underpin competence in the skills that Heyman does recognise, i.e. global choice and prudential rule-following. Equally Heyman underplays the addict's diachronic agency – only the most extreme cases are completely confined to a synchronic bubble. Addicts and healthy agents are often aware of plans or rules to achieve greater diachronic rewards yet still choose lesser rewards. This should not be possible on Heyman's account and he cannot explain why they would choose in this way. In Chapter 2 I consider two competing explanations of why people fail to pursue what they would most want.

Chapter 2 – Diachronic stability in action

Knowing of a beneficial rule, pattern of choice or plan at one time will be insufficient if that knowledge is forgotten or warped at times when the knowledge needs to be implemented. Such

failures of diachronic stability appear central in succumbing to temptation and suffering from addiction. George Ainslie (2005, 2011; Monterosso & Ainslie, 2009) and Richard Holton (2009) offer competing explanations of what causes this diachronic inconsistency and how agents overcome it. Ainslie claims that inconsistency is caused by hyperbolic discounting of future rewards. He argues that diachronic consistency is achieved by forming pre-commitments, and by treating current choices as precedents for future decisions of the same type (or what he calls 'test-case willpower'). Holton argues that 'everyday' temptations cause diachronic inconsistency by corrupting the agent's judgment through a process called judgment shift while the temptations in addiction may bypass the agent's judgment altogether. In either case, diachronic consistency can be achieved by forming intentions, which settle decisions in advance, and wielding a finite resource under executive control called muscle model willpower.

I argue in favour of Holton's account. Pre-commitments are too inflexible, too weak, or require too much planning to do the job alone. Ainslie recognises this and relies on test-case willpower for most diachronic stability. However, I explain why test-case willpower cannot account for the diachronic stability of normal agency. It is arational, cannot adequately stabilise unique plans, does not allow the diachronic coordination that comes from making decisions in advance, and is vulnerable to judgment shift. Intentions on the other hand are rational within certain constraints, allow for sophisticated diachronic coordination of plans, and can diachronically stabilise unique plans, prudential rules (or policies), and global choices. Pre-commitments, however, remain a useful tool for achieving diachronic stability and I argue that they should be added to Holton's account.

Chapter 3 – Normative planning agency and self-governance

In Chapter 3 I continue to argue against Ainslie but on the grounds that his reductive account effaces the agent and the possibility of self-governance. Ainslie describes all action as the result of extra-agential processes of reward expectation; the agent is just the passive outcome of those processes. I develop my arguments against Ainslie by drawing on a theory closely related to Holton's account – Michael Bratman's planning agency. On Bratman's account, the individual constitutes herself as a self-governing agent by creating a network of intentions (plans and policies) according to three norms of practical reason: diachronic stability of intentions, means-ends coherence in planning and ends-ends consistency among intentions. Normatively endorsed plans and polices have agential authority and the more closely she follows those plans and policies the greater self-governance she exhibits.

Bratman's account provides more explanatory power than Ainslie's account in three ways. First, it can describe the phenomenology of a struggle against temptation as a synchronic clash between a normatively endorsed intention and an intention that would break a norm. Ainslie assumes all rewards are commensurable and so there is never a rational basis to resist pursuit of one's current greatest reward expectation. Second, Bratman can explain *degrees* of self-governance with reference to the amount of willpower used, the strength of temptation, the importance of the intentions threatened, and diachronic efforts to avoid temptation arising. For Ainslie, action is determined by extra agential forces so *self*-governance is impossible; he is therefore forced to redescribe our established folk beliefs and practices based on self-governance. Third, Bratman can distinguish extra-agential changes in desire over time from agentially-driven changes in desire through plans and policies. Ainslie insists that all changes in desire are extra-agential even though we are far from understanding those changes in objective terms.

Chapter 4 – Narrative agency

I begin Chapter 4 by arguing that Bratmanian planning agency should be supplemented with a wider notion of self-concept because self-concept influences self-governance. Bratman holds the view that agents are synonymous with their network of intentions while their contrary desires are extra-agential. However, we typically consider our contrary desires to be as much a part of our self-concept as our intentions. Furthermore, our self-concepts include many contingent features of our identities and situations, e.g. our gender, race, bodily characteristics, where we were born and raised, who our parents are, the medley of accidents, windfalls and unexpected consequences that happen *to* us, and the future inevitabilities that we anticipate happening to us. Self-concept influences self-governance because our intentions have to make sense in light of our self-concepts; by interpreting our self-concepts in different ways we can adjust what intentions it makes most sense to hold.

I then go on to argue that agents' self-narratives connect their intentions, desires and contingencies so that they make sense in light of each other. I describe how agents control their lives through the process of narrative self-constitution and identify the constraints that apply to that process. The narrative view offers several explanatory improvements over the Bratmanian view. First, the motivational character and intensity of intentions, desires and contingencies is influenced by the prominence and connections we afford them in our narratives. Self-governance can therefore be enhanced through strategic narration and undermined by

detrimental self-narration. These effects and means of control do not appear on a Bratmanian account of agency. Second, established self-narratives can be difficult to change because they have an independent influence on motivation, i.e. self-narratives have momentum. Narrative momentum has two general sources. The established plot of the self-narrative limits the plot continuations that can make sense and agents are averse to living nonsensical lives (Velleman, 1989). Furthermore, the existing self-narrative pre-reflectively directs attentional focus and sets the context for experience and cognition which limit the agent's ability to recognise alternative interpretations of their life. There is no equivalent of narrative momentum on Bratman's account and so he underplays the difficulties that can be involved in making significant changes to one's life; it is not always as simple as changing one's plans and policies.

Having critically examined these various theories of agency in Part One, in Part Two I assess the adequacy of these theories for explaining addiction, recovery and relapse. For ease of presentation, I treat Heyman's and Ainslie's positions as versions of choice accounts and I treat Bratman's and Holton's positions as versions of planning accounts. In Chapter 5, I argue that planning accounts provide a number of explanatory benefits over choice accounts and in Chapter 6 I argue that the narrative account improves on planning accounts.

Chapter 5 – Choice accounts versus planning accounts

Choice accounts define action as pursuit of maximum available reward. Some trajectories of drug use and recovery involve a consistent pursuit of the greatest reward. However, many addicts, even those who recover relatively easily, report distress, ambivalence, and struggle synchronic to their drug use. These addicts are aware of preferable ways to live even as they use and, I argue, these people are the paradigmatic cases of addiction. The choice theorist is restricted to describing such addicts as deceptive, self-deceived, or behaving unintentionally; however, these explanations are inadequate. Similarly, choice accounts cannot explain why both recovered agents and clinicians refer to recovery as a process that demands the agent's *efforts* rather than merely a change in the available rewards. Planning accounts, in contrast, can explain the struggle and distress associated with addiction because attempts to improve or maintain self-governance are effortful. These accounts can also explain a variety of other addiction phenomena with reference to specific practical norms. First, the resigned fatalism of some addicts may result from a failure of means-ends planning to either *create* motivating goals or *connect* with known goals. Second, chronic ambivalence can be characterised as a failure to find ends-ends consistency among highly valued but incommensurate goals. Finally, the

planning account provides a way of understanding why treatments that support these practical norms work.

Chapter 6 – Planning accounts versus the narrative account

In the final chapter I argue that the narrative account reveals several factors relevant to addicts struggling for self-governance that do not appear in planning accounts. First, established self-narratives of addiction limit the recovery-directed intentions that appear plausible to the agent. These narratives undermine the adoption of potential recovery intentions even if those intentions would meet norms of practical reason. Self-narrative reinterpretation can overcome this effect. Second, the context created by a self-narrative influences the motivational effects of the narrative's constituents. Narrative contexts can, therefore, be developed to favour recovery-enhancing intentions, desires and affective responses. Third, the development of diverse narrative foci is essential for recovery but made more difficult if one's self-narrative has narrowed to focus exclusively on drug use, treatment and abstinence. The planning account underplays this effect by ignoring the importance of being able to link developing projections with past contingencies. Fourth, the narrative account provides a more plausible explanation than planning accounts for why the ends-ends inconsistent behaviour of addicts can be so entrenched. Fifth, the narrative account emphasises the co-authoring effects of other people and archetypal narratives. Addicts aiming to recover should engage with people who will best support their nascent recovery narrative. The planning theorist might expand his account to consider how others help us plan but this account cannot explain how others help us interpret our contingencies and connect them with our plans. Many of these narrative effects will be present to some extent in all struggles with addiction; however, in some cases, they arguably make the difference between continuing addiction and recovery.

I then outline how this way of explaining addiction entails a change from Bratman's view of self-governance to Christman's view. That is, in cases where agents have become alienated from their addiction self-narratives, self-governance is not merely a matter of enacting existing intentions (or narrative projections); rather, agents will exhibit greater self-governance, i.e. recover, if they re-narrate a life that better coheres with their diachronically stable evaluations. Given this relationship between self-narrative and self-governance, we should expect significant implications for the treatment of addiction. I conclude the thesis by considering some of these implications.

Chapter 1: Addiction: A disorder of choice?

Introduction

In this chapter I outline and critique Gene Heyman's recent work, *Addiction: A disease of choice* (Heyman, 2009),[4] in which Heyman argues that addiction emerges as a function of everyday choice and not from abnormal neurology or 'brain disease.' However, unlike proponents of the moral or liberal models, he believes that being addicted is a problematic state that agents would usually prefer to avoid. This critique serves two broad purposes. First, I begin to stake out where I stand in the debate over the aetiology of addiction. Second, I begin to more fully characterise both what is required for normal human agency and what might have gone wrong with it in addiction; this second goal occupies the majority of the thesis.

The first section of the chapter outlines Heyman's argument for the claim that addicted choices are ordinary, everyday choices. According to Heyman, ordinary choices need not involve any foresight; even the most synchronic choice is an ordinary choice if the agent believes it will maximise reward. The reason that ordinary choice leads some people to self-destructive addiction and not others is because of the 'toxic' properties of addictive rewards. Those toxic properties encourage the agent to choose so synchronically that they fail to see how a different pattern of choice would maximise reward across a more diachronic timeframe. Heyman claims that people recover from addiction and avoid addiction despite exposure to addictive rewards by following prudential rules, such as, 'all things in moderation.' Alternatively, addiction can be avoided by adopting a global perspective and choosing diachronic sequences of rewards but this is rare because it is so cognitively demanding.

I begin my critique of Heyman's position by arguing that he has set the standard for everyday agency far too low.[5] Consistent synchronic choice cannot be typical because it will preclude many of the things that people value most. Healthy agents (and many addicts) display much more diachronic sophistication in agency than Heyman acknowledges. In particular, agents successfully pursue unique diachronic goals, such as careers, hobbies, and holidays. I argue that the agential resources that Heyman refers to, rule-following and choosing sequences, are insufficient to support such diachronic goals. I go on to claim that successful pursuit of diachronic goals require a raft of planning skills and an awareness of one's unique diachronic

[4] I also refer to his more recent work, "Addiction: An emergent consequence of elementary choice principles" (2013), which builds on the same core arguments of the book.

[5] I appeal here to a folk-concept of what counts as normal, healthy human agency; agency that is by no means perfect but that comfortably avoids any pathological diagnosis. The points I will make are, I believe, general enough that my assumptions as to what counts as normal, healthy or typical will be uncontroversial.

context. Furthermore, these additional diachronic skills are also required for the competent use of the agential skills that Heyman considers.

The main problem for Heyman's explanation of addiction is that addicts continue to use despite knowing of prudential rules that, if followed, would provide more reward and despite the very high costs of continued use. Addicts seem to have an inability to apply such rules and so addiction is not simply a result of everyday choice but involves impaired agency. My analysis suggests that prudential rule-following is only one of many agential skills that might be impaired in addiction. I investigate these skills in more detail in the subsequent chapters.

Heyman's account

Heyman argues against the view that currently dominates medical opinion that addiction is a chronic relapsing brain disorder. He draws on a variety of empirical sources to support the alternate view that the large majority of addicts choose to use drugs and, when it no longer suits them, they choose to stop using drugs. The phenomena of addiction, recovery and relapse are all "a function of the rules of everyday choice" (Heyman, 2013, p. 428).

Voluntary behaviour

Heyman's first task is to define 'everyday' agency. To that end he distinguishes voluntary from involuntary behaviour as follows:

> "…The degree to which an activity is voluntary is the degree to which it systematically varies as a function of its consequences… If the factors that affect everyday decision are the same as those that affect drug use in addicts, it is voluntary. However, if the scale of these factors is such that they are not legitimate – say, only the threat of severe punishment brings drug use to a halt in addicts – then for practical purposes, drug use in addicts is involuntary" (Heyman, 2009, pp. 104-105).

So Heyman is clear that volition comes in degrees and the distinction between voluntary and involuntary behaviour is normatively defined.[6] Where on this continuum of consequence-

[6] However, as we see in his characterisation of 'local choice' outlined below, he sets this normative limit extremely low. Furthermore he doesn't consider any of the more demanding normative standards that distinguish degrees of control in voluntary behaviour. If we think that addiction often involves poorly controlled voluntary behaviour, rather than involuntary behaviour, then these other normative standards are likely to be more relevant. I discuss this in more depth in Chapter 3.

responsiveness does addicted behaviour fall? The evidence suggests that most agents' patterns of drug-use are quite responsive to consequences. For example, the drug-use of addicts is price-sensitive (Elster, 1999; Neale, 2002) and it even remains price sensitive after an initial dose (Fingarette, 1988, pp. 36-42). If we look at substance dependence as broadly as possible, i.e. including people who don't seek treatment, we find that, of people who meet the criteria for substance dependence during their lives, roughly 80% eventually recover, mostly before they reach 35 years of age.[78] Heyman proposes that this is because the consequences of drug-use change as one matures; the alternative rewards of a career and family develop, adult economic concerns arise, et cetera (Heyman, 2009, 2013). In the context of treatment, pilots and physicians have a high rate of recovery (80-90%) (Coombs, 1997) which Heyman claims is because they stand to lose their valuable careers (Heyman, 2009, p. 86). Recovery rates in more general treatment circumstances are lower, "…about 50-60% of patients begin re-using within six months following treatment cessation, regardless of the type of discharge, the patient characteristics or the particular substance(s) of abuse" (McLellan et al., 2005, p. 449). Heyman implies that this is because the average person doesn't have as much to lose as a pilot or physician and so drug-use might be worth the risks. However, he also notes that addicts in treatment have a higher rate of comorbid psychiatric illness which undermines their development of rewarding lifestyle alternatives to drug-use (Heyman, 2009, p. 85).

In summary, then, drug-use in addicts is as voluntary as any act because it is guided by costs and benefits. Drug-using individuals happen to like drug-use but otherwise they are normally motivated other concerns such as family, cultural values, self-esteem, fear of punishment, and so on. Drug-use may escalate to problematic levels but that depends largely on "the nature and magnitude of the available nondrug reinforcers" (Heyman, 2013, p. 436).[9] In other words even more extreme patterns of drug-use are voluntary when they are the best of a poor range of options.

Heyman realises there is a problem for his claim that addiction is voluntary – sometimes drug-users *don't* respond to the severe self-destructive consequences. What should we say, for example, about the 10-20% of physicians and pilots who don't swiftly recover despite the

[7] These figures come from several large surveys: The Epidemiological Catchment Area Study (1980-4); The National Comorbidity Survey (1990-2) and its replication (2001-3); The National Epidemiological Survey on Alcohol and Related Conditions (2001-2). Results can be seen in Anthony and Helzer (1991); Kessler et al. (2005) and (2005); Stinson et al. (2005, 2006); Warner et al. (1995).

[8] Recovery typically occurs without treatment; only 30% of people ever mention their substance use to a medical professional and only about 16% of addicts at any time are concurrently in treatment (Heyman, 2009, p. 70).

[9] Heyman notes that the magnitudes of non-drug reinforcers depend on the capacity of the consumer to take advantage of them and the capacity of the environment to provide them.

threat to their careers? Even more troubling are the people who continue to use drugs despite the consequences of alienating their family and friends, losing their job and home, and, in all too many cases, dying. Ongoing drug-use in such cases appears to more than meet Heyman's criterion for *involuntary* behaviour because it doesn't stop with the threat of severe punishment.

Nevertheless, Heyman maintains that drug use in these cases is voluntary because the agent can make two kinds of analyses of the costs and benefits available to them, one by framing the options globally, the other by framing them locally. Both global and local choice are forms of voluntary behaviour but choosing locally can lead to self-destructive patterns of drug-use despite poor longer-term consequences.[10]

Dynamic preferences, local choice and global choice

To understand Heyman's account of global and local choice first we need to understand preference dynamism. Choices and preferences (desires)[11] are mutually dependent. We choose the option that we most prefer and our preference ranking of options changes as a function of our choices. For example, if we binge on a particular attractive option we get sick of it, if we abstain from certain options our preference for them increases.[12] Every option available for choice has a preference ranking which changes over time, either becoming more preferred or less preferred (even aversive). Preferences change at different rates and according to different patterns. Many preferences oscillate – preference for an option increases over time so when that option is eventually chosen it is highly satisfying. However, choosing that option is satiating, decreasing the preference for it. Once satiation decreases the preference for that option below preferences for other options the agent changes their focus to those other options. In the meantime, the preference for that original option slowly begins building again. This oscillating pattern clearly describes preferences for food, water and sex but also other activities such as playing games, socialising or exercising. Other kinds of preference involve long term trends. Some activities are fairly unpleasant at first but become more and more

[10] Therefore Heyman's graded notion of voluntary behaviour should be rephrased as follows: the degree to which an activity is voluntary is the degree to which it systematically varies as a function of its *local* consequences; response to longer-term consequences is an optional extra.

[11] Heyman refers to 'preferences' which I take to be synonymous with desires. I will use 'preference' and 'desire' interchangeably.

[12] When we are inexperienced with certain options we don't know what we will get sick of and what we will prefer if we abstain and so, initially, our preferences don't change as a matter of choice itself. But once we learn the effects patterns of choice have on the satisfaction we derive from those options then our preferences change as a function of choice because we can anticipate the variation in satisfaction.

enjoyable in an open ended fashion, e.g. developing musical or sporting skills. In contrast some preferences we grow out of permanently, e.g. we stop enjoying certain musical styles, we become bored by games we enjoyed as children. More synchronic oscillations in preference are superimposed on longer-term trends, for example, even as the rewards from learning to play the piano increase in an open-ended way, one still gets tired of playing the piano for too long at any one time. We control these trends to some extent by how we choose, e.g. when we satiate appetites, when we let appetites develop, how often we practice the piano, et cetera.

Given preference dynamism, Heyman suggests that an agent can choose locally or globally. In local choice the agent chooses the option that will maximise the satisfaction of *current* desire. Local choice requires no memory or future projection, except the minimum retention and projection necessary for experience; it is a method available to even the most synchronic agents. Global choice is where the agent considers the effect of the present choice on their future preferences. They then choose an ordered series of options that will maximise the satisfaction of their dynamic preferences for that *set* of choice situations. Global choice is only available to an agent with the diachronic awareness necessary to predict their future options and preferences. To attain the global benefits after becoming aware of them, the agent must occasionally choose against their most preferred local option.[13]

Heyman illustrates the two modes of choice in the restaurant dilemma: Imagine you will eat out every night at either an Italian or Chinese restaurant and your initial highest preference is for Chinese food. However, habituation reduces your subsequent preference for a food type while dishabituation increases your preference for a food type so that your preference changes according to the restaurants you choose each night. Habituation and dishabituation processes are stronger for Chinese food.[14] How would you decide which one to choose each night under these conditions?

Heyman reports that the majority of respondents say they would choose based on the strength of their current preferences each time the choice arose (local choice); a small minority (that can be increased by prompting) spontaneously say they would choose with an eye to the subsequent choices and preferences they would have (global choice). Local agents choose a

[13] Heyman isn't clear as to whether the strongest local preference represents a temptation for the agent who has embarked on a global series of choices. We are left wondering if weakness of will is possible for the global chooser and, if so, how it should be characterised. Furthermore, Heyman isn't clear whether the global agent commits to their series of choices from the first choice (i.e. forms an *intention*) or if they reconsider it in each new choice situation. I return to these issues in Chapter 2.

[14] In the example numbers are attributed to initial preferences and (dis)habituation rates but here I leave them out for simplicity of exposition.

series of individual meals while global agents choose a meal series. The global choosers occasionally choose Italian even though their current preference is for Chinese. They do this because they know that by letting the preference for Chinese increase through dishabituation they will enjoy Chinese even more on a future night. In Heyman's example (where figures are assigned to initial preferences and (dis)habituation rates), a local chooser will reach an equilibrium where they eat nearly seven Chinese meals for every three Italian meals while the global equilibrium is about four Chinese meals to every six Italian meals.[15] In this case the global equilibrium provides an average rate of benefit, i.e. desire satisfaction, 20% higher than the local equilibrium.

Based on the restaurant example we can get an initial impression of the advantages and disadvantages of each model of choice. "The options in local choice are concrete and correspond to how things look. Local choices involve items that have clear physical outlines and activities that are easy to name. … In contrast, the aggregates of global choice have no naturally occurring boundaries, but are abstractions. …Put another way, local choice corresponds to the natural fracture lines of perception; global choice does not" (Heyman, 2009, pp. 138-139). So choosing locally is beneficial because local preferences are obvious to the agent (leaving aside unconscious desires) and, therefore, relatively fast and cognitively easy. These are significant benefits because we have limited resources of time and cognition. The local reward also tends to be more or less guaranteed once the agent decides to choose it. The downside is the reduced rate of benefit the agent accesses over time. Global choice, in contrast, provides access to those higher rates of benefits while the downside is the greater cognitive and time costs needed to work out the global strategy. So we can see that the agent with the cognitive potential for global choice will have to balance spending the time and cognitive effort to access global benefit with giving up those extra benefits to save on time and effort.

Heyman claims that because few people choose the global strategy in the restaurant example agents appear to pre-reflectively prefer to save on time and effort rather than access the global benefit. This is further supported by empirical studies that Heyman cites in which local choice seems typical (Herrnstein, 1997; Herrnstein et al., 1993; Heyman & Dunn, 2002; Kudadjie-Gyamfie & Rachlin, 1996; W. Mischel et al., 1992). Given this evidence, Heyman thinks that

[15] We can speak of equilibria in this case because the series of choices goes on indefinitely but finite series of choices also typically generate different global and local patterns of choice.

local choice is the default state and, although humans are capable of making occasional global choices, they rarely do so spontaneously.

Global/local choice and self-destructive addiction

If local choice is our default mode of choice then we will then tend to access sub-optimal rates of benefit by excessively consuming locally attractive rewards. Heyman argues, however, that this excess doesn't usually become self-destructive because our preferences naturally oscillate in response to typical rewards. For example, if our preference for Chinese decreases as we eat it, the greatest local preference then becomes, say, watching a movie and eating an ice-cream. If the preferences for the movie and the ice-cream remain high throughout, then we watch to the end of the movie and finish the ice-cream. We then find that the next local preference is to go home to bed, and so on. We may end up eating Chinese, going to the movies, and eating ice-creams a little too often, but our activities change; we don't find ourselves eating Chinese indefinitely.

Addictive rewards, however, are not like typical rewards; they have a range of properties that send locally choosing agents into self-destructive spirals. First, typical local rewards are directly satiating; we feel full when we eat, we get fatigued when we exercise, we become bored with mental exercises, et cetera. Addictive rewards do not satiate so directly and so people binge on them; tolerance develops over time but has a less immediate effect. Second, typical local rewards do not undermine the reward to be had from other activities, e.g. there's nothing about eating Chinese that makes movies less appealing. Addictive rewards, however, are "behaviourally toxic" – they significantly reduce the motivation and time available for other activities. Finally, addictive rewards provide almost instantaneous and certain pleasure but delayed and somewhat uncertain costs. Given this combination of properties, the local preference for addictive rewards only very rarely dips below that of other rewards. This situation may be further exacerbated by intoxication which stunts the cognitive powers needed to recognise the longer-term consequences and make a global choice.[16]

[16] The evidence suggests that habitual substance use doesn't usually create any long-term cognitive deficiencies (Lyons et al., 2004; Toomey et al., 2003) but even relatively short term deficiencies will be sufficient to prevent behaviour change.

> "Not all addictive drugs express each of these properties to the same extent, but stimulants, opiates, and alcohol produce these effects to a greater extent than do other substances and activities"[17] (Heyman, 2009, p. 152).

Despite addictive rewards having long-term self-destructive consequences, their 'toxic' properties ensure that they remain the most preferable local option much of the time. Therefore the distinction between local and global choice combined with the properties of addictive rewards provides an explanation for voluntary drug-use that happens to be self-destructive.[18]

This, though, leads us to another quandary. If addictive rewards have these properties, and people choose locally by default, then addiction should inevitably result when agents are exposed to addictive rewards; yet this isn't the case. Roughly 95% of Americans who have used an addictive drug at least once did not become addicted while about 85% who drank alcohol did not become addicted (Conway et al., 2006; Hasin et al., 2007; Stinson et al., 2005; Substance Abuse and Mental Health Services Administration, 2013). If we assume, for the moment, that addictive rewards have the properties Heyman suggests,[19] then we must conclude that most people are not choosing locally, at least not when considering addictive rewards.

Heyman surveys a few possible explanations: some people may have genetic profiles that make drug-use locally less preferable (Tsuang et al., 2001), those in close relationships might be subject to intense local opprobrium if they use drugs (Robins & Regier, 1991), and differences in impulsivity and cognitive skill may make some people more able to take the global perspective (Heyman & Dunn, 2002; Heyman & Gibb, 2006; Kirby et al., 1999; Vuchinich & Simpson, 1998). But he admits that these explanations can only partially explain the discrepancy between the exposed population and the addicted population.

> "There are large numbers of married addicts and even larger numbers of individuals who are single who nevertheless are not addicted. … Some addicts learn the global equilibrium, and some wait for the larger, later reward. Conversely, many nonaddicts

[17] I will set the case of cigarettes aside here because they are a special case: "Cigarettes are not intoxicating or particularly rewarding, relative to other drugs, but they make up for these 'deficiencies' by filling a niche that until recently had no competition" (Heyman, 2009, p. 152).

[18] The self-destruction itself is not voluntary because it appears on the global time-frame so the agent doesn't consider it.

[19] The difference between normal rewards and addictive rewards may not be this stark; but if addiction is not caused by a special class of reward then we must look elsewhere. Impaired agency would explain abnormal action towards otherwise normal rewards but this goes against Heyman's thesis. I revisit this question at the end of the chapter.

> fail to learn the more efficient global-choice solution and choose the sooner, smaller rewards. Indeed, the responses to the restaurant problem and other observations imply that most of the time most individuals adopt a local frame of reference when making choices, yet most do not become addicts" (Heyman, 2009, pp. 160-161).

This admission raises two problems. First, people seem to commonly choose in ways which are consistent with the recommendations of global choice despite the supposed rarity of global choice itself and the limited local pressures countering addictive rewards. Second, local and global choice styles are general so a capacity to approximate a global perspective should be applicable across the board. Why then, do we see people approximating global choice in some domains, such as household finance, and not others, such as diet? To explain this within-agent variation in diachronic control by domain, Heyman introduces the notion of prudential rules.

Prudential rules

Prudential rules provide domain specific approximations of global choice without the high cognitive demand. The domain specificity of prudential rules is a function of their variety. Some appear explicitly as rules, from the general – 'enjoy things in moderation', 'don't burn your bridges,' 'don't cry wolf' – to the specific – 'look both ways before crossing the road,' 'roast until the juices run clear.' Other prudential rules appear as ideals relevant to certain roles, such as, being a mother, worker, or child. Yet others define social identities, such as being a 'wild and crazy,' 'sober,' or 'religious' person.[20] Prudential rules are distinguished from role and identity descriptors more generally (e.g. paedophile, sloth) because they are normatively endorsed by society or some societal sub-group. "Prudential rules are inextricably linked with values, and values are embedded in signs of approval and disapproval that accompany social interactions" (Heyman, 2009, pp. 162-163). Because all prudential rules are held in the body of cultural knowledge and are passed on in discourse, exposure to certain prudential rules depends on the discursive environments one inhabits.

Prudential rules approximate global choice by guiding our natural appetites; they indicate when, where, who with, and how to, relax, be patient, eat, sleep, drink, express emotion, or have sex. Since our natural appetites are a strong influence on our local preferences it's plausible that prudential rules help limit excessively local agency. In the restaurant dilemma,

[20] We should bear in mind that an extremely wide variety of conflicting rules enjoy *some* normative endorsement and many more are normatively neutral so the agent faces a lot of work to judge which rules will be beneficial to her personal situation. I return to the issue of selecting rules below.

for example, if the agent applies the rule, 'all things in moderation,' he should decrease his frequency of eating Chinese (assumedly this involves ignoring or inhibiting his appetite for Chinese). The better he does this the closer he approximates global choice.[21]

This approximation of global choice requires less cognitive effort than global choice for three reasons. First, the individual doesn't have to create the rule; that work has already been done by someone else. Second the individual doesn't have to assess the rule closely to see if it provides increased benefit. Its very existence is a good indicator, if not a guarantee, that it tends to increase benefits over time for most people.[22] Third, the rule-follower doesn't need to make any predictions of the future or consider the past, they just have to assess whether this current situation is one where the rule applies or not. Or, as Heyman puts it, "[global choice] involves abstract thought processes, such as keeping track of variables, considering alternatives, and tracking down contingencies. In contrast applying rules is a matter of judging similarities. Which rule best fits the case at hand?" (2009, p. 166).

But why would the agent follow a prudential rule when the action it advises clashes with his highest local preference? Following prudential rules tends to improve the benefits we attain over time and those benefits reinforce rule-following.[23] This still leaves the issue of getting rule following off the ground, why would someone follow a prudential rule for the first time? Here Heyman suggests that normative pressure does the job. Locally applied, or anticipated, social approval or disapproval motivates initial rule-following.[24] More generally this fits with our folk understanding of teaching children – when children are young rules have to be socially enforced but eventually children begin to follow rules because they understand the merit of the rule. Normative pressures do not disappear but presumably, maturing agents eventually start to try out new rules even without any particular normative pressure because of their positive experience with earlier rules.

[21] Of course he could go too far and fail to maximise desire satisfaction by eating Chinese too infrequently. To try and counter such cases we have other prudential rules such as, 'treat yourself sometimes' and 'live for the moment.' Again, this raises the problem of how to balance contradictory rules.

[22] Certain prudential rules might survive because they benefit society rather than the individual rule-follower. Other rules may even survive as selfish memes and merely provide an illusion of benefit. I will set these issues aside on the assumption that most prudential rules benefit the rule user when used correctly.

[23] This assumes the agent has the diachronic awareness to make the association between the benefit and following the rule. I consider this in my analysis below.

[24] Heyman's view is informed by Ellingsen and Johannesson's study (2008) where local choices, to share a monetary windfall with an anonymous stranger or not, were strongly influenced by anticipated social approval or disapproval in the form of a letter from that stranger. Money winners that were told of the letter (despite knowing they would remain anonymous) were nearly twice as likely to share their winnings evenly with the stranger than those who weren't told of the letter.

Heyman suggests that global choice and prudential rules interact. Because prudential rules are relatively vague any rule accommodates a variety of behaviour patterns. Agents occasionally use global thinking to tailor prudential rules to their own case. For example, one might start with the general rule, 'all things in moderation,' and from that develop the more personal rule, 'only two Chinese meals a week,' based on the global analysis of experience ('last week I had three Chinese meals and the third one just wasn't that satisfying'). Although Heyman doesn't explain where prudential rules come from, I assume they develop from the reverse of this process. An individual designs a pattern of global choice for themselves and as it is taught or appropriated the rule is tailored to be more general so that it is valuable to a greater range of agents.

Prudential rules and addiction

Let's return now to the original motivation for positing prudential rules. Global choice is too difficult to do very often yet the majority of people who are exposed to addictive rewards approximate global choice and don't become addicted. Furthermore, we see intra-agential variation; some people do become seriously addicted but manage to approximate global choice in many domains while others don't become addicted yet choose locally in many domains.

Prudential rules can explain both observations. The majority of people exposed to addictive rewards follow prudential rule(s) that guide an approximation of global choice when they are considering those rewards. There is some empirical evidence to support this claim – exposure to religious, conservative and spiritual rules has been shown to be protective against addiction (Galaif & Newcomb, 1999; Galaif et al., 2007; Newcomb et al., 1999). The intra-agential variation is explained by the domain specificity of prudential rules. I could know of and follow a rule targeting low alcohol consumption but not know of, or not follow, a rule targeting cocaine, fatty foods, paying rent on time, saving for retirement, et cetera.[25] So, Heyman's theory predicts that addicts do not follow prudential rules that are relevant to their self-destructive drug use, although they may still use a range of prudential rules that keep them approximating global choice in every other facet of their lives. In contrast, non-addicts might still choose in excessively local ways in many scenarios but will not become self-

[25] Heyman glosses over the agent's responsibility in learning rules or applying those he knows of. If the agent cannot usually think globally, then he cannot recognise the value of a rule for himself. Therefore rule adoption usually comes down to social pressure. This makes the agent relatively passive in addiction onset and recovery. I consider this below.

destructive addicts if they can just consistently follow prudential rules relevant to addictive rewards. So, prima facie, prudential rules are suitable to do the explanatory job Heyman requires of them and there is some evidence that they are indeed doing this job.[26]

Summary

The picture of agency we are left with is as follows. Agents choose locally by default; it's rare that they choose globally. However they follow a vast variety of prudential rules that allow them to approximate global choice with relatively little cognitive cost. The agent initially follows whatever prudential rules they are socially pressured into. As the mature they voluntarily trial new rules and may occasionally tailor rules to fit their personal circumstance using global thinking.

If we accept this picture it has some interesting consequences. First, the balance of local and global choice doesn't play a large role in distinguishing healthy agents from addicts. Addiction and recovery depend much more on the prudential rules agents follow (Heyman, 2009, p. 167). Second, diachronic thinking is barely needed to live a healthy, typical, human life. All that is needed is an appropriate set of prudential rules. Agents can think highly synchronically most of the time because diachronic thought is superfluous for both local choice and the application of prudential rules.

Critique of Heyman

The central theme of my critique is that, contrary to Heyman's account, typical human agency depends heavily on diachronic cognition and control. Local choice, synchronic rule-following, and rare instances of global choice are insufficient to explain typical agency. I begin by pointing out that local choice is more damaging to agency than Heyman realises – it will cause self-destructive behaviour even in the absence of addictive rewards. As a result, diachronic control is not only needed to avoid addiction but to pursue many of the most fundamental goods that agents value, such as skill development and relationships. Typical

[26] However, I think it is hard to believe that most people who have problems with drug-use do not know of prudential rules guiding moderation or are not exposed to social pressures to conform to them. I return to this below,

agents are often successful in these endeavours so an account of agency is under pressure to explain how they do so.

With this in mind I reassess the resources of diachronic control in Heyman's account beginning with global choice. I find that, although global choice is cognitively demanding in general, as Heyman suggests, it is also significantly affected by epistemic limitations. As a result, it is relatively easy in cyclically recurring situations where knowledge of relevant outcomes of choice is more easily developed. In significantly novel, non-recurrent situations, however, a lack of knowledge undermines the global choice calculus making it outright impossible. I then turn to prudential rules but find that these too are of limited help in non-cyclical situations because their generality underdetermines choice.

Non-cyclical scenarios happen to be very common in human lives because every life has a unique trajectory. Yet typical agents manage the associated epistemic limitations and make neither local nor global choices but what I call, 'singular diachronic choices.' Singular diachronic choices target unique future states of affairs (not bundles of familiar options) and include some of our most important goals such as intimate relationships and careers. To understand how agents make singular diachronic choices we need to go beyond the explanatory tools Heyman provides. In distinguishing singular diachronic goals from global and local choice we find that they depend on two core features of human agency – planfulness and temporally extended self-conception (Bratman, 2007, p. 21) – that are missing from Heyman's account.[27]

I then argue that competent rule-following cannot rely on synchronic cognition alone as Heyman claims. Competent rule-following depends on the diachronic reflection required for global choice as well as the planfulness and temporally extended self-conception needed for singular diachronic goals. Finally I outline some of the implications of my critique for Heyman's view of addiction. The typical agent has a much greater involvement in shaping their temporally extended life than Heyman envisages, therefore there is a much wider range of things that might be going wrong in addicted agency than merely a lack of specific prudential rules.

[27] Bratman refers to three core features, the other being reflectiveness. I don't concentrate on reflectiveness because Heyman has included it, if in a limited capacity, in global choice.

Local choice outside the restaurant

The restaurant dilemma is a deliberately simplified case to bring out the distinction between local and global choice, however it glosses over significant aspects of 'real world' decision-making. If we look at local choice in the real world we see that it is much more destructive than it appears in the restaurant case. Local choice favours short-term benefits despite greater long-term costs and it rules out choosing long-term benefits if they entail short-term costs. In other words the local chooser always acts in accordance with their immediate strongest desire. We have seen why repeated local choice provides a lesser rate of benefit in the restaurant example, but outside the restaurant example the implications are far greater than just not enjoying your Chinese meal as much as you might have.

Chronic local choice wouldn't just result in excessive, sleeping, eating, drinking, drug-use and sex, it would prevent persistence in the face of challenges or temptations that undermine the things that people tend to value most, e.g. learning new skills, working on relationships, longer term health concerns, careers, hobbies, et cetera. Heyman hopes that as you get sick of one activity, another naturally becomes more interesting. That might be so but there are two problems with this view. First if you want to succeed, say, in your career, you often have to persist *despite* being (locally) sick of it. Second, the cycle of local choice this natural preference change would entail is only going to involve locally attractive activities such as sex, food, watching TV, and sleep. Other, less immediately satisfying, activities wouldn't get into the cycle.

In other words, pure local choice would result in a severely imbalanced life by typical human standards. The local agent doesn't just miss out on the maximum rate of benefits; he misses out on many of the most value-laden activities in their entirety.[28] In the real world, an agent choosing purely locally would clearly suffer from impaired agency. Korsgaard gives an example which nicely brings out the impairment in agency of such locally focused choice for human agents:

> "Jeremy settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at a book. Before he finds it, however, he meets his friend Neil, who

[28] One may still want to describe this in terms of a reduced rate of overall benefit rather than admit that some rewards are incommensurable but, even if we take reward to be commensurable, the overall rate of benefit appears to be massively reduced.

> invites him to join some of the other kids at the bar next door for a beer. Jeremy decides he can afford to have just one, and goes with Neil to the bar. When he arrives there, however, he finds that the noise gives him a headache, and he decides to return home without having a beer. He is now, however, in too much pain to study. So Jeremy doesn't study for his examination, hardly gets a walk, doesn't buy a book, and doesn't drink a beer" (2008, pp. 116-117).

Jeremy fails to complete any action or, I assume, get much benefit from any of his half-completed actions, due to his pattern of local choice throughout the evening. He would have been able to see these actions to completion despite choosing locally if only his desire to complete that action happened to remain highest throughout the action. But, as we saw, his greatest desire at any moment was dictated by contingencies, e.g. noticing a book title, running into a friend. If an action would only ever be completed if it were uninterrupted by contingencies then happening to complete an act is itself contingent.[29] As Korsgaard argues, in these cases the agent doesn't *will* anything, he doesn't try to shape himself in the face of contingencies.[30]

If Jeremy doesn't stop choosing locally the negative effects are going to go well beyond failing to study, have a beer or buy a book. *Any* commitment he makes is at high risk of being unfulfilled and this will prevent him from achieving any longer-term project. Most seriously it will undermine all cooperative and loving relationships; nobody wants to employ, work for, or enter an intimate relationship with someone who cannot keep even the most basic commitments.

Agents who have been pursuing long-term goals despite short-term costs, or who have the capacity to do so, will destroy those long-term aspirations by choosing exclusively on a local basis. Therefore, local choice cannot be glossed as merely suboptimal but otherwise typical voluntary behaviour for humans. If typical humans choose locally at all, it must be relatively constrained lest they become like Jeremy. Heyman may still be right to call local choice voluntary behaviour in order to distinguish it from various involuntary experiences, such as epileptic seizures and anxiety attacks, but it is the most minimal form of voluntary behaviour. A human agent who only chooses locally is agentially impaired.[31]

---

[29] I take it that being assailed by contingencies like this is a familiar experience. Jeremy isn't the victim of an unnaturally tempting environment – this is the kind of environment that normal agents negotiate.

[30] Frankfurt (1971) made a similar point and labelled such agents, 'wanton.'

[31] This is not to say that all local choice will be detrimental, just that it cannot be the consistent, default means of mature human agency. Local choice will be harmless and even beneficial in certain contexts, contexts that the agent can learn and create. One might also object that many choices are best made spontaneously and intuitively.

Now that we see the full implications of chronic local choice we can see that the stakes are much higher than Heyman realises. Chronic local choice doesn't just leave us susceptible to addictive rewards but renders life absent of many of the things that people tend to value the most. This puts pressure on the remaining tools in Heyman's account, global choice and prudential rules, to explain how humans achieve the typical agential control they do. As we will see in the rest of the chapter these tools leave out some important aspects of diachronic agency.

Before we move on, however, there is a mystery for my account of chronic local choice. If chronic local choice is as dangerous as I say, why do people choose locally in the restaurant dilemma and in the other experiments Heyman cites? I speculate that in these experiments people judge, largely implicitly, that it won't hurt to take the cognitively easy local option *within that context*. They thereby preserve cognitive power for other tasks that they judge more important. This is likely to be a beneficial strategy given the ego-depletion literature which suggests our higher cognitive powers draw on a finite resource (Baumeister, 2002; Baumeister et al., 1998). I suspect that the agent develops a skill of knowing when to use cognitive resources through long term inter-subjective training involving (largely implicit) reference to their values and context. If this is true then we wouldn't expect agents to exert diachronic control in cognitive science experiments because such contexts are detached from their lives and so lack the typical cues that the agent uses to know when to exert diachronic control. Furthermore this attitude is usually justified because the experimental choices will have little impact on what the agent values most.[32]

Global choice outside the restaurant and epistemic limitations

It is rare for people to make global choices spontaneously, at least in experimental conditions (Heyman, 2013, p. 437). However, people are more likely to choose globally when a global pattern is scaffolded in a way that reduces the relevant cognitive costs. Rates of global choice are increased, for example, by presenting choices in patterns that are temporally grouped into sets, e.g. three choices temporally close to each other, then a longer time period before another set of three choices (Kudadjie-Gyamfie & Rachlin, 1996). Based on this evidence,

This is true but spontaneous choice is not necessarily local choice. A choice can be made spontaneously yet still be implicitly informed by diachronic considerations. Equally one can carefully, or excessively, reflect on a local choice.

[32] I investigate the role of self-conception and value in diachronic agency in more detail later in the thesis.

Heyman concludes that the low rate of spontaneous global choice is due to it being too cognitively demanding.[33]

This is only part of the picture, however. Another reason global choice, as Heyman describes it, is rare in everyday life is because agents simply lack the necessary information; they face epistemic limitations on top of their cognitive limitations. We see these epistemic limitations when we consider real world choices where the information needed for global choice is much more difficult, or even impossible, to access. In the restaurant dilemma, number values are given to preferences and (dis)habituation rates so that the relevant self-knowledge is made fully transparent.[34] However our desires, especially our future desires, and our rates of (dis)habituation, are *not* transparent. Sometimes we have very little, or even no idea, about what our preferences will be in the future. This epistemic limitation on self-knowledge also applies to knowledge of others and the world. In the restaurant example it is just stipulated that the choice between Chinese and Italian will come up once every day; no other decision can take priority over it because there are no other decisions. Obviously in the real world the possibilities are not artificially narrowed in this way; unexpected options sometimes arise while expected options fail to eventuate. The ability to predict future options depends on the ability to predict the behaviour of others and the world. The further we stray from our knowledge of ourselves and the world, the more difficult a global choice is. At some point, no matter how much cognitive effort the agent invests, they will not be able to come up with a globally rational choice because they lack information. Therefore, global choice is harder than Heyman claims because it isn't just about the cognitive manipulation of available information, it is about predicting future scenarios from limited information and this information can be so limited that global choice becomes impossible.

In defence of global choice one might hope that we sufficiently overcome the epistemic limitations to make enough global choices to avoid Jeremy's predicament. It's true that

[33] Heyman is a little inconsistent, however. For example, he suggests that, "when drug users regret their past behaviour (or anticipate future relapses), they too are taking a global perspective" (2009, p. 128). Regretting past behaviour is common and hardly cognitively demanding but equally it isn't clear that people are bundling together sets of local choices when they regret. Below I suggest that our diachronic awareness isn't so closely tied to global choice as Heyman assumes. In Chapter 4 I argue that diachronic awareness is closely linked to skills of self-narration. Self-narration comes more easily to most people than the accounting required for global choice.

[34] It also becomes possible to arrive at a global choice by mathematical calculation. Explicit decision making in real life involves mathematics when number values are appropriate descriptors, e.g. when we try to work out the most economical option for, say, commuting or shopping. Preferences and (dis)habituation rates, however, are not the kind of things that we usually attribute precise numbers and, even if we were forced to, it's not obvious how useful that would be. In any event, our decision-making, global or local, does not usually proceed by way of mathematical calculation. Perhaps *implicit* decision-making can be understood using numbers values for preferences in a kind of Bayesian analysis. But as long as we see a non-supervenient role for explicit thought, any model of human agency will have to accommodate the non-mathematical form of that thought.

epistemic limitations can be partially managed and controlled. We can limit the temporal extension and/or specificity of our predictions so that the global choices based on them are protected from greater uncertainties. Most importantly, we can selectively invest time and effort in developing knowledge of certain situations. Where we have developed a better understanding of ourselves, others and the world we will be able to make more temporally extended and detailed predictions on which to base global choice. These techniques, combined with naturally occurring cyclic experiences can make global choice relatively easy in some situations; however, in many important choice situations it remains impossible despite our best efforts. Some examples will help illustrate these points.

Natural cycles allow us to develop knowledge about the kind of options that arise and the desires we tend to have in response to them. For example, most adult humans have learnt that they will feel hungrier the longer they go without eating and the more activity they undertake. Conversely, if they eat too much at once the latter part of the meal isn't so enjoyable and they feel ill and unable to do much. Similarly, they have learnt that they will get more tired the longer they go without sleep, but sleep too much at one time and you feel lethargic. In such cyclic cases the possibility of global choice stares us in the face – space out your eating and sleeping patterns to better enjoy life. Cyclical situations also provide ample opportunity for trial-and-error learning which can help develop specificity in global choice. Through some experimentation the agent might settle on eating three times a day and sleeping once a day, having found that other patterns didn't satisfy their preferences so well. One person might experiment with an afternoon nap but find that it throws out sleeping patterns and ultimately reduces the satisfaction of preferences; other agents might find it improves their activities later in the day.[35] When we are presented with novel cycles, e.g. being put on the night-shift, we can begin to develop the knowledge we need and after a while global choice with respect to that situation becomes possible.

Of course, we aren't just at the mercy of contingent cycles; we create cycles that suit our interests, cycles of exercise, practice, socialisation and work. Where we set up those cycles we get the knowledge needed to make better global choices. For example, I find I should go swimming the day after playing soccer because I still need some exercise to concentrate well at my work but low impact exercise is better for my body after the high impact of soccer. Or, I make sure I go camping at least once a month because I have found that that frequency of

[35] If people are still slow to realise the possibility of global choice in these cases, others are often on hand to point it out to them, for example, 'that cake will ruin your appetite,' 'you better get an early night's sleep for your big day tomorrow.' Perhaps Heyman would count these suggestions as forms of prudential rules; in any case it seems that global choice is inter-subjectively scaffolded.

reconnection with the countryside improves my mental health and helps me cope with life in the city.

My intuition is that, because these kinds of example are familiar and common, global choices in these situations are actually very common. The more frequent the cycles, the better our knowledge of the situation becomes, and the more detailed and temporally extended our global choices can be. The opposite, however, also holds – where cyclic experience does not occur and we cannot develop cycles, or do not want to, we are left without the knowledge required to choose globally.

Humans are biological organisms with a finite life span. Our finite lives entail that we have to make many choices that are too rare to be grouped into global choices.[36] Some opportunities only knock once so you have to make your decision based on rough guesses, e.g. job offers, marriage proposals. Some choices the agent wants to ensure are once-in-a-lifetime, e.g. when one decides to marry. Even when somewhat similar choices recur, 'should I *re*marry?' or, 'should we have *another* child?,' important details of the present choice are often too dissimilar to prior choices for our knowledge of prior choice outcomes to be useful. For example, the range of factors relevant to the choice, 'should I have another child?,' are so diverse and variable that there is no point looking for an ideal pattern of child birth.[37] We learn to judge the weight we should give to past choices in present deliberation but we are fallible. Sometimes we assume past choices are relevant when they are not and attempt global choice where we lack information. At other times we judge past choices irrelevant when they would have informed beneficial global choice.

One of the most important factors generating unique choice situations are the trajectories of meaning in our lives. I expand on this in Chapter 4 when discussing narrative agency but, for now, the important point is that "one and the same increment in one's momentary well-being may have greater or lesser effect on the value of one's life, depending on when and how it occurs" (Velleman, 2000, p. 61). This clearly impacts on decision-making. The decision, for example, to go through with a pregnancy is significantly changed by the mother's position in her biological and career trajectories (among other things). Is she 15, 30 or 40 years old? Does she come from a wealthy family or is she trapped in poverty? Even when choices repeat, they

[36] Perhaps if we lived infinitely long lives and had infinitely long memories then we might find that all choice situations recur. If they did recur we might then hope to find global choice patterns for every choice situation.
[37] At least in modern Western society. A nomadic hunter-gatherer, on the other hand, might make a global choice. She might balance the benefit of having a child every year to develop the strength of the tribe with the need for mobility. Because the mother can only carry one child she must wait until her youngest child can keep up with the group on their own. In this case the choice situation might remain constant enough that prior experience is a good guide.

may be significantly different at each time because of their changing position in a diachronic context. In order to make global choices, the effect of overarching trajectories has to be negligible across the series of options chosen.

In summary, global choice can protect us from the damage of chronic local choice but only in cases where choice situations repeat on a regular basis. In this aspect, Heyman's account is doing better than he gave it credit for. He all but wrote off global choice as having a role in human agency but it might actually be quite prevalent in recurring situations. However, much in human life is non-recurrent and in these cases unmitigated local choice will still be dangerous. Consistently making local choices to marry or divorce, raise children or adopt them out, accept jobs or resign from them will leave the agent's life in disarray. Typical agents' lives are not in disarray so it seems they are not choosing locally despite the epistemic challenges of non-cyclic choice situations. Heyman didn't expect global choice to be available in these cases anyway so his account appears no worse off for this critique. Unfortunately the remaining resources that Heyman *does* rely on to do the work of diachronic agency, prudential rules, are also limited in non-cyclical situations.

Prudential rules and non-cyclical choice

The use of prudential rules raises the same problem as global choice in non-cyclical situations because they necessarily have a general application. Of course it should be no surprise that prudential rules inherit the limitations of global choice if we assume that they are essentially a subset of ancestral global choices that have been popular enough to remain in circulation. Indeed, they are likely to have a *more* general focus than the average global choice because, to survive, they need to appeal to a wide range of agents.

An example can bring out the limitation of prudential rules in novel situations. When faced with a non-cyclic choice such as, 'should I marry *this* person?' there is no prudential rule specific to that person, e.g. 'don't marry Brian.' In any case, your own idiosyncratic characteristics would entail that specific rules made by others were irrelevant; one could justifiably think, 'in my case it's different.' That is not to say that a variety of prudential rules applicable to *general* situations that cover this case are useless. Rules such as, 'marry somebody you love,' 'live together first,' and 'don't marry too young,' help guide the decision, but their generality necessarily leaves the choice underdetermined. The agent might not love their fiancé, may not have lived with them, and they might both happen to be young, yet they might still be better off married because of factors specific to their case. Perhaps the

marriage protects one of them from being deported to a dangerous country. Likewise, they might love each other, live together, and be middle-aged but marriage might not be the best decision. Perhaps the formal ties of marriage would ruin the excitement of the relationship.

Heyman's account of typical agency therefore has a significant explanatory gap in it. Typical agents make diachronically informed decisions in non-cyclical choice situations but neither global choice nor prudential rule following can account for this kind of diachronic agency; something is missing from the picture. This explanatory gap is filled, at least in part, by what I refer to as singular diachronic goals.

Singular diachronic goals

Singular diachronic goals are novel, future states of affairs that agents cannot bring about just by action at a time; they require diachronically organised action across time. Such goals are everywhere in human lives, for example, whether to further a certain career, skill or relationship, what new meal to cook, where to go on holiday, or purpose for one's life. When agents pursue singular diachronic goals they do so despite the epistemic limitations that prevent global choices. To better define singular diachronic goals I distinguish them in turn from local choice and global choice. In so doing we find that pursuit of singular diachronic goals depends on what Michael Bratman refers to as our core features of human agency: our planfulness (which includes means-ends coherence, diachronic stability and ends-ends consistency) and our conception of agency as temporally extended (2007).[38] These features of agency entail an awareness of diachronic context and a capacity to imagine novel futures despite the epistemic limitations that prevent global choice.

When we make singular diachronic choices we face issues of diachronic coordination that don't arise for local choices – we have to plan. For example, I am at work and I decide I want to eat chocolate after dinner tonight. I don't have the desire to eat chocolate now but I know that I often have that desire after dinner. I also know that if I don't get the chocolate ahead of time I am usually too lazy to go out and get some and so my desire will be frustrated if I do not prepare in advance. To make sure I don't forget, I put chocolate on my shopping list, when it is time to leave I find my bike, I peddle my bike to the supermarket, I find the chocolate aisle, I pay for the chocolate, and so on. My goal of eating chocolate in the future

[38] For Bratman these core features of agency are intimately connected with forming and acting on intentions. In Chapter 2 I will go on to argue that singular diachronic goals, rule-following and patterns of global choice all need to be intentions if they are to have diachronic stability.

entails a number of interrelated sub-goals that have to be correctly ordered across time. If I cannot develop appropriate means to access the chocolate I won't be able to succeed in my goal. That's to say, singular diachronic goals involve some degree of means-ends coherence; without committing to some set of coherent means one cannot really be said to have the end (Bratman, 1981).[39] Our singular diachronic plans also need to be *ends-ends consistent*; if I try to become both a professional football player and a concert pianist I will fail at both. There just isn't enough time in the day to achieve both ends; trying to do both will undermine each end. This issue doesn't usually arise for local choice because the choice tends to lead immediately to the goal. For example, a child is allowed one of two desserts, ideally he would like both but his parents won't let him.[40] The child picks a dessert and even though he regrets not being able to have both, his choice has guaranteed one of them. Of course equivocation may still prevent him from getting either dessert if his parents say, 'sorry you were too slow to choose.' However, there is much less scope for such ends-ends inconsistency in the short timeframe of local choice.

It normally goes without saying that agents who undertake singular diachronic goals see themselves as temporally extended agents. "In the middle of the project I see myself as the agent who began the project and (I hope) the agent who will complete it. Upon completion, I take pride in the fact that *I* began, worked on and completed this essay" (Bratman, 2007, p. 28). Heyman seems to assume this diachronic self-conception when considering global choice because only a diachronically extended agent could benefit from the global increase in preference satisfaction. As Millgram notes, "for first-person practical deliberation to have a point, the deliberating agent must be presumed to be around in the future in which the plans and policies that are deliberatively arrived at are to be implemented [or benefited from]" (1997, p. 66). However, such a diachronic self-conception is not needed for local choice because the reward is experienced almost immediately by the chooser. Therefore Heyman's claim, that most agents choose locally most of the time, seems to come close to saying most agents don't have much of a diachronic self-conception and that just seems false.[41]

To commit to singular diachronic goals, design coherent means and ensure ends-ends consistency we also need awareness of our diachronic context.[42] For example, starting a plan

[39] There may not be a clear point in the continuum of means-end coherence that divides local choice from singular diachronic choice but local choice requires extremely little planning.
[40] When Buridan's Ass cases arise in local choice, agents typically have some heuristic to break the deadlock.
[41] In the next chapter I argue against George Ainslie's attempt to develop diachronic agency out of time-slice agents.
[42] At the most basic level we need to judge whether there is time to achieve a goal before we start, or, during a plan, to judge whether we have mismanaged our time and need cut our losses and change goals.

to become a professional football player is not a wise goal if one is already 30 years old. It seems humans have some capacity to imagine non-cyclical, novel futures because they do so in the everyday planning situations where they form singular diachronic goals. Awareness of our diachronic context requires not only memory but anticipation of the future. We know not to try and become a professional football player at the age of thirty because we don't anticipate becoming quicker and more nimble as we age. This isn't because we have been through the cycle before but because others have and we have heard their stories. But because our lives won't be just like those of others we need to incorporate our personal idiosyncrasies into our narrative projections. I develop this view further in Chapter 4 but the idea is that these narratives give us a way of predicting and manipulating novel futures.

In successfully pursuing singular diachronic goals we prove that we don't need to be able to bundle future local choices into groups to predict and control our futures. It's worth noting that these diachronic skills of planning unique possible futures and awareness of diachronic context are often quite easy. It's not unusual to find ourselves reminiscing about the past or day-dreaming about our futures; sometimes we do it without even trying. Of course our day dreaming may be fairly inaccurate but with a little more effort we can canvas a range of possible plans for this evening, the weekend, next week, the holidays, retirement et cetera. We can then choose and successfully pursue one of those unique options despite the unfolding set of actions required to meet the goal being novel to us in many ways.

Furthermore, singular diachronic choice requires diachronic stability; we need to act on such choices even when they are no longer supported by the strongest desire of the moment. If my desire for chocolate does not remain my strongest desire throughout my voyage to get it, say because I am distracted or tempted by something else, then I risk failing to follow through on the goal. The challenge is magnified as goals become more diachronic, for example, pursuit of a career or maintenance of a happy marriage require *consistent* efforts of diachronic agency. As it happens these longer-term goals also tend to be some of our most highly valued. I don't really mind being distracted from my chocolate eating plan when I run into a friend say - no doubt I'll enjoy some later - but ruining a valued career or marriage through excessive local choice would be much more upsetting. Local choice does not require diachronic stability but this is exactly what makes it insufficient for healthy agency.[43] Much more can be said about

[43] Global choice on the other hand does require diachronic stability because if the agent changes their global pattern of choice as they go they risk undermining the benefits.

diachronic stability and I return to it in the next chapter where I discuss intentions, pre-commitments and intra-personal bargaining.

Distinguishing singular diachronic choice from global choice is much simpler than distinguishing it from local choice. Singular diachronic goals are different from global choices because they target novel future goals, not patterns of familiar choices. When deciding whether to marry a particular person there is no global choice one can make because marriages with this person, and the alternatives – remaining unmarried, marrying other people – just aren't cyclically recurring options. An agent cannot hope to maximise preference satisfaction by finding just the right balance of marriage to this person versus the other options. Similarly, when deciding whether to have a third child an agent cannot maximise preference satisfaction by having a third child three days of the week and then going without the other four days. The agent either has the child or she doesn't, and either way she commits to the lifetime of consequences. She can still make a diachronically informed decision despite the lack of a global pattern because she can (at least roughly) imagine non-cyclical futures. As she imagines the two possible futures, married or unmarried, third child or no third child, she weighs them up to see which will best realise her values and integrate with the other long-term commitments in her life.

In summary, singular diachronic goals are a distinct feature of typical human diachronic agency that depend on planning skills (including imagining unique possible futures) and awareness of one's current position in a temporally extended life. I now return to consider prudential rules in more detail. Prudential rule following is less demanding than global choice but upon close inspection I find that competent use of prudential rules requires the same diachronic skills just highlighted for singular diachronic goals.

Prudential rules and diachronic cognition

I agree with Heyman that prudential rules[44] are an important source of diachronic control because they set beneficial parameters for local choice (and singular diachronic choice[45]). This point holds despite my argument that they underdetermine choice in non-cyclic situations. However I disagree with Heyman's claim that prudential rules require almost *no*

[44] For simplicity I will limit my discussion of prudential rules to those that take a clear rule-like form, e.g. 'look before you leap,' as opposed to social role and identities, e.g. 'mother' and 'religious.' However, I assume all the points I make apply to prudential rules considered broadly.

[45] Over the next two chapters I investigate the importance of mutually supportive plans and rules (or policies).

diachronic cognition. Here I argue that competence in rule-following depends on a range of diachronic skills.

First, although the cognitive demands of prudential rule-following are less than global choice, they still require the same kind of reflection on diachronic benefit. Heyman plausibly suggests that agents initially adopt prudential rules because social pressures make following them locally attractive. Presumably, if social pressure was maintained through the threat of immediate praise or punishment, the agent could follow those rules indefinitely while only using local choice. They would then make the maximum saving in terms of cognitive effort. Clearly, though, agents develop beyond this point and Heyman claims this is because the agent comes to see the global benefits of rule-following for themselves. This also seems plausible; but what Heyman doesn't acknowledge is that to judge the benefit of a rule for himself, the agent must take the total benefit of a series of past rule-following actions and compare that with the total of a series of actions without following the rule. This is a complex cognitive task but agents typically learn to do even better than that. Sometimes agents adopt rules because they judge that the benefits of following the rule are (or will be) better than a series of counterfactual events where they didn't follow the rule. If the agent can think in this abstract and diachronic fashion then he can choose beneficial rules without having to suffer the negative effects of trialling detrimental rules. We can see then that adopting and rejecting prudential rules so that they are most beneficial to one's idiosyncratic situation requires an analysis of diachronic benefits just as when choosing globally. Prudential rules do, however, allow for savings in diachronic cognition because this diachronic cost-benefit analysis only needs to be made at rule adoption or when rules clash.[46] Global choice, in comparison, requires such a diachronic cost-benefit analysis every time it is deployed (unless the agent turns the results of global thinking into a rule).

Second, typical prudential rule-following depends on the diachronic cognition highlighted above for pursuit of singular diachronic goals – planning and awareness of one's temporal extension and diachronic context. Heyman tells us that once we are following a rule we need only assess the temporally local situation to see whether it applies or not. But this is only true of absolute rules that hold at all times, such as, 'never drink.' Other rules require an awareness of diachronic context to know how to implement them and, the better the diachronic awareness, the better they can be implemented. Take the rule 'all things in moderation'; to know whether you have indulged in a moderate amount of a substance and might be about to

[46] Although, there needs to be some diachronic monitoring in order to notice if a rule has stopped providing benefit.

do it excessively, you need to know how much you have already had. This might be relatively synchronic such as when eating from a packet of chips, but moderation in some substances might be measured in frequency per week, or month. An agent can't follow this rule without remembering how much of a substance she has already had over a certain timescale. Furthermore, she might do even better at following the rule if she can anticipate future occasions when, say, fast food is likely to be particularly beneficial or detrimental. For example, anticipating having no time to cook tomorrow, an agent might decide not to eat fast food tonight since it would be better to have it tomorrow. If does eat fast food tonight then she will put her rule-following in jeopardy by exposing herself to strong temptation. The kind of diachronic planning that was needed for singular diachronic goals is, therefore, relevant to proficiency with many prudential rules. Sometimes these anticipated future situations will be unique but agents typically manage to apply the correct rules anyway. Meeting the in-laws for the first time, for example, signals the need to strictly enforce one's rules of politesse. On a related point, some rules should only be applied in certain diachronic contexts, for example, the rules associated with Ramadan and Lent, 'no alcohol before 4pm', or 'stay active in your retirement.' The agent needs to be aware of the time of day, year, or of his life to follow these rules competently. We also need to be alert to others' diachronic contexts to know how to follow rules such as, 'always be polite' and 'love thy neighbour,' because polite conversation topics, whether to congratulate or commiserate, the kind of practical and emotional support needed, et cetera, all depend on the those contexts.

Third, like singular diachronic plans, one's adopted prudential rules sometimes clash with each other; they can be ends-ends inconsistent. For example, if one suffers an injury then the rule, 'exercise three times a week,' might clash with the rule, 'don't exacerbate an injury.' In this case the agent can keep both rules but he must work out which rule should trump the other. To decide on the best hierarchy he needs to compare the diachronic benefit profiles of each possible hierarchy. If such a clash has not occurred regularly before then he finds himself imagining a variety of unique possible futures in order to make his decision.

Finally, like singular diachronic plans, agents need diachronic stability in the rules they have judged to be beneficial. If the agent decides to follow the rule 'no alcohol before 4pm' but sees that a large benefit will derive from drinking today at noon they might think that the situation demands an exception. Perhaps it does; but if they make an exception *every* day then they aren't following the rule at all despite knowing in a non-tempted moment that the rule will maximise benefits. In summary, prudential rules offer the benefits of global choice with a

lower cognitive cost; however, to be used competently, prudential rules still require a wide range of diachronic cognition including those skills used in planning and global choice.

Implications for Heyman's account of addiction

Recall that Heyman essentially explained addiction in terms of the failure to follow prudential rules. If you follow prudential rules that protect against addictive rewards you won't engage in self-destructive local choice. If you don't have such rules then exposure to addictive rewards will all but guarantee your addiction. This chapter has raised some serious concerns for his view. I canvas those concerns here but I delay my main argument to Chapter 5 while I develop an account of human agency that can provide a competing explanation.

Most addicts will be well aware of both the greater rewards associated with quitting and the prudential rules that would help. In fact, many addicts go through cycles of using for several months and abstaining for several months and so, arguably, the agent should have the information needed to make a global choice about the pattern of use that maximises reward. However, people go back to using even though they know it will be detrimental.

> "I lost it all. I lost my house and my wife and kids and I finally got my son back and it's happened again. I met somebody else, had another child and it happened again. …Another separation and lost another child, another partner because of alcohol" (FAL-004, our interviewee[47]).

In any case, just telling an addict to follow the rule 'don't use drugs' or pointing out the benefits of a drug-free lifestyle are rarely effective. Yet, according to Heyman, this knowledge should be all they need for recovery. It is also puzzling on Heyman's account why somebody should relapse after an extended period of abstinence. The agent must be following beneficial prudential rules to abstain and the benefits of those rules should reinforce them. They are familiar with the costs of addiction. The only plausible reason to stop following the rules, on Heyman's account, is that non-drug incentives have disappeared and so the rules have stopped providing benefit. This happens to some unlucky folk but there are still plenty of cases where the relapse occurs while non-drug using incentives remain.

> "I was cool for about six months, then ran into somebody that I hadn't seen for a while, had a taste, that was pretty much it. That went on for another 18 months,

[47] Throughout the thesis I draw on interview material from the ongoing Australian Research Council funded discovery project, 'Addiction, moral identity and moral agency' (J. Kennett et al., commenced 2010).

> …There was a recurring thing for 12 years where it would be fine for maybe nine months, 18 months, but then I'd run into somebody and make the mistake of having a taste. Sometimes it wouldn't lead into a habitual use again, but more often than not it would" (Interviewee 104, our pilot study).

Given the chronicity of many cases of addiction and that relapse is a common occurrence, it seems there must be factors other than knowing prudential rules that explain addiction.[48]

This chapter has raised a number of candidates that have the potential to explain at least some of these cases of addiction. Recovery from addiction is a kind of singular diachronic goal, as are such things as careers and intimate relationships that tend to characterise healthy lives and are often absent from addicted lives. So perhaps addicts are struggling to create or pursue singular diachronic goals. If so, the solution may be to target the underpinning diachronic skills: means-ends planning, diachronic stability, ends-ends consistency, and awareness of one's temporal extension and diachronic context. Because these skills also underpin successful prudential rule-following, limitations in these domains may also explain why some people manage to follow protective rules but others just can't seem to put them into practice.

Finally, on Heyman's account, addiction, recovery and relapse appear to depend solely on contingent aspects of one's environment – exposure to behaviourally 'toxic' addictive rewards in the absence of prudential rules or normative pressure to begin following those rules. By offloading the explanation for addiction from agency to extra-agential forces, Heyman's position ends up much closer that of the 'brain disease' theorist than he might like. Brain disease theorists will be more than happy to champion the role of 'toxic' reward types causing addiction.[49] After all, certain rewards can only be 'toxic' if we have a neurology that responds to them in those detrimental ways. So the distinction between Heyman's position and the brain disease camp essentially boils down to a disagreement over the role of prudential rules and social pressure to follow them. Heyman claims these are independent factors while the 'brain disease' theorist hopes to reductively explain the inability to follow prudential rules as further effects of 'toxic' rewards. In neither theoretical position does the agent play a crucial

[48] In the face of this we might be tempted to revert to a physiological model of addiction and think that the drug-using behaviour of these hard core addicts is not action at all but non-intentional. However this is an unattractive option because of the weight of evidence that supports the intentionality of these actions (Levy, 2006). Although, as Kennett (2013a) argues, we may opt for a physiological explanation of behaviour while maintaining that the behaviour in question still counts as intentional.

[49] Heyman makes a general case for the distinction between addictive and non-addictive reward types and explains that certain drugs tend to cause addictions because they are the most toxic (2009, pp. 149-150). However, when we look at some of the more exotic targets of addicted behaviour, e.g. shopping, tanning, video game playing, the distinction either becomes harder to defend or we need a new term for these other kinds of addiction that don't seem to stem from toxicity.

role in addiction onset, recovery or relapse; they are swept along by extra-agential forces.[50] Heyman's claim that addiction is a disorder of choice is, therefore, misleading. On his account, the choices involved in addiction are no different in kind to those of non-addicted lifestyles. My analysis, in contrast, begins to uncover conceptual space for qualities of agency to have an explanatory role in addiction; addicted agency is impaired in some way that non-addicted agency is not. By acknowledging agential impairment, treatment does not have to be paternalistic; the addict can play an important role in their recovery and be held somewhat responsible for their relapses.

Conclusion

Heyman's view of typical agency is that people choose locally most of the time but under the varying constraints of prudential rules. Those prudential rules are applied synchronically and their uptake depends on contingent social pressures. The only difference between non-addicted and addicted agents is that addicted agents have been exposed to addictive rewards while lacking protective prudential rules to follow. As long as addictive rewards remain available, recovery requires adopting these protective rules; addicts that don't recover are not adopting these rules.

I have argued that Heyman oversimplifies typical human agency. By ignoring the ubiquitous dangers of local choice in the human case he fails to see how important diachronic agency is. He considers two forms of diachronic control – global choice and prudential rule following. After ruling out global choice he relies on prudential rules to provide all the diachronic support agents need. However neither rules nor global choice enable singular diachronic plans. Singular diachronic plans are essential for what most people would take to be a good life. They are needed to achieve any unique future goal such as maintaining relationships, pursuing careers, building a house, raising children, et cetera.

Success with singular diachronic plans requires planning agency, i.e. the capacity for developing means-ends coherence, ends-ends consistency and diachronic stability. The agent must see himself as temporally extended and be aware of his current diachronic context. Upon analysis of competent prudential rule-following we find that this too is underpinned by the diachronic skills involved in global choice and singular diachronic plans.

[50] Except Heyman allows, while brain disease theorists do not, that rare cases of global choice would allow the agent to choose or abstain from addictive substance use.

The more complex picture of typical agency developed in this chapter has implications for our understanding of addicted agency, recovery and relapse. Exposure to addictive rewards and prudential rules is only a small part of the picture. We also need to investigate why agents fail to adopt beneficial rules when they are available and why they don't manage to develop singular diachronic plans that they value. This initial critique has highlighted a few places to look: addicts might have difficulty planning (developing means-ends coherence, ends-ends consistency), they may struggle to imaginatively engage with the future, or they may be unable to ensure the diachronic stability of the plans or rules they value.

In the next chapter I focus specifically on this issue of diachronic stability.[51] I draw on the work of George Ainslie and Richard Holton who each offer opposing views on how typical agents achieve diachronic stability and how that is disrupted in addiction. I argue in favour of Holton's view on several grounds. Perhaps most importantly, Holton's account comfortably explains how singular diachronic plans, prudential rules and global choices can all have (and lose) diachronic stability. In contrast, Ainslie's view cannot explain the diachronic stability of singular diachronic plans and, as we saw in this chapter, this is a serious oversight given the value healthy and addicted agents place on such plans. However, I think Holton's view can be supplemented with two aspects of Ainslie's account. First, I agree with Ainslie that pre-commitment strategies are important for diachronic consistency especially in the context of addiction. Second, although Ainslie ultimately makes too much of it, some diachronic consistency can be achieved through treating present decisions as precedents for future decisions.

[51] In Chapter 3, I consider the importance of planning for agency in general and in Chapter 5 I focus on the relevance of planning in addiction and recovery.

Chapter 2: Diachronic stability in action

Introduction

We saw in Chapter 1 that patterns of global choice, prudential rules, and singular diachronic plans are structures for diachronic action; they are tools the agent can use to avoid the chronic local choice that can result in addiction. However, we finished the chapter with an unanswered question – addicts tend to be aware of these tools and use them successfully in various domains, so why don't they use them to recover from addiction?

In this chapter I look at two potential answers to this question, one from George Ainslie the other from Richard Holton. Both realise that just forming diachronic goals will be insufficient if those goals are forgotten or warped by temptation at crucial moments. If I redefine the pattern of global choice I pursue, the rules I follow, or my singular diachronic plans whenever tempted, I will never act in accordance with any of them. Therefore, agents must be able to accord these structures a degree of *diachronic stability*. Failure to achieve diachronic stability appears central in succumbing to temptation and suffering from addiction.

My central argument in this chapter is that Holton's model of diachronically stable agency is more plausible than Ainslie's model. A defence of Holton's view is an important stepping stone in my wider project for two reasons. First, Holton claims an independent role for the agent in setting diachronic goals. I build on this in Chapter 3 where I argue against Ainslie's metaphysics on the grounds that it effaces the agent. Second, in Chapter 4 I claim that self-narratives provide diachronic stability in essentially the same way that intentions do. This ultimately leads to my view that addiction is influenced by self-narrative and the agent can often do something about it.

I begin this chapter by describing Ainslie's explanation of diachronic instability in agency and how it can be countered. His view is a logical extension of Heyman's in that Ainslie believes the agent will necessarily do what she judges will provide the most reward.[1] Unlike Heyman, however, Ainslie is aware of the problem that diachronic inconsistency poses for agency. He describes diachronic inconsistency in terms of an innate tendency, shared by all animals, to hyperbolically discount future rewards. He proposes that hyperbolic discounting is typically overcome through adopting pre-commitments and, more importantly, what I will call 'test-case willpower.'[2] The

[1] Heyman described this as choosing the option which one most prefers or acting so as to ensure the best, typically local, consequences.

[2] Both Ainslie and Holton refer to distinct concepts of 'willpower.' I call Ainslie's version of willpower 'test case willpower' and Holton's version, 'muscle model willpower.'

latter is where one treats a present decision as a precedent or test-case intended to influence the choices of one's future selves.

I largely agree with Ainslie regarding pre-commitments but I present a range of objection against his view that test-case willpower is an all-purpose, flexible skill for achieving diachronic stability. I then outline Holton's alternative account, pointing out how it improves on Ainslie's account as I go. According to Holton, diachronic instability is the result of desires contrary to best judgment (temptations) overwhelming judgment. He claims that diachronic consistency can be achieved by forming intentions, which settle decisions in advance, and wielding a finite resource under executive control called muscle model willpower. Intentions have a number of advantages over test-case willpower: They are versatile, adequately stabilising singular diachronic goals, prudential rules and sequences of choice; they do not rely on arational manipulations; and, they efficiently spread cognitive effort over time and enhance diachronic coordination in action.

Ainslie on diachronic agency

Ainslie is committed to a reductionist account of agency. In his view, agents are born with a range of pre-existing reward expectations which compete against each other to control action. Those reward expectations target such things as food, water, warmth and the presence of one's mother. Reward expectations develop over time to include a greater diversity of targets, some of them much longer-term, but they are always the result of bottom-up processes (Ainslie, 2011, p. 71).[3] Over time, reward expectations group into compatible factions of increasing diachronic stability. Typically they achieve the kind of coherence that we refer to as a person. What one finds rewarding is objectively set; the agent might become more aware of what they find rewarding but they cannot change those fundamental reward expectations. Ainslie avoids all top-down descriptions of agency

[3] Bottom-up processes here refers to the competition between the sub-agential motivations that result in macro-effects such as action and social rules. Top-down processes would include things such as the agent or society controlling sub-agential motivations. Top-down processes don't really occur on Ainslie's view because the agent actions and social pressures are reducible to earlier bottom-up processes.

because he assumes that commits one to seeing the agent as a supernatural homunculus (2005, p. 643).

This subsequently commits Ainslie to a reward[4] theory of motivation whereby the "individual is constrained to choose the option with the greatest expected reward of all those she considers" (Monterosso & Ainslie, 2009, p. 116). Therefore, it is impossible for the agent to be more motivated to pursue a lesser perceived reward over a greater reward when both are available to her. If she doesn't choose something it is because she does not prefer it or believes it to be unavailable. All action is thus explicable in terms of expected utility. This also entails that all rewards are *commensurable* and only distinguished by size and time of availability.[5]

The agent is, therefore, a reward maximiser by definition. Maximising reward would be easy if the apparent sizes of expected rewards constantly tracked the amount of reward they would actually provide. Unfortunately the process is complicated (for all animals) by hyperbolic discounting whereby the *apparent* sizes of rewards are discounted at a rate that approximates a hyperbolic curve (Ainslie, 1975; Green & Myerson, 2004; Green et al., 2005; Kirby, 1997; Mazur, 2001). Hyperbolic discounting fools us into reversing our preferences in a way that undermines reward maximisation. An example can make this process clear. Take two mutually exclusive rewards, a smaller sooner (SS) reward such as going out drinking on Friday night and a larger later (LL) reward such as going skiing on Saturday (which will require getting up early). When both rewards are far in the future they are roughly equally discounted and so the agent knows that the skiing reward is larger. But, as the possibility of drinks on Friday comes temporally close, it becomes significantly less discounted than it was, while the skiing reward remains roughly as discounted as it was. At this time the SS drinking reward *appears* to be greater than the LL skiing reward; Ainslie describes this as temptation.[6] The agent, being caused to pursue that which they think will maximise reward, undergoes a 'preference reversal' and so goes out drinking and sleeps-in on Saturday morning. The appetite for a night out drinking is then temporarily satiated so that the next possible night out is again well in the future (next Friday). The next chance to go skiing is also well in the future (next Saturday). Now that both rewards are again roughly equally discounted, the agent can again see

[4] Ainslie uses the term 'reward' which I take to be synonymous with preference or desire satisfaction.
[5] If rewards were incommensurable then the agent would need an additional means of choosing between them that the reward account doesn't provide.
[6] In the next chapter, I dispute whether this description properly captures the phenomenology of temptation. It seems to preclude the phenomenology of conflict often inherent in temptation. Later in this chapter I argue that hyperbolic discounting underplays the corrupting effect temptation has on judgment.

that the drinking reward is less than the skiing reward. They then regret their action because they realise how it undermined reward maximisation. They might then begin to pursue the next available skiing reward but, without taking any further measure, they will be doomed to suffer preference reversal again.[7] Clashes between SS rewards mutually exclusive of LL rewards are extremely common in the human world. Typical SS rewards include sleeping in, procrastinating, extra helpings of food and drink, et cetera. LL rewards are typically to do with health, career, hobbies, and relationships.[8]

This natural diachronic inconsistency complicates the reductive picture of agency. "The inherent instability of preference creates separate, temporally-defined agents within the unit that classical economics has always seen as basic, the individual person" (Monterosso & Ainslie, 2009, p. 118). However, Ainslie does not subscribe to a pure time-slice view of agents. He thinks that even time-slices have longer-term interests that they want to see realised, e.g. the agent on Wednesday wants to go skiing on Saturday. In some sense, the current time-slice believes they will be around to enjoy future rewards (or they have a strong altruistic relationship with their future selves). Indeed, time-slices must have some interest in the future agent or all diachronic pursuits would constantly be sabotaged.[9] If chronic preference reversal is the default state, what can the current time-slice do to counter the conflicting interests of future selves and develop diachronic consistency in pursuit of the LL rewards that the time-slices share? Ainslie divides the possible techniques into two categories: pre-commitments and test-case willpower.

Pre-commitments

Ainslie divides pre-commitments into external and internal pre-commitments. External pre-commitments include tying oneself to a mast when sailing past sirens, methadone, gastric bypass,

[7] Ainslie assumes that, aside from the temporary illusion created by hyperbolic discounting, the agent's LL goals have a diachronically stable existence. In other words the size of rewards is objectively fixed – the agent doesn't (cannot) do anything to adjust their non-discounted size. In contrast, on Holton's view the agent also plays an important role in creating and maintaining the LL rewards themselves. I discuss the agent's role in creating LL rewards in Chapters 3 and 5.

[8] The conflict can operate on short-term timescales as well. In the last chapter we saw that Jeremy failed to achieve the LL rewards of going for a walk because he got distracted by a book, and he failed to appreciate the LL reward of the book by being distracted by a friend, and so on.

[9] Schechtman argues that an adequate theory of diachronic agency requires that the agent can adopt both a synchronic and a diachronic perspective. "In order to unify ourselves as agents we need to treat past and future selves as others, but to motivate this endeavour we need to think of ourselves as temporally extended agents, and so identify with past and future selves" (2008, abstract). She refers to Ainslie's theory as a case that allows switching between synchronic and diachronic practical perspectives (2008, p. 421).

illiquid investments, eliciting the potential praise or opprobrium of others. I take external pre-commitments to be relatively uncontroversial; however it's worth sketching their main benefits and limitations. External pre-commitments may be fully binding, making it impossible to access an SS reward, or partially binding, making it more costly or less beneficial to access the SS reward. More binding pre-commitments render the agent less flexible in the face of unexpected circumstance. For example, Ulysses tied to the mast simply cannot chase the sirens but he may regret the strength of his pre-commitment if the boat begins to sink. If he had merely made a side bet with another sailor that he wouldn't chase the sirens he would have been less protected from the sirens but more able to deal with other circumstances. Pre-commitments also require a certain amount of prior planning and so the agent can be caught out when temptations arise unexpectedly or when the necessary resources for the pre-commitment are lacking (e.g. methadone is unavailable).

Internal pre-commitments include attentional and emotional control. Attention control involves directing attention away from temptations before they gain sufficient power to control action, either consciously (Metcalfe & Mischel, 1999) or in the implicit Freudian defence mechanisms of suppression, repression, or denial. Emotional control involves cultivating or inhibiting emotions, either consciously or in the implicit defence mechanisms of isolation or reversal of affect. Internal pre-commitments are more flexible than external ones because they can be applied to any form of temptation and they can be applied synchronically, without much preparation. If you accidentally find yourself outside a bakery you might direct your gaze away from the cakes. One might try to see a tempting marshmallow as an emotionally 'cool' puffy cloud rather than a delicious sweet (H. N. Mischel & Mischel, 1983).[10] According to Ainslie, this works because, by controlling your attention and emotion, an SS reward that is temporally close may be made to seem less immediately available and thus more discounted. The main limitation of synchronic attentional and emotional control is that they are not stable over even moderate periods of time; if the SS reward remains available the agent will eventually give in (Monterosso & Ainslie, 2009, p. 121).[11]

[10] The agent still needs to be familiar with the clash between particular SS and LL rewards to be able to distinguish an SS reward that will undermine reward maximisation from one that is compatible with reward maximisation. The first few times you eat the cake you might think it is compatible with reward maximisation. It's only after some experience that you realise the cake is undermining greater rewards of good health and so needs to be resisted. This becomes an issue when trying to diachronically stabilise singular diachronic plans as I argue below.

[11] The role of the synchronic techniques of emotional and attentional control is contested. Holton refers to these techniques as 'muscle model willpower' and takes them to play a much more central role in agency as we will see below.

More diachronically one might cultivate or inhibit certain emotional or attentional responses that will help avoid preference reversal, for example, not letting one-self wallow in self-pity or gaze at cakes. If the agent develops habits of overcoming self-pity and directing attention they will tend to avoid SS rewards without realising it. These techniques are limited in their effectiveness however, because they require long-term persistence to be effective and, even then, it seems that humans just cannot cultivate or inhibit certain habits beyond certain limits. [12]

In summary, external pre-commitment requires foreknowledge of temptations and the more effective it is the more inflexible it renders the agent. Diachronic internal pre-commitments allow the agent to avoid some SS rewards altogether but they must be carefully established through training and practice over long time periods. Synchronic internal pre-commitments are highly flexible but can only secure diachronic consistency for short periods of time and so temptations often exceed them. For these reasons, Ainslie considers pre-commitments alone to be insufficient to secure typical diachronic stability. He goes on to argue that test-case willpower can ensure diachronic stability when pre-commitments cannot or are undesirable.

Willpower – test cases, personal rules and inter-temporal bargaining

Ainslie claims that test-case willpower is "at once the strongest and most versatile" tactic of achieving diachronic consistency (2005, p. 640). To use willpower the agent treats the current decision as a test-case or precedent for all future decisions of the same type. I might be prepared not to show up at work tomorrow for this heroin now, but I'm not prepared to give up work altogether for heroin use. The decision to refuse heroin now sends a bargaining signal to future selves, either increasing or eroding their confidence in intra-personal cooperation. Choosing the LL reward now provides your near future selves with evidence that *further* future selves are likely to be cooperative, so continuing to pursue the LL reward won't be in vain. In contrast, choosing the SS reward (heroin) now provides near future selves with evidence that further future selves are likely to be *uncooperative,* so pursuit of the LL reward will be in vain.[13] Time-slices are in 'limited

[12] Ainslie claims that another limitation of internal pre-commitments is that they can be just as easily used to concentrate on the SS rewards (once preference reversal has occurred) and to develop bad habits that tend to make SS rewards obvious. Internal pre-commitment is therefore only useful for maximising reward if the agent can predict what habits undermine reward maximisation and which SS rewards threaten damaging preference reversal.

[13] Therefore, Ainslie is not endorsing 'one boxer' reasoning because the present choice is still meant to have a causal effect on future action. 'One boxers' choose just the second box when faced with Nozick's (1969) task. You face two boxes and can choose both or just the second. The first box has $1000 in it; the second has either $1 000 000 in it or

warfare' (Schelling, 1960, pp. 20-80), with each other; each time-slice tries to find a balance between accessing SS rewards specific to itself and passing up those rewards in order to cooperate with future time-slices in pursuit of mutually valued LL rewards.

To treat a choice as a test-case the agent must hold two beliefs – that the present choice will be *necessary* in ensuring the same choice will be made in all similar situations and that the present choice will be *effective* in ensuring the same choice will be made in all similar situations. For example, to quit smoking, I must believe that if I smoke *this* cigarette I will always smoke when presented with the choice to smoke (SS reward) or abstain (LL reward). This is the belief that choosing the LL reward now is *necessary* to quit smoking. I must also believe that if I do not smoke this cigarette I will never smoke again when presented with this choice so choosing the LL reward will be *effective* in quitting smoking.

The strength of belief in this connection between the present choice and all future choices of the same kind puts the 'power' in willpower. Once the agent believes that so much reward hangs on this decision, they are motivated by accessing that reward and by the fear of missing out on it (Ainslie, 2011, p. 69).

Ainslie tells us that treating the present choice as a test-case for a particular range of cases is the equivalent of adopting a *personal rule* in that range of cases. For example, treating the present chance to smoke as a test case means that to smoke now is always to smoke and not to smoke now is never to smoke again. If never smoking again is preferable to smoking forever then the agent effectively follows the rule 'never smoke.' But other rules might be formed that allow SS rewards on occasion, such as 'only smoke socially,' or 'only smoke two cigarettes per week,' which might be compatible with also accessing the later rewards from low levels of controlled smoking.

The exercise of test-case willpower, or 'following' of personal rules, is usually implicit.

> "…The intrapersonal bargaining situation is usually not perceived in explicit terms and the personal rule that provides a truce line is intuited rather than stated. The result is that an individual looks as if she is following principles, but except in the most deliberate cases

nothing. An extremely reliable predictor has put the money in the second box if they have predicted you will choose just the second box and nothing in it if they have predicted you will choose both. But at the time of your decision they have already acted. A 'one boxer' chooses just the second box since that provides evidence that they are a one boxer and so the predictor would have placed the money inside. 'One boxer' reasoning appears irrational because, although it arguably provides *evidence* that the money will be there, it has no causal impact on the outcome. A 'two boxer' assumes that what is done is done and so you may as well take both boxes.

> usually cannot put the principle into words. She completes a chore and feels virtuous, or fails to get out of bed at the usual point in the program on her clock radio and feels a vague sense of loss, but in neither case consciously thinks she is testing a principle" (Monterosso & Ainslie, 2009, p. 124).

Therefore, most of the time, we exert test-case willpower, or fail to, without realising it. The rules we appear to follow are the result of bottom-up processes of competition between reward expectations. Sometimes we recognise the patterns in our behaviour and in that subset of cases we can verbalise the rule we are following.[14]

Test-case willpower involves recursive momentum because the agent's belief that this choice will be a precedent is influenced by their prior decision history. Thus, "every lapse reduces your ability to follow a personal rule" (Ainslie, 2005, p. 644) because confidence in future selves is decreased. One begins to think, 'I'm not the kind of person who will follow this rule in the future so there's no point trying now.' On the other hand, every observance of a rule increases your ability to follow that rule. One begins to think, 'I am the kind of person who follows this rule which means I will access LL rewards as long as I continue to follow the rule.' Therefore "…the more you believe that you will keep [a rule], the more you *can* keep it, and the more you will subsequently believe; the less you believe you will keep it, the less you can keep it, and so on" (Ainslie, 2011, p. 68).

If the agent faces a runaway decrease in confidence they can halt it by defining the specific details of the case where they lapsed. This is called creating a 'lapse district.' For example, if I had a panic attack while public speaking and that panic threatened to recursively undermine my confidence in all social encounters, I might say, 'I just can't speak publicly.' This draws a line in the sand where I give up on having confidence in public speaking in order to prevent expectations of panic in other cases. Ainslie can therefore explain the domain-specific failings of diachronic consistency with reference to the agent's domain-specific decision history and their lapse districts.[15] These recursive effects driven by decision history raise the question of how an agent would ever *reverse* a runaway process or remove a lapse district – I address this concern below.

[14] Ainslie position here is ambiguous. He sometimes describes the 'rule' as merely describing the ongoing bargaining process that happens to have a rule-like consistency. At other times, as we will see, he appears to assume that time-slices just follow (or rationalise) the rule rather than engaging in the thought processes of bargaining and precedent setting independent of whatever 'rule' may have been followed in the past.

[15] Recall that Heyman appealed to knowledge of prudential rules to explain domain specificity in control. Here we see that Ainslie can use lapse districts in personal rules to explain the same phenomena.

Underconfidence, overconfidence and rationalisation

When using willpower, diachronic stability is not threatened by SS rewards as much as by detrimental self-expectations (Monterosso & Ainslie, 2009, p. 125). Those dangers can be broken into overconfidence, underconfidence and rationalisation. If the agent is over-confident, then they incorrectly believe that their later selves will achieve an LL reward despite present choice of an SS reward. In fact, their choice of the SS reward now leads their future selves to choose SS rewards. If the agent is under-confident, then they incorrectly believe that their future selves will choose SS rewards even if they choose the LL reward this time. So they choose the SS reward now. In fact, if they had chosen the LL reward, it would have led their future selves to choose the LL reward. In each case they lack the beliefs needed to treat the present case as a test case and so they choose the SS reward now to their long-term detriment. Notice that if they were *correct* in their assessments of their future selves, then choosing the SS reward would be a good decision because it means they either access the SS now *and* achieve the LL reward later (accurate self-confidence) or they take the SS now knowing that the LL was never achievable anyway (accurate acceptance of one's limitations).[16]

Willpower can also be undermined by rationalisation. In rationalisation the agent either convinces himself that the present choice scenario is different in kind to the one where their personal rule applies, e.g. 'I can drink today because it's my birthday,' or, that the rule actually admits of the present SS reward, e.g. "my rule, 'only two drinks per day,' allows two pints of scotch." Personal rules can be more or less 'bright.' It is more difficult to rationalise exceptions to brighter rules. For example 'never drink' sets bright lines while 'only have two drinks' is less clear because the size of the drinks or the time between sets of two drinks is open for negotiation.[17] When lines are less

[16] This situation leaves room for another form of judgment error that Ainslie doesn't mention. The agent might overestimate the effect of their test-case willpower. They may judge the choice of an LL reward now to be sufficient to ensure future choices of LL rewards when in fact it won't be. When their future self comes to choose they just don't happen to be so impressed by that prior choice (assuming that there are more factors than recent choice history affecting future choice). In other words, just because we are convinced that a choice counts as a precedent today that doesn't guarantee our future selves will see that choice as a precedent or that they will see the similar choice they are facing as a precedent for future choices. Conversely they might despair that their imminent choice of an SS reward will ensure their future choices of SS rewards when in fact it won't. Although this shouldn't be possible on Ainslie's account because that very thought should cause the agent, being a reward maximiser, to choose the LL reward in the present. So willpower requires that the agent judge what to expect from their future selves and just how significant their current choice may appear to those future selves.

[17] This correlates with the observations from Chapter 1 that prudential rules that apply at all times such as 'don't drink,' require less diachronic cognition than those that only apply in specific situations. "Bright" rules include those that apply at all times and so one reason why they are less vulnerable to rationalisation may be that they are less cognitively demanding to apply.

bright there is more room to rationalise damaging exceptions to the rule; however such rules also give the agent the flexibility to take occasional SS rewards that do not undermine the LL rewards and thus to get the best of both worlds.

Negative side-effects of test-case willpower

Ainslie is clear that test-case willpower isn't an unmitigated good. Here I outline three of the more important downsides. *Rigid behaviour*: the agent can fail to live in the here and now because they connect the current decision so strongly to future trajectories. Rigidity is the flip side of rationalisation in that the agent fails to realise that certain cases *should* be exceptions. *Lapse magnification*: Because violating a rule is evidence that one won't be able to follow that rule in future circumstances, small violations can result in complete collapses in self-confident rule-following. This explains 'abstinence violation effects' (AVEs), where those trying to abstain follow an initial lapse with a binge.[18] *Self-deception*: "Personal rules depend heavily on perception--noticing and remembering your choices, the circumstances in which you made them, and their similarity to the circumstances of other choices" (Ainslie, 2005, p. 644). There is a short term motivation to ignore lapses because that way no efforts will be made to stop it. There is a long-term motivation to ignore having made a lapse since noticing would initiate the recursive erosion of willpower (you must also fail to acknowledge that you have wilfully ignored the lapse). Self-deception explains why money goes missing despite a budget and how people can put on weight despite claiming to eat little. Given these downsides, how can an agent use willpower safely?

Improvement of test-case willpower

Ainslie holds that test-case willpower can be improved to provide the benefits it offers while minimising the detriments. The treatments Ainslie considers to best exemplify what we might call, 'willpower therapy,' for addiction are Twelve-Step programs. Setting aside the disputed clinical efficacy of Twelve-Step programs, such programs might influence willpower by countering overconfidence, underconfidence and rationalisation. Twelve-Step members declare powerlessness over their addiction. This protects against overconfidence; a powerless agent won't

[18] For examples of AVEs, see (Collins & Lapp, 1991; Grilo & Shiffman, 1994; Hudson et al., 1992; W. G. Johnson et al., 1995; Marlatt & Gordon, 1980; Shiffman et al., 1997; Spanier et al., 1996; Ward et al., 1994).

think that they can take the SS reward now and leave their later selves to achieve the LL reward. Underconfidence is countered by replacing ambitious LL goals with more believable ones, e.g. 'one day at a time.' Confidence builds up as the days accumulate and a tally of sober days is kept by the member to serve as an explicit reminder. Because the increasing tally might lead to overconfidence again, the member is told that, 'every day brings you one day closer to your next relapse.' This highlights that test-case willpower is a constant balancing game – you must build some confidence in the cooperation of your future selves if willpower is to provide any diachronic stability. But you cannot build too much confidence because that will undermine diachronic stability. You must maintain a healthy doubt in your future self's ability because only then will your current decision to pursue the LL reward be necessary and effective. Rationalisation is countered by using rules with bright lines; abstinence must be absolute. There is no way to hedge in an exceptional drink because, 'one drink is a thousand drinks,' or, in other words, be afraid of a lapse because it will certainly lead to an abstinence violation effect (AVE). In the event of a lapse, of course, the agent is encouraged to forget the lapse, to try not to 'beat yourself up,' and to focus on one day at a time again. There is no attempt by such programs to counter excessive rigidity but that isn't surprising since, in these cases, rigidity is the lesser of two evils; yes, you might miss out on a couple of drinks on genuinely exceptional occasions but that's better than risking relapse. Based on this, we might conclude that addicts are less skilled in using willpower than non-addicts. If they developed this skill then they would better maximise reward by avoiding overconfidence, underconfidence, and rationalisation. Ultimately they would also be able to stop relying on heavy-handed pre-commitments.[19]

In summary the agent needs test-case willpower and pre-commitments to achieve diachronic stability in the face of hyperbolic discounting. But since pre-commitments are frequently unavailable, ineffective or undesirably inflexible the agent typically relies on test-case willpower. Unfortunately, using test-case willpower exposes the agent to certain downsides so they must use it carefully. If the agent can develop their test-case willpower there is hope that they can, for the most part, gain the benefits and eschew the dangers of rigidity, rationalisation, underconfidence (with associated AVEs), and overconfidence. An agent with willpower can more consistently

[19] There are reasons to suspect there is more to overcoming addiction than skill in test-case willpower but I delay these arguments until Chapter 5.

follow personal rules with less bright lines and thus access genuine exceptions to their personal rules and better maximise reward.

Analysis of Ainslie's Account

Objection 1: Test-case willpower entails false beliefs

One might object to test-case willpower on the basis that it requires the agent to hold false beliefs. Recall that to exercise willpower the agent has to believe that resisting the current temptation will be both effective and necessary in ensuring future pursuit of the LL rewards. The problem is that the agent can have good reason to think that these beliefs are false. For example, smokers often only give up after failing to give up several times, so not only do they have to believe that they can give up despite failing in the past, but prospective quitters now see that several failures are likely before success. The belief that not smoking now is necessary to give up therefore appears false (Holton, 2009, p. 118, note 9). There is also reason to believe that not smoking now will be *in*effective in stopping smoking later because abstaining gets harder as time goes on, at least at the beginning (ibid). How can the agent use test-case willpower in the face of the evidence against holding the required instrumental beliefs?

Ainslie can respond as follows. The beliefs one has to have are not outright false. Even if most people fail to quit before later succeeding and even if the agent has tried and failed in the past, it is still possible that *this* choice will be a necessary and effective precedent. It is therefore part of the skill of willpower to convince yourself of the truth of certain beliefs *despite* the contrary evidence. This might not be blatant self-deception because of new factors entering the equation. Perhaps this time the agent will benefit from prior experience of trying to quit smoking – they will use nicotine patches, or they will think, 'this time I am older and wiser.' Of course, if one fails to quit *again*, the negative effects of test-case willpower come to the fore. The strength of belief in the test-case will cause an AVE. The agent then has to go about *weakening* their belief that the test case necessarily and effectively condemns them to smoking forever. In other words, when strengthening the necessary beliefs one should ignore the fact that you or others have failed before, when weakening those beliefs one should focus on the fact that people have succeeded in regaining control after such failures. So getting benefits from willpower and avoiding the detriments depend

on being able to manipulate one's beliefs about the likely cooperation of your future selves through selective consideration and weighting of evidence; it doesn't require outright false beliefs.[20]

Objection 2: Test-case willpower does no work

Sometimes Ainslie makes it seem as if test-case willpower is entirely determined by the agent's decision history. This would be problematic because, in that case, willpower would merely be a way of explaining how decision histories influence action and so do no additional explanatory work. Ainslie can sidestep this objection if he describes willpower as belief manipulation that is only partially influenced by decision history.

Folk observations suggest that agents' decision histories *underdetermine* their subsequent choices. For example, a history of successful pursuit of certain LL rewards, such as a career, might result in any of the following: continued success in that career; excessive rigidity in pursuit of the career so that the agent fails to enjoy SS rewards that are not really mutually exclusive of career success, such as occasional holidays; or, overconfidence causing detrimental choice of SS rewards that are mutually exclusive of career success, such as coming to work late every day. To take another example, a history of successful career pursuit followed by a lapse, such as losing one's temper at the boss, might result in any of the following: an AVE resulting in a total loss of control in pursuing the career; the continued pursuit of the career but with the creation of a new lapse district, 'I cannot keep my temper in such conditions;' or, the agent could just brush off the lapse leaving their confidence in career pursuit undisturbed. Finally, a history of SS reward choices that has continually undermined career development might result in either of the following: permanent underconfidence and so never trying to pursue a career; or taking that first step in trying to launch a career again. If in all these cases the agent's decision history doesn't determine their subsequent behaviour, then some other factor(s) must be at play.

Ainslie could argue that variation in skills of belief manipulation can explain that variation in outcome. For example, after several rule-following successes it might be better to focus on the role of the current choice as a precedent rather than getting carried away by your good decision history

[20] This provides a way of understanding the resigned attitude of some addicts. After several cycles of building up a belief in a precedent only to have to give up that belief later on, one will start to struggle to see anything different about the present circumstance in order to build up belief in a new precedent yet again. There's only so many times one can convincingly say to oneself, '*this* time it really will be different,' when you know there isn't really anything different at all.

which might lure you into overconfidence. If you have just had a lapse after a period of steady rule-following it is better to focus on your total decision history where this lapse is just a blip rather than consider that lapse as a precedent which would encourage AVEs. When in low confidence after failed rule-following it is better to just try and set a new precedent without focussing on your total decision history. But once you begin to successfully follow the rule again then it will be beneficial to focus on the decision history since the latest precedent.

Objection 3: Test-case willpower is arational

This, however, leads us to a general concern, raised by Bratman, that test-case willpower is an *arational* way of controlling behaviour; setting a precedent doesn't give your future self a *reason* to act, you merely change your future self's affective situation.

> "The belief, on day 2, that may well be to some extent confirmed by the choice to abstain on day 1 is the belief that I tend to abstain. *But that is not a belief that normally gives me practical reason, on day 2, to choose to abstain on day 2*. The belief that if I had it on day 2 would give me reason to abstain on day 2, is the belief that if I abstain on day 2, then as a result I would continue to abstain. The problem is to say why that choice is confirmed by my choice to abstain on day 1" (1996, p. 300, my italics).

Test-case willpower can effectively increase the confidence of future selves so that they pursue LL rewards (or decrease confidence in cases of lapses). But, "to see my earlier failure as ineluctably leading to my later failure is to be guilty of a kind of 'bad faith' rather than to be functioning as a rational agent" (Bratman, 1996, p. 302). The same could be said if I see my future success as ineluctably leading from my past success. Test-case willpower is a means of generating arational psychological effects to our longer-term benefit but if we want *rational* diachronic stability in agency we need something more. Ainslie, of course, may not care about this; if we ultimately maximise reward then it doesn't matter if our technique fails to meet some other putative standard of rationality. In other words the reason, 'because it will maximise reward,' trumps all other reasons. It would be irrational to *not* use a technique, however it works, that one knows would maximise reward.

There is evidence that agential success is affected by beliefs in self-efficacy and so the manipulation of self-efficacy beliefs involved in test-case willpower might be expected to help.[21] However I agree with Bratman that it is a strange view of agency that makes those arational manipulations the *primary* means of achieving diachronic stability.

Objection 4: Stabilising singular diachronic goals

This objection follows on from the main point made in Chapter 1 where I argued that many typical values in human lives take the form of singular diachronic goals. Humans must, therefore, have a way of diachronically stabilising such goals. Numerous SS rewards are mutually exclusive of these goals but so too are other singular diachronic goals. We cannot succeed if we try to become a concert pianist one day and then a professional footballer the next. Nor can we devote ourselves to one exclusive long-term intimate relationship one day and then another the next day.

Pre-commitments are helpful supplements for diachronically stabilising singular diachronic plans but they cannot stabilise the specific plans themselves. External pre-commitments tend to be non-specific in their effects, for example, not keeping wine in the house might help better piano practice but it also favours a range of other behaviours that don't fit well with wine drinking; it doesn't, therefore, commit one to trying to become a concert pianist. Probably the most flexible external pre-commitment involves telling other people your plans so that you are exposed to opprobrium if you fail to carry them out. This can sometimes be as specific as needed but it is limited by the fact that other people aren't always around to be told, don't always care enough to deliver sufficient opprobrium, or don't care if your resulting activity is not exactly what you planned. Internal pre-commitments are either only effective in the short-term or too general to ensure that a particular plan rather than something of similar emotional valance is carried out. I assume Ainslie would fairly readily concede these points given that he also sees pre-commitments as supplementary strategies.

[21] See 'Empirical evidence for human intentions' below and Bandura (1997). Although the self-efficacy literature indicates that self-efficacy beliefs don't just narrowly influence belief in the power of the present choice to set a precedent but they influence the range of options the agent believes to be accessible. This suggests that test-case willpower could be used to increase the size of the SS and LL rewards available to the agent (i.e. to make the agent more ambitious). That wouldn't sit easily with Ainslie's theory, however, since he takes expectable reward to be fixed prior to the influence of test-case willpower.

However, test-case willpower is also inadequate for diachronically stabilising pursuit of singular diachronic goals because it requires the agent to envisage a series of future choices of *a certain kind* where rewards clash. In the case of singular diachronic goals the LL rewards along the way are often unique and so the clashes between SS rewards and LL rewards have no recognisable kind; there is no cyclical pattern of rewards or rule that can stabilise the goal. It doesn't make sense to marry a person now, despite a fear of commitment, because that will set a precedent for future positive responses to marriage proposals. Neither does it make sense to commit to having a child now, despite the worry it will take up all one's time, because that will tend to encourage future choices to have more children. Choosing to marry or have a child lead to a marriage or a birth which are unique LL rewards. This then leads on to an evolving series of unique rewards (both SS and LL) that develop as a function of pursuing the marriage or raising the child, e.g. seeing the baby take its first steps, celebrating a 25 year wedding anniversary.[22] These rewards don't always come in predictable sizes, at predictable times, or clash with other rewards in consistent ways. Without a predictable pattern, the agent cannot see what the present choice really commits them to and so they cannot access the motivation of test-case willpower.

What the agent can do, however, is connect this specific choice within a more general set of choices. This is what Ainslie describes as 'throwing in collateral,' although he refers to the ploy as a way of strengthening willpower in general (Ainslie, 2011, p. 68). For example, an 18th Century gentleman might think that it would be fantastic to see the transit of Venus. However the chance to do so requires months of ocean travel and the gentleman is terrified of the sea. Once he's down at the docks ready to board, the SS reward of staying safe on land will outweigh the discounted reward of seeing the transit of Venus. But he will never have another chance to see the transit and so he cannot motivate himself by thinking, 'if I don't go now I'll never see the transit again.' What he *can* do is think, 'if I board this ship I will always follow through on brave choices but if I don't I will never make brave choices,' or, the ultimate generalisation, 'if I commit to this I will always see through my commitments and if I don't I will never be able to commit to anything again.' If he can convince himself of the connection between the singular diachronic goal and the wider pattern of choice, where he can clearly see the rewards and costs, then he should have the motivation to board the ship. So is test-case willpower suitable for stabilising singular diachronic plans after all?

[22] Some of the developing choice situations do recur and in those cases test-case willpower can work, e.g. balancing the SS rewards of sleeping versus the LL reward of taking your turn to get the baby back to sleep and maintaining a good relationship with your partner.

I think we can admit that test-case willpower can stabilise singular diachronic plans in this way but it remains inadequate because, rather than committing to the value of the singular diachronic plan itself, it bases the commitment on the value of something else. Say I want to commit to a marriage based on a rough estimate that it will provide significant LL rewards. Given hyperbolic discounting, I know I will struggle for diachronic stability in that commitment so I decide to stake some collateral on my commitment. I convince myself that if I cannot stick to this marriage I will never be able to make a relationship commitment again. But this makes my commitment to *this* marriage irrationally dependant on my ability to make relationship commitments in general. This becomes clear if the marriage severely deteriorates but I daren't leave because I believe my ability to commit to *any* relationship is on the line. This might explain some people's excessive commitments to unhealthy relationships but it shouldn't be the norm of being able to commit to a specific relationship. Perhaps the problem is that I invested the wrong collateral, I should have staked something less on the marriage. But whichever way we cut it up there will be a situation in which it is rational to abandon the marriage but irrational to believe I have lost the collateral which, after all, is only artificially connected to the marriage.

Agents appear to commit to singular diachronic goals with a strength of commitment that roughly matches the value of the goal and without putting less relevant collateral on the line. So test-case willpower is surely not the *only* way of committing to singular diachronic goals; I explain Holton's alternative below when I discuss intentions and muscle model willpower.

Objection 5: Inefficient diachronic coordination

There is reason to doubt that test-case willpower is sufficient to achieve typical levels of cognitive efficiency and diachronic coordination because Ainslie insists that choices aren't made in advance but always in the present. Ainslie claims that, "there is no physical connection between current and future choices. You literally make one choice at a time" (Monterosso & Ainslie, 2009, p. 122).[23] That is to say, when a prior time-slice makes a choice that they hope will be a precedent, that choice

[23] See also the following passage: "But how does a person arrange to choose a whole series of rewards at once? In fact, he does not have to commit himself physically. The values of the alternative series of rewards … depend on his expectation of getting them. Assuming his is familiar with the expectable physical outcomes of his possible choices, the main element of uncertainty will be what he himself will actually choose" (Ainslie, 1992, pp. 147, 150). Bratman agrees that, "Ainslie understands the choice of a series not as a present choice of future action or rewards but as a series of future choices. To 'arrange to choose a whole series of rewards at once' is to arrange that there be a series of future choices" .

is registered in memory and thus, although it influences future choices, it still leaves those future choices to future agents. Once that future time slice gets their turn to choose, the prior choice is available for consideration in the present choice but it doesn't count for anything more than that.

A prior time-slice cannot make a choice in advance for a future time-slice (although pre-commitment might preclude choice or heavily bias it in a particular direction) and so the present time-slice always has to make a decision. Making decisions is cognitively demanding (especially on Ainslie's view[24]): the present time-slice has to judge what kind of choice situation(s) they are in, consider their decision history in that kind of choice situation and judge whether the present case is an exception or not. If the case is not an exception then, depending on how such choices have gone in the past, they may need to generate or decrease self-confidence in their future selves by selective consideration of the evidence in order to choose the LL reward now.[25] Of course, Ainslie claims that test-case willpower is a largely implicit, sub-agential process; therefore, the agent needn't have the cognitive resources to *consciously* calculate expected reward bundles and balance test-case willpower beliefs. But implicit cognitive processing is cognitive processing nonetheless and so the brain still has to do the required work.

Contra Ainslie, there is good reason to believe that people do make decisions ahead of time and enjoy the increased cognitive efficiency and diachronic coordination of doing so. For example, an agent might decide on an emergency plan if the house catches fire. With a pre-existing plan the agent does not need to draw on higher cognitive processes that may be limited by panic; while those cognitive resources that would have been devoted to planning escape are left free to help implement the plan or improvise if required. Furthermore, by making decisions in advance, we achieve greater coordination in diachronic agency in ways that are independent of cognitive power. "…All sorts of other actions will be dependent upon what we decide to do; and we will need to perform some of these actions in the meanwhile" (Holton, 2009, p. 3). To borrow Holton's example, if I decide to paint my front door this weekend but the paint store is only open during the week then I need to decide on the colour now so I can buy the paint in anticipation. On Ainslie's

[24] Making decisions would be easier if we didn't have to calculate and compare bundles of expected rewards. In Chapter 4 I consider a more intuitive means of decision-making where the agent judges what action best fits with the self-concept they want to have. I don't pursue this aspect of cognitive efficiency here, however.

[25] It might seem that Ainslie can make cognitive savings through personal rules the way Heyman did with prudential rules. As we saw in the last chapter, you can make cognitive savings through rule-following if you apply the rule without a full recalculation of expected rewards. However personal rules are claimed to be just a way of describing the underpinning sub-agential bargaining between factions of reward expectations. If the rule is merely descriptive of the underpinning choice calculations then the agent is never actually following the rule and so reaps none of the cognitive savings.

view, agents cannot achieve this cognitive efficiency or diachronic coordination. Human time-slices seem prepared to be more diachronically cooperative than Ainslie estimates; we trust our prior selves to the extent that we will give up our present choice in order to act on their decisions and we trust our future selves to give up their choices in order to act on decisions we make now. As we will see, Holton makes advance decision-making (i.e. intentions) central in his account and so naturally accommodates the cognitive efficiency and diachronic coordination that typical agents exhibit.

Objection 6: Unresolved tension between bottom-up and top-down effects

My final objection in this section is that Ainslie's account involves a metaphysical tension that he doesn't try to resolve or clarify. He clearly commits himself to a bottom-up picture whereby all explanation of the agent and action can be reduced to the sub-agential competition between reward expectations. Yet he refers to many processes that seem to require an agent who is semi-independent of their reward expectations and who is attempting to control them. For example, sometimes personal rules are used to describe patterns that have emerged from sub-agential competition between reward expectations but at other times the agent is said to be following rules, making rationalisations, choosing rules with 'bright lines,' defining lapse districts, and engaging in self-deception to avoid noticing lapses. Sometimes test-case willpower is painted as dependent on an objective prediction of likely future behaviour based on decision history but at other times the agent needs to manipulate their beliefs based on selective attention to evidence.

Because Ainslie clearly commits to a reductive view I assume he would resolve this metaphysical tension by saying that, in the cases that seem to imply top-down causation, 'the agent' is just short-hand for the current, dominant faction of reward expectations. For example, dominant factions of reward-seeking processes, not agents (as the folk think of them), find rules with bright lines easier to follow and harder to rationalise exceptions to. Dominant factions of reward-seeking processes, not agents, manipulate their self-efficacy beliefs when trying to exercise test-case willpower. The dominant faction of rewards and potentially other competitive factions play the role of an agent but at a sub-agential, scientifically respectable level. This response, however, tends to undermine some of the intuitive appeal of his account. It redescribes what I took to be *my* rationalisations and *my* (begrudging) adoption of rules with brighter lines as outcomes fully determined by sub-agential processes. This doesn't just offend the sensibilities of anyone who hopes to secure a more active

role for the agent in agency but it results in some more serious explanatory costs which I treat in detail in the next chapter.

Recall that Ainslie adopted a reductive approach in order to avoid positing a homuncular agent controlling the body independent of the causal network. The problem with reference to a homuncular agent is that it doesn't seem to explain action at all; we then just want to know why the controlling homunculus acted as he did. As we will see in the next chapter, Bratman gives us a way of navigating between excessive reductionism on the one hand and supernatural homunculi on the other.

Summary

To use test-case willpower without excessive rigidity or abstinence violation effects the agent must carefully balance the associated instrumental beliefs at each decision of the relevant kind. The agent's variable skills and efforts of belief manipulation distinguish the effects of test-case willpower from the passive affective influence of the agent's decision history. Test-case willpower works through having an arational effect on the decisions of one's future selves; it does not provide them a reason to act in a certain way. This is not a concern for Ainslie, perhaps, but it doesn't sit well with thinkers of a more rationalist persuasion like Bratman. As we will see below, Holton's account can better appease this concern.

More problematically, test-case willpower and pre-commitments are inadequate for diachronically stabilising many singular diachronic goals that typical humans pursue. Pre-commitments are usually not specific enough and/or not strong enough to do the job on their own. Test-case willpower on the other hand requires the agent to put *something else* on the line to secure the commitment. Subsequently the commitment isn't really to the singular diachronic goal itself, and the size of the commitment is unlikely to be exactly proportional to the estimated value of the singular diachronic goal. Furthermore, not only is diachronic coordination more efficient if the agent can spread the cognitive load of decision making across time but some extremely common forms of diachronic organisation require it. Ainslie seems to rule out making decisions ahead of time except through pre-commitments which are heavy-handed in nature. Finally, Ainslie's reductive metaphysics completely effaces the agent replacing her with sub-agential reward

expectations. This is an unattractive theory of agency but it also has explanatory costs that I elaborate on in the next chapter.

Considering these points it appears that test-case willpower can provide some diachronic consistency but it is unlikely to be the flexible all-purpose control technique that Ainslie claims. Typical agents must be relying on something in addition to pre-commitments and test-case willpower to achieve typical diachronic coordination. Whatever that is, it might also provide a clue to what addicts are lacking. With this in mind I now turn to Holton's alternative take on diachronic consistency in action which is based on intentions and muscle model willpower. Intentions can stabilise singular diachronic goals and unlike test-case willpower intentions they have the benefit of being a rational form of control. They also accommodate decisions being made ahead of time and so can explain the cognitive efficiency and diachronic coordination humans display. Furthermore, Holton gives the agent a central role in his account, not just because it is metaphysically appealing but in response to empirical evidence.

Holton on diachronic agency

Holton's perspective on agency could scarcely be more different to that of Ainslie. Ainslie claims that agency develops from complex interactions between pre-existing desires while Holton claims that it develops from the efforts of the agent to control those desires. For Holton, therefore, the agent exists somewhat independently of her desires; she can make partially desire-independent *judgments*. Desires fluctuate as a function of extra-agential contingency[26] but judgments indicate what the agent takes to be good and so better represent who the agent really is.[27] The agent strives to bring her desires (and other contingent aspects of her environment) in line with her judgments. This entails the need to enforce diachronic stability in judgment; if her judgments also changed in contingent ways, then those judgments could no longer be distinguished from extra-agential forces.

In the next chapter I argue that we must treat the agent somewhat independently of her desires unless we want to give up many of our highly valued folk-practices. However, I set those issues

[26] Holton isn't committed to any particular explanation of how desires change. Perhaps it is a result of delay discounting; perhaps other factors are also involved.

[27] This picture is muddied by cases where inauthentic judgments clash with more authentic desires. I set these cases aside. Regarding addiction, no doubt some drug-users are coerced into judging that their drug-use is worse than it is and so struggle inauthentically against their drug-use. However, in the majority of addiction cases the judgment that the desire to use drugs has become detrimental is an authentic judgment.

aside for the moment to concentrate on Holton's alternative account of temptation and diachronic stability. I explain how Holton's theory improves on Ainslie's theory as I go.

We begin by looking at temptation. Holton distinguishes two different kinds of temptation, 'everyday' and 'addictive' temptation. Everyday temptation involves 'judgment shift' where the agent's judgment is corrupted by the temptation so that the agent comes to judge the temptation to be their best option. I argue that judgment shift has more explanatory power than hyperbolic discounting, primarily because it can account for the distorted values attributed to *future* rewards. The more extensive effects of judgment shift will severely undermine the usefulness of test-case willpower. Addictive temptation is more extreme; it bypasses judgment and so can motivate actions the agent knows are bad for them. The evidence for this form of temptation is also problematic for Ainslie who rules out action not aimed at maximising reward by definition.

In both cases of temptation the agent's judgment at the time of temptation will often be compromised. Holton's solution to this problem draws on intentions, i.e. settled decisions for future action, formed prior to the temptation ('everyday' or addictive) arising. Intentions represent an improvement over test-case willpower in that they can diachronically stabilise singular diachronic plans, improve diachronic coordination of agency, and more efficiently spread the cognitive load of decision-making. Holton also argues that we supplement intentions with muscle model willpower – a finite, executively deployed means of resisting the effects of temptation when intentions are at risk of being overwhelmed. Holton's claims for both intentions and muscle model willpower are supported by empirical evidence.

Everyday temptation

On Holton's theoretical framework, temptation is where the agent's strongest desire of the moment is in conflict with what they judge best.[28] Judgment is not an all-powerful controller of motivation and so the agent might cede to the greater motivating forces of temptation; this can happen in two ways. First, the agent might choose to act on a temptation despite simultaneously judging the act to be at odds with their best judgment; this is akrasia. Akrasia involves the unpleasant experience

[28] Here a temptation is any motive that goes against best judgment so it doesn't just involve the typical appetites for food, drink and sex. For example, fear might tempt one to avoid facing something that one knows it is best to face. A strong desire to act on a lesser ranked principle also counts as temptation, so one might be tempted to give money to charity despite your best judgment to save it for your family.

of cognitive dissonance and so we are motivated to avoid it. This leads Holton to hypothesise that it is actually relatively rare (2009, p. 103).[29] That's not to say we don't give in to temptations but that we succumb through another process where cognitive dissonance is postponed, perhaps indefinitely – judgment shift. In judgment shift, temptation corrupts judgment by a process where judgment changes or 'shifts' to corroborate the temptation.[30] The agent is then fooled, permanently or temporarily, into judging the temptation to be best after all and so feels no cognitive dissonance at the time of action. Of course if the shift is only temporary then they will suffer dissonance when judgment shifts back, but, by that stage it is too late to change the action. Succumbing to temptation through judgment shift is, therefore, distinguished from akrasia; this is what Holton refers to as, 'weakness of will.'[31] But what evidence is there for the judgment shift hypothesis?

There is anecdotal evidence where we realise in hindsight that a temporary judgment shift has occurred, e.g. 'I always justify that slice of cake at the time.' A more permanent judgment shift is more complicated to detect since we might not be able to distinguish it from a genuine change of mind. However they do occur, e.g. 'I always defended my purchase of that Ferrari but I see now that I was lured into it by the associated social status, something I don't really value.' In addition to this Holton presents empirical evidence that appears to catch judgment shift in the act. In a study by Karniol and Miller (1983), children are asked to value two possible rewards (marshmallows and chewing gum) and then are left in a room with the two options visible. They are told that if they wait until the researcher returns they can have their first preference but, at any time, they can ring a bell and receive their second preference immediately. When the researcher returns, ten minutes later, the children who have not ceded to temptation are asked to revaluate the two rewards. The children report a decreased value for their first preference (although it remains above the second preference). Control groups who did not have a bell to ring or who did not have the choices left in front of them did not change the values they attributed. Holton interprets these results as follows. The children who decrease the value of their first choice do so because they are exposed to the maximum temptation (options visible and an easy, authorised means of getting them). They

[29] Levy (2011b, p. 138) discusses the possibility that akrasia may not really occur at all. Of course, anecdotal evidence suggests that akrasia does occur –we say things like 'I know I shouldn't be doing this.' 'I'm going to regret this.' One's position here depends on how accurate you believe first-person reports to be; can the agent really be sure they had a simultaneous occurrent judgment against their action?

[30] Judgment shift is, therefore, distinguished from rationally changing one's mind which is not primarily temptation driven.

[31] Although this is a term of art and doesn't fit the folk-concept of 'weakness of will' which appears to involve a cluster of factors: breaking with prior commitments (intentions), acting against best judgment (akrasia), and acting against social norms (Holton & May, 2012).

anticipate that they are going to give in to the temptation and so, to avoid cognitive dissonance, they begin to implicitly or explicitly convince themselves that their original judgment was incorrect. When they are asked to revaluate their preferences we are seeing a snapshot of their judgment shifting. We can speculate that, if the process was allowed to run longer, they would succumb to the temptation at the point that their judgments of the first choice's value finally dipped below their judgment of the second choice's value (Holton, 2009, p. 100).

Does judgment shift differ from hyperbolic discounting?

At this point one might object that Holton has just redescribed preference reversal caused by hyperbolic discounting. Indeed Karniol and Miller interpret the children's changes in preference to be caused by changes in the estimate of the time until the reward becomes available. That is, the children aren't changing their preference for the reward itself; they are changing their preference for the reward plus its time until availability. As they begin to believe their first choice (the LL reward) is actually more distant than they thought, their second choice (the SS reward) begins to appear larger.

There are three main differences on Holton's view. First, in judgment shift, the reward itself is revalued rather than the reward's value remaining constant while the value change depends entirely on the estimate of time-till-availability. Holton's explanation is arguably the simpler one; after all, the participants have been asked for the value they place on the thing itself. Of course, participants may be implicitly changing that value as a function of the estimated time-till-availability.[32]

Second, on Holton's view, the agent necessarily plays a complicit role in adjusting the value of the options so it is not just a result of the extra-agential process of hyperbolic discounting. As a result, Holton's agent does not immediately regret giving in to temptation. He won't regret it until his

[32] Another doubt is connected to the dynamic of preference change. If hyperbolic discounting explains the preference change in Karniol and Miller's experiment, it would seem more likely that the time until being able to access their first preference was greatest at the very beginning of the experiment. If so, then, if they were ever going to choose the SS reward, they would have done so immediately. This is speculation of course, perhaps estimates of the time until a reward will become available increases the longer one waits for it (the children didn't know how long they would have to wait). To better distinguish the two theories on this count we would need to study temptation in a group who knew how long they would have to wait to access their first choice. Holton's model would predict that one subset of participants would immediately take their second choice and not regret it much (judging the wait too long) while another subset would exhibit a slow change in preference while under temptation which for some might lead to judgment shift before the time was up. Ainslie's model would predict one subset immediately taking their second choice and not regretting it, another subset immediately taking their second choice and regretting it, and a third subset waiting until their first choice was available without preference change over time.

judgment returns to normal and he reflects on the prior choice. In fact, he may never regret it if his judgment remains permanently shifted. Holton claims that empirical evidence supports him here; subjects that succumb to a temptation then go on to choose that option again even when they are no longer in tempting scenarios (Wang et al., 2010).[33] On Ainslie's theory acceding to temptation typically entails almost immediate regret because the agent will realise he has failed to maximise reward as soon as the SS reward is no longer immediately available. Ainslie can explain longer-term absences of regret with reference to the creation of lapse districts and self-deception, but these are meant to be exceptions rather than the usual case.[34]

Third, temptation appears to have wide ranging effects on judgment that judgement shift easily accommodates but hyperbolic discounting does not. An experiment by Wheeler at al. (2007), showed that subjects exposed to temptation were significantly more convinced by weak arguments for counter-attitudinal statements than non-depleted subjects. Both groups were equally convinced by strong arguments.[35] Judgment shift predicts that susceptibility to weak arguments would increase one's tendency to accede to a temptation. This is one reason why continued exposure to a temptation increases the tendency to give in. Such a susceptibility is not obviously relevant if hyperbolic discounting is all there is to the process. Furthermore, Wang et al. (2010) have shown that judgment shift not only affects the apparent value of temporally local options but also influences the value assigned to temporally distal options. In this study one group of subjects was exposed to temptation before choosing a movie to watch a few days later.[36] A control group was not exposed to temptation and made the same choice. The group exposed to temptation chose low-brow movies more frequently than the control group. This evidence of a temporally distal effect cannot be accommodated by hyperbolic discounting which predicts that future LL rewards continue to appear larger than *future* SS rewards (only more immediate SS rewards appear disproportionately large). Judgment shift in contrast can explain temporally local *and* temporally distal changes in preference.

[33] The cognitive dissonance program also found that subjects came to strongly prefer a thing they have chosen, even if at the time of the choice the preferences were very close (Brehm, 1956).
[34] Although I am suspicious that, in so doing, he helps himself to top-down mechanisms that do not fit with his reductionist framework.
[35] Furthermore, Burkley (2008) found that resisting temptations in general and resisting persuasion both resulted in people becoming more easily persuadable.
[36] The period of temptation involved different choices than the subsequent movie choice. This raises another issue of how temptation can have general effects. I address this below when considering muscle-model willpower.

The results of these experiments also raise further problems for test-case willpower. Test-case willpower relies on sending a bargaining signal to future time-slices. If we expect that our future time-slices will be convinced by weak argument then we cannot trust them to take our bargaining signal properly into account. Test-case willpower also depends on future rewards being relatively equally discounted. If temptation also distorts the value of *future* LL rewards to appear less than *future* SS rewards then the agent won't know which series of rewards is more valuable even when he bundles them together. As a result he may not realise that he needs to exercise control. Ainslie just assumes that LL rewards are stable but, if temptation works the way Holton claims, then the agent also needs a means of stabilising the value they attribute to long-term goals.

Addictive temptations

Holton suggests a way that addictive temptations might differ in *kind* from everyday temptations (Holton, 2009; Holton & Berridge, 2013). Holton suspects that addictive temptation might not involve judgment shift but bypass judgment altogether. His view is partially motivated by evidence that there is no tendency to judgment shift when there is a large disparity in value between the options (Karniol & Miller, 1983). If judgment was bypassed altogether then that would explain how people could chronically fail to act on their best judgments even when they valued the focus of their addiction far less than recovery. So even if akrasia is relatively rare in everyday temptation it may characterise addictive temptation.

Of course in many cases of addiction the addict doesn't have alternatives that are *clearly* better options than continued drug-use so, in those cases, judgment shift could well be sufficient to maintain addiction. In yet other cases the alternatives are so poor that the agent judges that, for the moment, continued drug-use is their best option.[37] So a special kind of addictive temptation may not be needed to explain all cases of addiction but it may be required to understand a subset of particularly strong addiction nonetheless.[38] If it is, then that is a problem for Ainslie because akrasia

[37] This group would include people who Kennett refers to as 'resigned addicts.' "They are not, like the unwilling addict, doing battle with their desires. But they do not, like the willing addict, endorse them or endow them with normative force either. They have given up. From their own point of view (as well as from a third person perspective) it is true to say that their lives are in an important sense out of their control" (2013b).

[38] One might claim that all addiction is driven by judgment shift, not akrasia, as Levy (2006, 2011a, 2014) seems to. Such positions face the issue of explaining addicted action in cases where values (drug use versus recovery) diverge too significantly for judgment shift to be expected to do the work. Of course it is difficult to objectively rank an agent's values so there may be no easy way to settle the debate.

is impossible on his view. For Ainslie, action is the pursuit of the greatest expected reward by definition; to pursue anything less is non-intentional behaviour. I criticise this aspect of Ainslie's view directly in Chapter 5 with reference to addicted action.

What evidence is there that judgment dissociation actually occurs? At the neurological level there is some suggestive evidence, albeit in rat models. Robinson and Berridge (1998) have shown that a rat's neurological system of motivation, or 'wanting', can be dissociated from its neurological system for experiencing pleasure, or 'liking.' One way to dissociate these systems is to expose the rat to amphetamines in conjunction with a cue (e.g. a noise) while making a sugar reward available. The amphetamine apparently sensitises the 'wanting system' because upon hearing the cue, and thus expecting the sugar reward, the rat exhibits a much stronger 'wanting' response (more lever presses) than controls who have not been exposed to amphetamine. But the same amphetamine dosed rats show no increased 'liking' response to the reward (measured by an innate facial response) and no increased 'wanting' response when the reward is available without the cue (Robinson & Berridge, 2003). Ten days after amphetamines had left the system the 'wanting' response was still double normal indicating that sensitisation involves a relatively permanent change in the brain.

Assuming the same thing happens in humans, when the addict encounters a cue connected with the circumstance of drug-use, they will feel a strong desire for rewards associated with that cue, typically the drug itself. Given the decoupling of 'wanting' and 'liking' this desire will be disproportionately large given the expected pleasure to be gained from the reward. However the situation is more complicated in humans because humans also have a system of judgment. Like expected pleasure, judgment usually moderates the 'wanting' system, for example, a desire to buy a clever garlic peeler is undermined by finding out it doesn't work. If a decoupling effect is to explain addiction in humans then it must also decouple *judgment* from wanting. If this further decoupling occurs then it makes addiction particularly worrying because it means that, not only will the agent be motivated to use drugs when it is not that pleasurable, but they will be motivated to use drugs when convinced that drug-use is not the best course of action.[39] If judgment cannot get sufficient grip on motivation then the agent may accede to addictive temptation despite their judgment and without judgment shift.

[39] This might then be considered a form of akrasia but unlike the typical conception of akrasia it is not associated with expectation of pleasure.

Indeed there is some first-person evidence of both kinds of decoupling. It is common to hear addicts say that they are well aware their continuing drug use is against their best judgment.

> "I desperately wanted to quit alcohol and drugs for many years. And I could continue to want to quit even as I was lifting the pipe to my mouth" (Sartwell, 2008a).

Addicts also sometimes claim that they want a drug even though they are not motivated by pleasure.

> "Yeah but now it's just… it's not even fun anymore really, it just sort of becomes a… I don't know, more or less like a chore I suppose but yeah I just… I want to get away from it" (R29, in J. Kennett et al., 2013, p. 8).

Of course there is always the pain of withdrawal to be avoided but if that was a strong motivator it is strange that so many addicts get through the withdrawal and then begin using again soon after.

If judgment and desire can split in this way, it might seem that the prospects of achieving diachronic stability when addicted are quite grim. Indeed, if addictive desires were so strong that they constantly and completely bypassed judgment then the agent would have no chance. Fortunately, however, addictive desires are not always the strongest desires of the moment even for chronic addicts. Holton argues that agents can resist both addictive and 'everyday' temptation through use of intentions and muscle model willpower.[40]

Intentions

Intentions are formed whenever the agent decides in advance on a future action.[41] Holton provides the following example of forming an intention:

> "Suppose that you have been wondering what colour to paint your front door. You have narrowed the options to two: dark red, or dark blue. Both would be nice; both are available. But time is pressing, and you need to decide. So you make your choice. Blue it is. As a result of your choice you have acquired a new mental state. You still think that both colours would be nice; you still think that both are available. In addition, though, you are now in a

[40] Holton acknowledges that pre-commitments can be useful but we don't always need to resort to such extreme measures. "We should not let our interest in these exotic methods blind us to the fact that very often intention *is* enough" (2009, p. 10).

[41] Holton argues that intentions are also formed when we decide on a present action (2009, pp. 12-14). This may be true but, as this is peripheral to my current focus, I set the issue aside.

> state that does not look like either a belief or a desire. You have an *intention* to paint the door blue" (Holton, 2009, p. 1).

There are several advantages to the agent in forming intentions, including improving coordination with oneself and others over time, preserving the cognitive effort of having made a decision, and combating temptation.[42] All three advantages are improvements over test-case willpower.

Making settled decisions in advance can improve diachronic coordination with oneself and others. As mentioned above when discussing the limitations of test-case willpower, I will improve my diachronic coordination with myself if I can, on Friday, decide what colour to paint the door over the weekend, because I need to know which colour paint to buy on Friday. I can also improve my diachronic coordination with others; if I have decided on the door colour on Friday, then I can already invite my Sunday dinner guests by directing them to my house by referring to the colour of the door. Typical human lives are filled with diachronic coordination that depends on settled choices. I decide what to cook during the week so I can get all the groceries in one trip to the supermarket. I schedule an appointment with a doctor when I have decided I will already be in the neighbourhood for a prior commitment. I agree my friend can stay at my place and so turn down invitations to events out of town over those days. I tell my friend at the market to meet me back at a certain place and time in case my phone battery dies.

Intentions also help to distribute cognitive effort more efficiently. I might know that I will get flustered at the paint shop if I leave the decision until then. If I can consider my choice in a quiet moment, perhaps when I can look at the front of my house and discuss options with my wife, then I know I am likely to make a better decision. When I get to the paint shop I just have to act on the prior decision, saving the cognitive resources at the time for other things. To get this benefit of intentions I need to know when I am in a good position to make a decision and when I will not be. Again, people frequently make decisions on this basis. Plans are decided upon in the case of fires, floods, earthquakes and tsunamis. People research what they want to buy before going to purchase to avoid being dazzled by excess options and details.

Finally, intentions can protect us from both 'everyday' and addictive temptations by setting our decisions prior to the effects of judgment shift or akrasia. Perhaps when I get to the paint shop there

[42] Another benefit that becomes more relevant in the next chapter is the capacity to break a deadlock between actions that are equally desired (Buridan's Ass cases) or of high but incommensurable desire. Intentions are therefore sometimes the result of executive choices that are underdetermined by beliefs and desires (a possibility that Ainslie doesn't allow since he assumes all rewards are commensurable).

is a great deal on a different colour (that I would regret buying). Usually I am extremely tempted to save money but if I act on my intention, and so don't reconsider the decision, then I won't be prone to judgment shift. Intentions may, therefore, be formed specifically for the purpose of combatting foreseen temptation, e.g. 'I won't have a drink tonight,' 'I'll never drink again,' or, 'I will do a bungy jump.' When they are, Holton suggests we call them 'resolutions.' As when improving cognitive efficiency, I need to form intentions when I am free from judgment-impairing temptation. I also need to anticipate tempting situations so that I can have the intention formed ahead of time.

Theoretical simplicity

For the theorist of diachronic action, intentions have the additional advantage of providing a single means of explaining the diachronic stability of any kind of decision, singular diachronic plan, prudential rule, or global choice pattern. The only limits on the agent forming an intention are the limits they face in making a settled decision. In the next two chapters I look in more detail at how agents go about creating and deciding on singular diachronic plans; however, if we grant that agents can decide to pursue singular diachronic plans then there is nothing to stop them stabilising those plans with intentions.

What Heyman refers to as 'prudential rules' can be stabilised by making them standing intentions, or policies, where the agent decides in advance to always act in a certain way when a certain kind of decision situation arises. Rules are only prone to rationalisation if the agent reconsiders their commitment to them; intentions diachronically stabilise the rule against reconsideration. The agent might develop a policy himself, particularly if he finds himself often forming similar singular diachronic goals. For example, if the agent often formed the intention, 'don't get a dessert after dinner tonight,' he might realise that he would benefit from the standing intention not to eat cake This is cognitively more efficient because it saves on the resources required to form a series of individual intentions. Another way to develop a policy, I suggest, is through the thinking involved in test-case willpower. One recognises the difference between a series of SS rewards and a series of LL rewards and decides on a rule to choose the LL rewards each time. However, the stability of these rules isn't provided by test-case willpower but by refusing to reconsider the rule. Alternately

the situation might be as Heyman envisages; the agent hears of a prudential rule and tries it out. Prudential rules are essentially socially popular standing intentions.[43]

Finally, Heyman's patterns of global choice would be prone to judgment shift but they too can be stabilised as intentions if I refuse to reconsider the pattern I have chosen. Say I have developed a pattern of global choice to maximise reward from eating at Chinese and Italian restaurants but without forming an intention. Under judgment shift, I then think, although my choice pattern had recommended I eat Italian tonight, the Chinese looks too good to resist; I'll eat more Italian later in the week. I could have avoided this inconsistency if I had made my specific pattern of choices an intention and so refused to reconsider it. This pattern could be a singular diachronic plan or, if required, I could make it a policy, e.g. 'if I have eaten Chinese the last two nights I must choose Italian tonight.'

Rationality

One issue with test-case willpower was that it is an arational means of providing diachronic stability. Intentions have the potential to provide rational diachronic stability. Here I briefly discuss the conditions intentions have to meet to be rational.

First we should note that forming an intention does not give the agent an additional reason to *act* on their decision. If it did this would have irrational results; I could give myself a reason to do anything at all simply by forming the intention to do it. This is called the bootstrapping problem (Bratman, 1987, p. 24ff). We can avoid the bootstrapping problem by linking the rationality of an intention to the decision where it was formed. Intentions produced by irrational judgment are irrational so anything that undermines the rationality of a decision (e.g. judgment shift) cannot be compensated for by forming an intention. The intention doesn't adjust the rationality of the decision. What the intention does do is provide a reason not to reconsider the decision (Holton, 2004, pp. 514-515). Similarly Hinchman argues that an intention is not advisorial but executive. Advice is intended as an additional factor to be weighed in future decision-making but forming an intention requires the future agent to see the decision as already settled (2003, p. 34).

[43] We should also note that even individuals' simple intentions (i.e. singular diachronic plans) are visible in the discursive realm and although another agent cannot use that exact intention himself he might use it for inspiration in the creation of his own intentions.

If acting on intentions is to be rational then treating a decision as settled appears to lead to another problem, 'the problem of akratic resolution.' For example, I have an intention (or resolution) to go skiing on Saturday morning and so to go to bed early on Friday night. However, on Friday night my best judgment recommends going out drinking. So on Friday night it seems that my resolution is akratic; it requires I act against my best judgment and, even if that will be good for me, to do so is irrational. It appears that intentions are inherently irrational after all. This problem can also be sidestepped if we take intentions to give us a reason not to reconsider. I would act akratically *if* I reconsidered my decision, made a new judgment that drinking was best yet still acted on my intention. However, if I do not reconsider then I don't make that new judgment and so I don't act against it when I follow the intention.

Now, even if acting on intentions is not irrationally rigid per se, it clearly would be irrational in certain circumstances. For example, if I find out on Friday afternoon that the ski field is closed for the season then that is the kind of radical change in circumstance that should trigger a reconsideration of my intention to go skiing. Clearly, the agent needs to strike a balance between reconsideration and non-reconsideration because sometimes circumstances change to make non-reconsideration irrational. As a rough guide, it is reasonable to have tendencies not to reconsider when the intention was made in circumstances that allowed clear thought and now circumstances impede clear thought, or where the intention was made expressly to counter temptations now being felt. It is reasonable to have tendencies to reconsider when the circumstances are so different from those expected that they defeat the purpose of having an intention, the intention can no longer be carried out, or one now believes that the intention will lead to great suffering (Holton, 2009, p. 160). The judgment of whether to reconsider an intention becomes more difficult when one aspect of a situation favours reconsideration while another favours non-reconsideration e.g. where a resolution to not to overeat is at odds with showing appreciation for the in-laws' cooking and politeness in the face of their insistence on second helpings. Sometimes there is no (epistemically accessible) fact of the matter as to whether reconsideration or non-reconsideration is best. These kind of hard judgments come with the territory for epistemically limited creatures like ourselves.[44]

[44] Cohen and Handfield (2010) endorse a view very similar to Holton's. They argue further that the ability to balance reconsideration and non-reconsideration of intentions is not just useful for maximising beneficial consequences but it is necessary for human practical agency. Therefore we should resist reconsideration of some intentions even in a world where ready reconsideration regularly leads to greater rewards. Likewise we should reconsider intentions that are clearly no longer rational even in a world where excessive stubbornness leads to greater rewards.

Hinchman (2003) describes the correct balance in terms of self-trust among agent time-slices. "…The earlier self must believe that the later self will follow through on the intention through an exercise of reasonable trust. And … the earlier self must believe that it will appear trustworthy on the matter at hand" (Hinchman, 2003, p. 35). When circumstances have changed unpredictably or new information has become available, the earlier agent may no longer be a trustworthy judge of the best action. On the other hand, if you judge that your earlier self did know enough about the unfolding circumstance then you should continue to trust their judgment and so act on the intention.

So, as with test-case willpower, there is a balancing act involved in setting the strength of one's intentions. However the balancing act in forming intentions is less precarious than that involved in test-case willpower. When one has an intention, say, to not smoke, the strength of the intention isn't necessarily disturbed by failing to reconsider when you should have or reconsidering when you shouldn't have. The agent can take that failure into account and choose to adjust the parameters governing when to reconsider if they think it will be worthwhile. If Hinchman is right this involves adjusting the parameters around when self-trust is or is not reasonable. This allows the agent to achieve a stability in their stance because the strength of their intention is partially isolated from their decision history; the odd failure doesn't risk abstinence violation effects and consistent success does not breed overconfidence. When using test-case willpower, in contrast, one cannot act *without* changing the strength of willpower because the choice of an SS or LL reward has an inevitable recursive effect. To help maintain a stable stance, the agent will sometimes need to rebalance their test-case willpower through making choices she would not otherwise approve of, e.g. to prevent overconfidence.

In a further complication, Holton rightly recognises that even when we act on intentions the non-reconsideration of the decision cannot entail a complete lack of cognitive awareness in regards to the intention. To not consider the intention at all when enacting it would reduce the action to arational, automatic behaviour. Intentions, particularly resolutions, tend to be used in cases where one wants to *revise* automatic behaviour. To stop smoking one has to monitor behaviour to realise when one has a cigarette in one's hand and then one has to remember and implement the resolution. Implementing an intention requires awareness of that intention (including perhaps the reasons for it) and the conditions in which it is relevant. It seems reasonable, I think, that this level of reflective awareness would not entail a full reconsideration of the decision. Holton refers to this consideration without *re*consideration as '*rehearsal*.' The distinction between reconsideration and rehearsal is not sharp. When rehearsing you go some way into the process of deliberation but stop before

moving all the way to making a new decision. At some stage you realise that you have already formed a relevant intention, you note that conditions are such that reconsideration is unreasonable, and so you cut off deliberation at that point, and enact the intention. Because the process has several stages you can catch yourself moving past rehearsal towards full deliberation and stop that slide (Holton, 2009, pp. 123-124).

Muscle model willpower

Holton argues that action is not determined just by the agent's beliefs, desires and intentions but also by their muscle model willpower. Muscle model willpower, he suggests, is a means of preventing reconsideration of one's intentions even when seriously threatened by temptation.

> "Agents whose willpower is strong can stick by their resolutions even in the face of strong contrary desires; agents whose willpower is weak readily abandon their resolutions even when the contrary desires are relatively weak" (Holton, 2009, p. 113).

As it happens there has been extensive empirical research which postulates a form of willpower that does just this (perhaps amongst other things). This research indicates that various forms of self-regulation depend on a common limited resource that is depleted through various acts of executive control (Baumeister et al., 1994). It is called the 'muscle model' of willpower because, like a muscle, it can only resist a finite load for a finite time, i.e. strength and duration of temptation, and it can be strengthened by use.

There is significant empirical evidence for muscle model willpower. A typical experiment to show muscle model willpower involves a test group engaging in a self-control task for a set period of time, e.g. not smiling during a funny movie. That is, they are exposed to temptation and try to resist it. They are subsequently said to be 'ego-depleted'. Meanwhile the control group has to engage in a task that is matched in terms of cognitive demand but that does not require self-control in the face of temptation, e.g. ranking options according to desirability. Then both groups try to persist at another task requiring self-control in the face of temptation, such as using a handgrip exerciser for as long as possible or persisting at an unsolvable puzzle. The ego-depleted group persists for a shorter time at the subsequent self-control task than the control group (Baumeister et al., 1998; Muraven et al., 1998). Presumably the agents in these experiments have formed the intention to use the handgrip exerciser for as long as possible or to solve the puzzle. They are tempted to give

up but resist by using their remaining muscle model willpower which either then runs out or they refuse to deplete it any further, and they give up. In other studies the subjects more clearly have an intention that they care about being threatened by temptation. For example, in studies of dieters, the ego-depleted group, when subsequently tempted with food, break their resolutions and eat more than non-depleted dieters (Hagger et al., 2013; Kahan et al., 2003; Vohs & Heatherton, 2000).[45] The muscle metaphor isn't just appropriate because the agent has a finite strength that runs out but because they can build it up through exercise. Strengthening willpower through practice was shown in a study where the group who practiced self-regulatory exercises, such as developing a better posture, showed less tendency to suffer ego depletion (Muraven et al., 1999).

Muscle model willpower is essentially what Ainslie refers to as the synchronic internal pre-commitments of attention and emotion control. Ainslie dismisses the importance of this capacity given that it cannot provide stability in the face of temporally extended temptation. This isn't a serious problem for Holton's view, however, because he considers intentions to do most of the work in diachronically stabilising action while muscle model willpower is an emergency back-up. The reason typical agents do not frequently run out of muscle model willpower is that they are protected by their sets of intentions. If I have a penchant for unhealthy baked goods my intentions can prevent me from being exposed to that temptation. I may have a standing intention to not go into bakeries but, more importantly, I form a variety of other intentions as I go about my life that just don't include going into bakeries. I will only end up in front of the bakery having to use my willpower if I reconsider those intentions. In cases where agents cannot avoid persistent temptations they do indeed tend to succumb and that is consistent with them relying on finite muscle model willpower. So, although Ainslie is right that a skill like muscle model willpower alone is insufficient for diachronic agency, when it is combined with intentions it is a useful emergency back-up.

Holton's overall picture of diachronic stability then is as follows: The agent forms intentions with supportive means (implementation intentions) to resist temptations but also to coordinate their activity over time. They are constantly developing their sense of when it is right to reconsider an intention and when it is not. This calibrates the strength of their intentions and can be varied according to domain. Occasionally intentions come under strong threat from temptation and then

[45] As well as these experiments we know that depression, anxiety and tiredness all reduce one's ability to be resolute. In these states reformed alcoholics, dieters and people trying to give up smoking are more likely to relapse (Baumeister et al., 1994).

the agent has recourse to a finite store of muscle model willpower to resist reconsideration. As that finite store runs low they begin to accept weak arguments for taking the temptation, their judgment shifts and they succumb; or, if they suffer from addictive desires, they give up on their intention to avoid drug-use and begin to form new drug-using intentions that go against their best judgment.

A loss of explanatory power?

I conclude the chapter by reconsidering the explanatory successes Ainslie claims for his theory. Recall that Ainslie has a means of explaining the domain-specific levels of control an agent exhibits and why abstinence violation effects (AVEs) and excessive rigidity in rule following occur. Do we lose explanatory power by adopting Holton's position? I don't think so.

Ainslie correctly claims that muscle model willpower is a general power that cannot explain domain-specific levels of diachronic control (Ainslie, 2005, p. 643). Holton can agree on that count but still explain domain-specific variation with reference to the domain-specific intentions formed (or not formed) and the domain-specific strengths of tempting desire. Agents will exhibit more control in domains where tempting desires happen to be weaker and where they have carefully developed protective intentions. The agent will exhibit less control in domains where they experience strong tempting desires and they have failed to develop protective intentions.

What about AVEs? Holton's agent might lose confidence in their ability to control their action in a domain through a series of failures to diachronically stabilise their action. A build-up of failure in a domain will eventually cross a threshold where the agent loses confidence in their abilities; they give up and an AVE eventuates. However we can just as easily understand this as a loss of confidence in forming and committing to an intention as a loss of confidence in believing this choice will be a precedent. Overconfidence bred by success might create a false belief that future selves will choose the LL reward despite choosing the SS reward now, but it can equally be explained as a laziness whereby means-ends planning becomes dangerously incoherent. Finally, excessive rigidity might develop through concern about the present choice being a precedent. However it might also indicate a limited ability to know when it is best to reconsider intentions or a lack of imagination and social scaffolding required to develop new ways of responding to the environment.

Conclusion

In this chapter I have drawn on Ainslie and Holton's theories of how agents maintain diachronically consistent action in the face of temptation. Ainslie claims that diachronic inconsistency is caused by hyperbolic discounting and that agents use pre-commitments and test-case willpower to counter it. Pre-commitments, however, are too inflexible, too weak, or require too much planning to do the job alone, so most of the work falls to the more flexible test-case willpower. I argue that, even if Ainslie correctly characterises temptation, test-case willpower is an arational means of achieving stability, it cannot adequately stabilise singular diachronic goals, nor can it explain the diachronic coordination we see in typical human action.

More problems for Ainslie's account become apparent when we consider Holton's alternate description of 'everyday' temptation. Holton's view of temptation involves judgment shift and better fits the empirical data than hyperbolic discounting alone. If everyday temptation does involve judgment shift then this seriously undermines test-case willpower because it affects the agent's ability to compare the value of two *future* rewards; subsequently the tempted agent doesn't really know what the LL reward is. Furthermore, judgment shift involves an increased susceptibility to weak arguments, so when a tempted agent considers a bargaining signal sent from his earlier self he may not weigh that signal as strongly as he should.

Holton also describes a different kind of temptation, 'addictive temptation,' that can directly bypass the agent's judgment rather than shift it. The result is akratic action that goes against the agent's best judgment. This form of temptation can explain some of the first-person reports addicts make and some of their otherwise mysterious action. Ainslie's reward maximisation account struggles to explain these phenomena. I pursue this argument in Chapter 5.

Whatever the kind of temptation faced, Holton claims our primary tools for diachronically stabilising action are intentions. Intentions provide stability by settling the decision before temptation corrupts or bypasses judgment. I have argued that intentions can also provide better diachronic coordination and cognitive efficiency by spreading their decision-making over time. Intentions are particularly appealing from the stand-point of developing a theory of agency because they provide a single means of explaining the diachronic stability of singular diachronic goals, policies such as prudential rules, and global choice patterns. Furthermore, intentions have the

potential to provide *rational* diachronic stability neither overruling best judgments nor preventing judgment from occurring.

When temptations are particularly strong or unexpected, muscle model willpower can provide additional support. Although Holton doesn't focus on it, it also makes sense to further support your intentions with a range of pre-commitments (balancing their strength against their inflexibility) because by avoiding temptation altogether one protects one's limited supply of muscle model willpower. Pre-commitments, even quite inflexible ones, are likely to be particularly useful in the early stages of recovery from addiction when diachronic inconsistency is most prevalent.

We're now ready to move on to Chapter 3 where I continue to argue against Ainslie but with an emphasis on the role of the agent. The problem with Ainslie's reductive view is that it effaces the agent, replacing him with sub-agential forces. I outline a range of explanatory problems this generates for Ainslie: He cannot account for the typical phenomenology of temptation, degrees of control over action, or distinguish long-term changes in desire that are attributable to the agent rather than extra-agential forces. I argue that this creates a number of explanatory. He finds himself in this situation because he has dismissed any views that see the agent as a crucial player in agency as positing a homunculus to explain action rather than looking to scientifically respectable entities. The choice between his reductive account and an account with a homunculus is a false dichotomy. I draw in Michael Bratman's work to argue that we can posit a semi-independent agent without committing to any supernatural entities. The benefit of doing so, in addition to recovering a more palatable view of agency, is that we can then explain temptation phenomenology, degrees of control, and different sources of long-term change in desires.

Chapter 3: Normative planning agency and self-governance

Introduction

In this chapter I argue that we need a normative account of agency if we are to accommodate an active role of the agent, the related phenomenology of effort, degrees of control over action, and distinguish agential from extra-agential developments in desire. I develop my arguments in favour of normative theories drawing on Michael Bratman's account of planning agency. On Bratman's account, the individual constitutes herself as a self-governing[1] agent by following three norms of practical reason. The first of those norms – the norm of diachronic stability – we saw in Holton's account; intentions must be sufficiently diachronically stable to allow present decisions to organise future action but not so stable that they cause irrational inflexibility (Holton, 2009, p. 89). I discussed the other two norms briefly in Chapter 1 when considering the requirements for forming singular diachronic plans. These are the norms of means-ends coherence in planning and ends-ends consistency among plans and policies. On a normative account of agency, the individual is responsible for responding to the claims norms make on her and she, therefore, plays an active role in agency. Bratman's account describes how the human agent progressively creates a network of plans and policies by following these norms. The better she follows the norms of practical reason when developing that network, the better her plans and policies interlock and the more cross-temporal connections of personal identity this network ensures. Therefore, normatively endorsed plans and polices have agential authority and the more closely an agent follows those plans and policies the greater self-governance she achieves.

Ainslie defends a non-normative account of agency so I continue to use his position as a foil. Ainslie's account describes action and all changes in desire as the results of objective, causal processes of reward expectation. This leaves no room for self-governance, or degrees of self-governance, which is the root of the explanatory limitations I describe here. This chapter, therefore, extends the critique of Ainslie's position from the last chapter where we saw that test-case willpower could not adequately stabilise singular diachronic plans, required the constant rebalancing of arational sources of motivation, and was vulnerable to judgment shift. Holton's account of the techniques humans use to overcome diachronic instability, intentions aided by muscle model willpower, was shown to be superior and also happens to be compatible with the

[1] Bratman's view of self-governance will become clearer as we go but as an initial description it involves the agent directing and governing her practical thought and action. The agent directs action when she thinks and acts using plans and policies that have agential authority. Plans and policies have agential authority when they create cross-temporal connections and thus constitute a diachronic identity.

normative account developed here. In Chapter 5 I show how the points developed in Chapters 1 through 3 are relevant to addiction; I argue that a normative planning account of agency informed by Bratman and Holton can help us better understand addiction than the non-normative ('choice') account favoured by Heyman and Ainslie.

I begin this chapter by reiterating Ainslie's position. I then set out Bratman's view of planning agency before arguing for its explanatory benefits in turn. First, Bratman can accommodate the phenomenology of temptation – an effortful struggle to act on one's normatively endorsed plans or policies or to cede to the current strongest contrary desire. This phenomenology is a mystery on Ainslie's account where the agent has no rational basis on which to oppose their strongest desire. Second, the planning theorist can explain degrees of self-governance with reference to the amount of effort required to act in accordance with normatively endorsed plans and policies in the face of contrary desires of varying intensities. For Ainslie, action is either under perfect control or it is an unintentional behaviour that does not count as action. This clashes with our folk-understanding of self-governance. Finally, Bratman can attribute long-term changes in desire to the agent's efforts of self-governance rather than extra-agential forces when they stem from plans and policies designed to adjust those desires. On the non-normative account all changes in desire are extra-agential; again, this clashes with out folk understanding.

Ainslie and normativity

Ainslie takes himself to be developing a systematic causal hypothesis for action.[2] He claims that an action is only carried out if that action promises the greatest expected reward out of all possible actions open to that agent. Expected reward in this sense is causally determining, not normative, so that "…the individual is constrained to choose the option with the greatest expected reward of

[2] At least I assume he does since he criticises his opponents, supporters of muscle model willpower, as failing to provide systematic causal hypotheses (Ainslie, 2005, p. 636).

all those she considers" (Monterosso & Ainslie, 2009, p. 116).[3] The reward expectations that ultimately control behaviour have out-competed all others in a sub-agential 'marketplace.'

> "…In order to prevail an option has not only to promise more than its competitors, but to act strategically to keep the competitors from later undermining it. The behaviours that are shaped by the competing rewards must deal not only with obstacles to getting their reward if chosen, but with the danger of being un-chosen in favour of imminent alternatives (Ainslie, 2005, p. 637).

Ainslie assumes that agents and the strategies that provide diachronic stability in action such as precommitments and test-case willpower can be reductively explained in terms of competition between expected rewards. The agent is not in any way separable from this competition and so has no independent influence over the process and subsequent action.

> "…If choice is determined in a marketplace of competing interests, 'she' [the agent] is just the resultant of their activities, and stable choice has to be achieved as it is in the kind of markets that don't have governors" (Ainslie, 2005, p. 642).

The agent, qua dominant set of reward expectations, can predict the way they will perceive future rewards and take action to manipulate those perceptions using precommitments and test case willpower. When these strategies successfully result in relatively stable sets of expected rewards then something sufficiently unified to be called an agent has emerged but that is not an achievement of the agent. The agent themselves doesn't add any more stability to the process which is driven only by bottom-up forces; the diachronic stability only depends on the dominant reward expectations being dominant enough to consistently outcompete the others. Precommitments and test-case willpower are, therefore, just ways of describing the details of the underlying market forces at play; they are not executively wielded techniques. The agent cannot change the value of rewards as a governor might manipulate a market since the jostling of expected rewards *is* the governor and so she cannot but endorse what the market endorses. The rewards the agent pursues typically change over time but that change can only be a function of interactions between desires and extra-agential forces like biological maturation and external contingency. Expected rewards

[3] It's worth noting that Ainslie doesn't actually treat this as a hypothesis (i.e. as defeasible). He considers it an incontrovertible principle of all action. This principle becomes suspicious when agents pursue extremely poor rewards, as addicts quite often do; I make more of this in Chapter 5

and their dynamics are, therefore, assumed to have an objective existence; they develop, disappear, and interact upstream of the agent's awareness and judgment.

Part of what motivates Ainslie's position is that he thinks that normative accounts of action are committed to positing an agent that somehow manipulates action from outside the causal system of desires like a 'homunculus.'[4] Explanations of agency that appeal to a homunculus fail to really explain because we then want to know what are the homunculus's desires, norms, actions, et cetera. Ainslie, however, does not just remove the homunculus from the picture but any role for the agent at all. I argue that, as a result he loses more explanatory power than he gains. As we will see Bratman claims to provide a way of distinguishing the agent without removing them from the causal order. Whether he really succeeds in this is debatable but, due to space constraints, I will not try to settle that debate. Rather, I aim to set out the considerable explanatory power that comes from distinguishing the normatively responsive agent from their desires. A metaphysical concern remains as to how an agent semi-independent from their desires can be reconciled with a causally closed universe. I do not address this problem but what I say should at least strongly motivate the development of a metaphysics that makes room for normativity.

Bratman and normative agency

Michael Bratman develops an account of self-constituting and self-governing agency where the agent remains fully embedded in the causal order. Bratman's project is to steer a course between accounts such as Watson's which, in his opinion, link agency too closely to intersubjective standards of the good,[5] and naturalistic accounts, such as Frankfurtian hierarchies of desire (and Ainslie's account), that tend to efface the agent and his values altogether.

According to Bratman, the agent constitutes himself as a self-governing agent by developing an interlocking network of plans and policies that ensure inter-temporal (Lockean) connections that define his diachronic identity and guide deliberation. Plans and policies will only interlock and create stable inter-temporal connections if the agent follows norms of practical reason. Those

[4] He also avoids having to explain how norms, values and reasons can have an existence independent of more scientifically respectable forms of motivation such as biologically-driven desires.

[5] Bratman is at least partially motivated by a commitment to pluralism. Agents often make commitments that they do not assume would meet intersubjective standards of the good. For example, an agent who decides to eschew sexuality as part of a religious ideal can be said to govern his life accordingly without believing that such a life would be best for everybody or even everybody in his position (Bratman, 2007, pp. 233-234).

norms demand plans and policies to have diachronic stability, means-ends coherence, and ends-ends consistency. Certain policies are designed specifically to improve inter-temporal connections – self-governing policies. The more closely the agent follows their self-governing policies the more their action is self-governed.[6]

We can begin to understand Bratman's account as a response to an influential debate between Frankfurt and Watson. Central to that debate are the distinctions between judging an end to be valuable (or good), actually valuing that end, and merely desiring that end.

Judging valuable, valuing, desiring and agential authority

Frankfurt (1971) famously developed a view of agential authority based on a hierarchy of desires. Roughly, an agent acts in accordance with her will when she acts on a first-order desire while having a second-order desire that that first-order desire move her to so act (this is what Frankfurt calls having a 'second-order volition'). The higher-order volition must be uncontested or, if contested, supported by the highest-order desire in the hierarchy. Watson (1975) responded with several criticisms but the most relevant here is that it is not clear why higher-order desires should represent the agent any more than first-order desires. Higher-order desires are results of what Bratman calls, 'weak reflectiveness.' Such desires are just further desires in the 'psychic stew' and don't, therefore, have any more agential authority than first-order desires. Strong reflection, in contrast, would represent where the agent herself stands rather than just another pro or con attitude (2007, pp. 23-24).

Watson's attempted solution appealed to what the agent judges to be valuable (or good). He claimed that the agent's standpoint is constituted by their evaluational system (1975, p. 216) and not by the fluctuation of their desires (even those of higher order). However, Watson realised that the early formulation of his view faces a problem because, "judging good has no invariable connection with motivation. … One can in an important sense fail to value what one judges valuable" (Watson, 1987, p. 150).

[6] By dissociating self-governance from intersubjective standards of the good Bratman makes room for 'evil' people to be self-governing. He leaves it open as to whether a complete account of autonomy would require the capacity to judge correctly about the good (2007, p. 9).

As a result, the agent's evaluations can be inconsistent with his actions and we then need a way of saying why the evaluation rather than the action represents the agent. Similarly, evaluations can be inconsistent, as we saw in cases of judgment shift. A further issue highlighted in the work of Holton and Bratman is that judgment often cannot distinguish the value of two or more valuable actions. In such cases the agent must form an intention to pursue one or the other to break the deadlock (Bratman, 2007, p. 262). It is then intuitively correct to say that the agent values the action chosen but we cannot say that he chose that way because he judged it best. In all these cases we have a problem defining where the agent stands. What the agent evaluates as best can come apart from what they actually value through their actions.

Bratman's view of self-governance is presented as a solution to this. He claims that where the agent really stands is defined by what they actively value and not by their evaluative judgment. "The intention-based theory sees the agential authority of volitionally infused valuings as a matter of their organising roles in our temporally extended agency, and not primarily as a matter of a good-tracking role" (Bratman, 2007, p. 248). What the agent values is defined by what they commit to with plans and policies.[7] But not just any plans and policies will do; the more the agent's plans and policies are developed in accordance with norms of practical reason the better they reinforce a temporally extended identity and therefore the more agential authority they have.

Norms of practical reason

There are three norms of practical reason. The first norm requires rational diachronic stability of intentions. As we saw in the last chapter, intentions qua intentions require a certain diachronic stability without being overly inflexible. If you form an intention to avoid a second glass of wine at dinner tonight or a policy to avoid second glasses of wine in general, that intention should guide the refusal of a second glass. If it never has an effect on whether you take that second glass (when circumstances turn out as expected when the intention was formed) then you don't really have that

[7] Bratman's claim is controversial, however. One might still want to claim that the agent's judgment is required to ground the agential authority and value of plans and policies. We can see some version of this claim in 'cognitivist' theories such as those of Korsgaard, Wallace and Watson. I set this debate aside because its outcome doesn't affect my argument against non-normative accounts like Ainslie's. However I return to this concern in Chapter 5 – it appears that many addicts evaluative their coordinated plans and policies negatively and it seems those judgments have a good claim to agential authority. This is a problem for Bratman's account.

intention. If you cannot achieve a norm of diachronic stability in *any* of your intentions then you will be reduced to a time-slice agent.[8]

The second norm requires the agent to intend means coherent with their ends. If you will an end then you must will some means to that end or you will never achieve it. For example, I will fail to become a professional soccer player if I never intend to go to training or join a team, and so I can hardly claim to have professional soccer playing as one of my ends without intending those means. Those means in turn entail other means. If I intend to join a team then I must intend to get in contact with a team, find out when and where training is, et cetera. Following the norm of means-ends coherence, therefore, results in a proliferation of detail in one's network of plans and policies. Of course the agent cannot, and should not, always set out the means for an end in full detail straight away, particularly for long-term plans. Excessively detailing one's plans in advance will waste cognitive resources because natural epistemic limitations in predicting the future entail the need to respond flexibly to unexpected contingencies.

The third norm requires that the agent's plans and policies be ends-ends consistent; the agent should not commit to plans and policies that are not co-possible. This is most obviously the case when one plan requires an action that the other forbids, for example, to become a Catholic priest and a husband. Other policies and plans are incompatible because of the demands they make on the agent's time and effort, for example it would be impossible for most people to meet the standards of a professional soccer player and a concert pianist. Many clashes can be resolved by creating hierarchical structures, for example, the policy of enjoying the sensual pleasures of life clashes with the policy of financial prudence because each regularly recommends a different course of action. But, if the agent subordinates one to the other, the clash can be resolved. One might, for example, only act to achieve sensual pleasure when such action is within the bounds of financial prudence. In order to meet the normative requirement of ends-ends consistency agents need to carefully select their plans and polices and organise them into hierarchies so that they are mutually attainable.

[8] We also saw that intentions cannot be too inflexible. To not reconsider an intention despite a significant change in circumstance leads to irrational action. An agent who was excessively stubborn in all intentions might count as a kind of diachronic agent although one with serious agential impairment.

The three norms of practical rationality tend to support each other. Working out means to ends supports diachronic stability because the agent gets clear on what they have to do and when.[9] Working out means to ends also helps ends-ends consistency because it reveals potential clashes between ends, e.g. one might only notice schedule clashes upon detailing the means to different ends. Ends that are more consistent with each other are easier to stabilise diachronically because they don't have to be planned around each other so much. For example, one might decide to get a breed of dog that will also be able to come jogging so that one can more efficiently satisfy the plan to care for a dog and the policy to maintain one's fitness. According to Bratman, the agent constitutes themselves by following these norms of practical reason. The resulting plans and policies tend to form an interlocking network that supports cross-temporal organisation of practical thought and action in ways that involve Lockean ties of cross-reference and continuity. Those plans and policies have agential authority; they represent the diachronic entity that is the agent (2007, pp. 32-33).[10] Poorly following these norms will tend to break down those cross-temporal ties, damage the agent's temporally extended identity, and reduce the agential authority of the plans and policies they have committed to.

Self-governing policies

Bratman claims that agents direct their action when they act on plans and policies. Self-governance and strong reflectiveness, however, require something more. That something more is provided by policies that the agent adopts specifically in order to improve cross temporal organisation – self-governing policies. Self-governing policies are designed to help the agent meet the norms of diachronic stability, coherence and consistency by controlling the weight that desires should be given in deliberation, i.e. what desires should be treated as justifying reasons.

> "We should understand an agent's endorsement of a desire in terms, roughly, of a self-governing policy in favour of the agent's treatment of that desire as providing a justifying

[9] As supported by empirical evidence shown in the last chapter where forming implementation intentions increased the chances of acting on goal intentions.

[10] Of course we aren't born with the immediate capability to recognise norms of practical reason and we clearly need to achieve a degree of unity before we can get to that point. Here I assume Bratman would agree with Levy who suggests that the initial process of unification is sub-agential. Through a causal process a coalition of sub-agential processes develops that can think of itself as a single being. It is minimally unified at this point and capable of action. As the agent develops more sophisticated capacities for planning, the norms of practical reason become recognisable. "Sub-agential mechanisms build the self that will then continue to shape itself" (Levy, 2006, p. 439).

> reason in motivationally efficacious practical reasoning. And it is, roughly, a policy against treating a certain desire as providing a justifying reason in motivationally efficacious practical reasoning that is characteristic of the … unwilling addict, and the like – cases paradigmatic of the agent's rejection of a desire" (Bratman, 2007, pp. 39-40).[11]

By attempting to control one's desires in this way the agent reflectively aims at increasing the number and stability of inter-temporal connections that make up their diachronic identity. For example, you might support your intentions to be a fair player of competitive sports and to be a good role model for your son by creating a policy to not treat the desire to engage in petty arguments with authority figures as a reason in deliberation.[12] Other examples Bratman cites are not so obviously focused on any particular desire: developing a concern with honesty in writing, trying to be more playful, and trying to be less impatient with others (2007, p. 33). Bratman also suggests that quasi-policies where one tries to fit within the parameters of certain roles, such as being a good citizen, may play the same self-governing role (2007, pp. 42-44).

Like plans and policies in general, self-governing policies have to be ends-ends consistent. The goal here does not have to be to avoid absolutely all clashes between self-governing policies but to be 'satisfied.'

> "We can say that one is satisfied with such [cross-temporally organising] attitudes when, roughly, other relevant attitudes of that agent at that time do not exert significant pressure on that agent for change of those attitudes" (Bratman, 2007, p. 268).

Organising self-governing policies so that one is satisfied tends to prevent the agent from holding self-governing policies from which they are estranged.[13] It also blocks the need to cite an infinite

[11] Bratman clarifies that an unwilling addict may still endorse satisfying a desire for drugs because of a general coping policy – 'I must rid myself of this desire if I am to achieve any of my goals.' To do so is not the same as having a policy to satisfy the desire for the drug for its own sake which would characterise a form of willing addiction (2007, p. 38). Although cases of addiction cannot always be so neatly differentiated. Some people appear to adopt sophisticated drug-using plans and policies for their own sake and yet also appear to lack self-governance; I return to this issue in Chapter 6.

[12] To have self-governing policies is therefore to be a reason-responder and not just a reason-tracker (Jones, 2003, pp. 189-191). Reason-trackers can behave in accordance with reasons without reflective concern for their temporally extended identity and so they do not treat those reasons as reasons. Reasons-responders treat reasons as reasons; they can detect when reasons are outweighed or defeated and respond flexibly. In Bratman's terms, a reason is outweighed or defeated when some other self-governing policy will provide more inter-temporal connections.

[13] However some cases of addiction pose a problem for Bratman. An agent might have a lone self-governing policy not to treat desire for a drug as reason-giving in deliberation. Yet all their other plans and policies might have been developed around a drug-using lifestyle. They could most easily achieve satisfaction by dropping the policy against drug-using desires so why don't they? And why would we typically see that as a *failure* of self-governance?

regress of even higher-order self-governing policies that endorse the lower-order ones. We can also say where the agent stands without positing a homunculus – the agent's self-governing policies, with which they are satisfied, represent where they stand. The targets of their policies ultimately stem from a 'single marketplace' of desire but those desires underdetermine how what the agent commits to with plans and policies and how they decide to organise them. If the agent breaks with their self-governing policies they undermine their diachronic existence and, ultimately, their claim to a standpoint independent of the fluctuations of the marketplace of desire.

Self-governing policies versus singular diachronic plans

I disagree with Bratman's emphasis on self-governing policies over other normatively vetted, singular diachronic plans. If plans were typically 'first-order' so that the agent committed to an end based on merely having a desire for that end then, clearly, self-governance would depend on something more reflective and higher-order, such as self-governing policies. But plans create inter-temporal connections and are usually only adopted once the agent is satisfied they fit with existing plans and policies, they therefore have the same potential for agential authority. For example, one might commit to a certain career path because it provides a steady income, an income that supports multiple other plans, e.g. buying a house, planning a holiday, starting a family. That career plan also implicitly suggests how desires should be governed in a range of circumstances, for example, resist the desire to insult the boss. Cullity and Gerrans (2004) have made a similar point and, in response, Bratman partially agrees, conceding that singular commitments can be self-governing. However he maintains that self-governing policies have more agential authority because singular "commitments involve weaker connections to temporally extended agency" (2007, p. 189). So the issue of how much agential authority a particular plan or policy has seems to be a matter of degree depending on how much diachronic connectivity that plan or policy affords. I suspect that some plans will provide as much if not more diachronic interconnectivity than many self-governing policies. But however this issue is resolved it doesn't have a serious impact on the debate between normative views like Bratman's and Ainslie's non-normative view.

Summary

We can conclude that some plans and policies provide more numerous and more robust inter-temporal connections than others. Furthermore, some plans and policies support the inter-temporal connections of other plans and policies better than others. Those plans and policies that are more central to the diachronic stability of the overall network can be said to have more agential authority and be more valued; when the agent acts in accordance with them they are more highly self-governed. Likewise, plans and policies that are more peripheral to the network because they provide less diachronic stability or undermine aspects of the network are less valued and have less agential authority; the agent exhibits less self-governance when they act on them when that involves acting against more central plans and policies.[14] Bratman's account, therefore, provides a gradient of self-governance according to how well the agent manages to follow the plans and policies that fit with norms of practical rationality.

Introduction to the benefits of normative accounts

With Ainslie and Bratman's positions laid out I'm now ready to address the arguments in favour of the explanatory power of normative accounts, but first, a brief introduction for those arguments:

1. The phenomenology of temptation involves a sensation of effort. Temptations have to be effortfully resisted, while succumbing to a temptation typically comes with a certain relief. Normative accounts can explain the effort of temptation because the option that best fits with current plans and policies may diverge from the option that is currently most desired. On Ainslie's account, rewards vary in size and timing but they all provide the same 'stuff.' If all rewards are commensurable the agent has no rational basis to oppose the pursuit of the greatest reward they take to be available. They can therefore regret past failures of

[14] This can ground a response to the following objection. Kennett (2001, p. 67) argues that hierarchical accounts that place no evaluative constraints on the content of higher order desires are too permissive in their attributions of weakness of will. If I had a first order dislike for strawberries but a higher order policy (that I woke up with one day) to like strawberries then it appears I am weak willed when I turn down strawberries. "…Not only has my second-order desire [or policy] not moved me to action, we feel that it has not provided me with a reason to act, either. It has not provided me with a consideration in favour of eating something I dislike" (2001, p. 67, my brackets). Bratman can respond that one is more weak-willed when failing to act on a policy that supports more inter-temporal connections. Because a whim to develop a taste for strawberries is unlikely to ensure many inter-temporal connections, it is barely weak to contravene it. In the case where the agent developed the (self-governing) policy because they were going to have to run the family strawberry growing business then contravening it will be more damaging to diachronic identity and, therefore, more clearly a case of weakness of will.

judgment – when they realise that a greater reward was available but they did not notice – but never wrestle against their current strongest desire in the moment.

2. Our folk-understanding of agency considers agential control over action to come in degrees. To fail in self-governance does not necessarily reduce an action to unintentional behaviour. Actions vary in their self-governance depending on such factors as the nature of the plans and policies relevant to the situation, the strengths of temptation present, and the agent's willpower. Distinguishing degrees of control in action is essential for our fundamental folk-practices, e.g. attribution of praise, blame, weakness or strength of will. Ainslie's view, in contrast, provides no room for degrees of self-governance. Behaviour counts as action when it aims at maximising reward; this category combines controlled action with reckless and weak-willed action. Behaviour that aims at anything less than maximising rewards is considered unintentional movement and includes akratic behaviour.
3. Our folk-practices work on the assumption that the agent can adjust her desires by working on them. This is a central feature in normative accounts – the agent tries to govern her desires using plans and policies. These acts are attributable to the agent and can be distinguished from changes in desire that are the result of extra-agential forces (e.g. biological maturation). For Ainslie, in contrast, agents cannot exert any control over desires because the agent just *is* the dominant faction of desires of the moment; they have no independence from those desires. Therefore long-term desire change has to be explained by biological and/or environmental factors.

Phenomenology of temptation

We can distinguish three interrelated aspects of the phenomenology of temptation that need to be explained. First, resisting temptation involves a synchronic, effortful struggle while succumbing to the temptation relieves that effort. Second, temptations have a persistent affective or visceral appeal despite the agent knowing they are worse in some sense. Third, the tempting option often has unique qualities that the option judged best lacks and so there is something that can be regretted about *not* taking the temptation. Normative theories can better accommodate each of these phenomenological aspects than non-normative theories.

Normative agency and temptation

Normative theories describe the struggle of temptation as follows: Our normatively endorsed plans and policies favour one action but our strongest desire of the moment motivates another action that would require going against norms of practical reason. The struggle can play out in two phases.

In the first phase, as discussed in the previous chapter, the agent is tempted to *reconsider* plans and policies that preclude the temptation. There is a synchronic conflict because the agent is simultaneously aware of the action consistent with current plans and policies and the possibility of alternatives involving the temptation.[15] The agent struggles not to reconsider their intentions in order to meet norms of diachronic stability, thereby protecting their temporally extended identity and power of self-governance. As we saw in the last chapter, this experience of the struggle for self-control fits neatly with the executively controlled depletion of one's finite resources of muscle-model willpower.[16] "…sticking with a resolution does require ongoing effort. It requires a difficult form of mental control (Holton, 2009, pp. 178-179).

If the agent gives in at this stage and reconsiders what to do, there are three possible outcomes: (1) They convince themselves under conditions of judgment shift that taking the temptation is best. In outcomes (2) and (3), judgment shift is insufficient to result in the temptation being ranked above the original target of their normatively endorsed intention but the agent is still at risk of akratic action. In outcome (2) akrasia is also avoided because anticipation of cognitive dissonance ensures that the newly formed intention stays in line with best judgment. In outcome (3) the tempting desire is strong enough to override cognitive dissonance and motivate the formation of an akratic intention. If the latter situation occurs then there is room for a second phase of struggle.

When akrasia threatens, arguably the agent can effortfully tip the motivational balance away from the temptation and toward an action he judges best. This can be done either by increasing desire for the action judged best or decreasing desire for the temptation. There are a variety of stories as to how the agent can synchronically act against their strongest desire once they have reconsidered their decision. Addressing them in detail would take us too far afield but they are worth covering in brief. The agent might have previously developed dispositions that will take effect without them having to act against their strongest desire. Tempting cookies, for example, might then be seen as

[15] The alternative actions targeting the temptation might be fairly inchoate given that the agent hasn't yet reconsidered their choice or formed a new intention with new means.

[16] Although we aren't that good at distinguishing effort dispensed (i.e. fatigue) in self-control tasks from non-self-control tasks (Clarkson et al., 2011).

lumps of fat congealing in one's stomach (J. Kennett & Smith, 1996). This non-actional form of self-control, however, is unlikely to involve a phenomenology of struggle. Alternately, the agent might use an ancillary action (Mele, 1997). This method eschews trying to change the balance in the primary motivational conflict, e.g. the desire to act on best judgment and stick to the diet versus the tempting desire to eat the cookies. Rather, the agent tries to win an ancillary competition, perhaps to state out loud, "those cookies are just lumps of fat!" versus the desire to not utter that statement. This method would work when the outcome of the ancillary competition is sufficient to sway the outcome of the primary competition. Perhaps this is a possible means of control but it relies on the temptation being unable to assert itself across means-ends relevant desire competitions.

Sripada (2014) promotes an even stronger account of how the agent can directly overcome his strongest desire through action.[17] He recognises that his view requires a divided mind thesis where the mind is seen as having two motivational compartments – one available to judgment the other more automatic and visceral. The agent can overcome his strongest visceral desire in direct competition by drawing on the motivation available to judgment and just willing himself not to give in. A separate compartment of motivation available to judgment would also allow the agent to reinterpret the cookies as lumps of fat in the moment rather than rely on prior dispositions. Talk of 'compartments' is clearly metaphorical; their existence may solve the conceptual problem of how a strongest desire can be overcome but it isn't clear how to cash them out in psychological let alone neurological terms. Muscle-model willpower goes some way towards this, however; it represents a store of motivation available to judgment and compartmentalised from visceral desire. Judgment can thereby use muscle model willpower to directly overcome the desires of the more visceral system. As we have seen, there is empirical evidence for the existence of a finite resource of muscle model willpower although it is still uncertain how that works at the neurological level.[18]

If muscle model willpower is also needed for the synchronic resistance of akrasia then, if the agent failed in phase one (non-reconsideration) because they ran out of willpower, they will have nothing left for phase two. However, if they reconsidered their intention because they didn't bother using

[17] As he notes, his view is therefore similar to Holton's, in that he holds that "the exercise of willpower is a specific mental action … that agents actively employ" (Sripada, 2014, note 25). However, Holton focusses on non-reconsideration of intention while Sripada focuses on cases where a prior intention may have already been reconsidered.

[18] Some evidence suggests that error-related negativity signals generated in the anterior cingulate cortex are weakened in ego-depleted individuals. This suggests that the ego-depleted agent is worse at noticing discrepancies between their intentions and their actual behaviour (Inzlicht & Gutsell, 2007).

willpower (through laziness or in hope of conserving it), the immediate threat of akrasia might then motivate them to resist more effortfully in phase two. So, on the normative account we have two potential phases of synchronic conflict, one of trying not to reconsider an intention and one of trying to resist akrasia. Executive effort is required at each phase if the agent is to act in line with practical norms and to be self-governing.

The normative view of resisting temptation is also supported by evidence from dual-process theory. In dual process theory, system 1 is a collection of evolutionarily conserved mechanisms that respond to information automatically and rapidly. These mechanisms are domain specific and run in parallel. System 2 processes are relatively conscious, slow, and run serially. They are deliberate and require cognitive effort. System 2 appears to roughly correspond with a global workspace view of consciousness where the most important outputs from informationally encapsulated, system 1 modules are considered; the results of global thinking are then communicated back to the modules as their specific services are required (Levy, 2007, pp. 240-243). System 2 appears to regulate some of the output of system 1 but when system 2 is not deliberately employed, system 1 takes over.

> "All of this should immediately make us think of the ego depletion paradigm. Self-control, too, is slow, demanding and draining of cognitive resources. It is weakened or lost under conditions which look for all the world like the conditions which make agents switch from system 2 to system 1" (Levy, 2011b, p. 145).

In other words, the experience of temptations is often the result of some fast, automatic, evolutionarily conserved system 1 processes that occur independently of higher cognition. It is then up to the more executive, globally integrated (all-things-considered) processes of system 2 to inhibit those system 1 processes in order to follow norms of practical reason.

The second aspect of temptation phenomenology, the persistent affective or visceral appeal of temptations, can also derive some support from dual process theory. The visceral aspect of temptations is likely to result from the influence of system 1 processes because such processes have evolved to quickly and automatically generate motivation. Normative endorsement, on the other hand, relies on the slower system 2 processes that wait on cognitive effort and more globally integrated cognition.[19] Not only are these system 2 processes slower but they might only ever

[19] That's not to say that the normatively endorsed option doesn't have some affective appeal and the temptation doesn't make some cognitive sense. The ability of system 1 and 2 processes to influence each other is one reason why sex can be used to sell virtually anything.

counter system 1 processes without ever totally silencing them. This would explain why merely recognising that a temptation is incompatible with existing plans and policies would not automatically or completely override that contrary motivation.

The normative account also distinguishes choice situations involving temptation from those that are purely cognitive and affectively 'cool.' In purely cognitive choice situations none of the choices elicits much visceral appeal, for example, selecting a multi-choice answer to a maths problem is a purely cognitive choice. If someone can demonstrate that your answer is wrong and another is right then you are immediately motivated to choose the new answer and no longer have any motivation to choose the wrong answer. This distinction is important because, as we will see, Ainslie effectively reduces all choices to the cognitive variety.[20]

The third phenomenological aspect of temptation (and a range of other choice situations) is that the mutually exclusive options have incommensurable aspects. For example, the benefits of the healthy eating plan can't be substituted for by an unusually excellent dessert. The temptation and the normatively sanctioned option are often just different things.[21] To take an addiction related example, Russell Brand, an ex-heroin user, recently remarked:

> "It doesn't matter that I was sat in that flat in Hackney and now I'm in the Savoy. I'm jealous of me then. It doesn't make a difference to me. The money, the fame, the power, the sex, the women – none of it. I'd rather be a drug addict" (Brand in, Wilson, 2012).

The fact that Brand has not relapsed suggests that this statement does not really reflect a normative endorsement of being an addict. Rather, it suggests that there is something about drug-use, some value or quality of experience it yields, that his current success cannot replace.

The incommensurable aspects of different options can explain a couple of phenomena. First, it explains why I can continue to regret *something* about a temptation I successfully resisted ('that dessert would have been so nice!') even if I would still try to resist it if I had the choice again. Second, if the desire for that incommensurate aspect of the temptation remains even well after the fact, we can assume that it plays a role in making the temptation appealing at the time when it is

[20] On Ainslie's view, mutually exclusive options may each have a visceral element to the reward they promise but there is no way for a lesser reward with more visceral appeal to outcompete a larger reward with less visceral appeal. This is because Ainslie treats all reward as commensurable; the agent has to bias to favour the visceral over the non-visceral.

[21] Although this is not always the case, if the agent can take one marshmallow now or have two later then these options are not incommensurable.

still available. If this assumption is right it helps explain the appeal of temptations even when they go against endorsed plans and policies.

So the normative account explains the experience of temptation as a divergence between the normative endorsement of globally integrated cognition and a more viscerally appealing option usually with incommensurable qualities. In temptation both normative endorsement and strongest visceral desire underdetermine our action leaving it to the agent to struggle between two live options. Resisting temptation requires effort and succumbing to temptation allows one to stop putting in that effort. That saving in effort comes with the costs of cognitive dissonance and regret, either immediately in cases of akrasia or (usually) sometime later in cases of judgment shift. In contrast to temptation, other kinds of choices are more purely cognitive and are settled as soon as a judgment is made (if it can be made[22]). I now turn to Ainslie to see if he can explain these aspects of phenomenology.

Ainslie's view on temptation phenomenology

Ainslie's theory of action cannot easily capture the struggle of temptation because he considers all reward to be commensurable; the only relevant variables are reward size and timing. Visceral appeal, when it occurs, is one contributor among many to the size of the reward. The compatibility of different ends is important but the task of achieving compatibility doesn't fall to a semi-independent agent trying to achieve a norm, rather the reward expectations themselves form factions with the faction that promises the greatest expected reward causing action. Judgment is not something the agent does but the outcome of a 'marketplace without a governor.' Through this process some faction of reward expectations always wins out over the others (setting aside Buridan's Ass cases).[23] The resulting picture is a clear distinction between action, i.e. pursuit of greatest expected reward, and unintentional behaviour, pursuit of anything less than greatest expected reward. Presumably action comes with a phenomenology of being in control of one's

[22] Other choices are hard because making a judgment is hard, not because of temptation. Consider, for example, resenting the long menu at a restaurant and asking your wife just to order something for you; or choosing to pursue the career of a pianist or a football player. Interestingly, making these choices where judgment doesn't easily yield a clear favourite is also phenomenologically effortful and evidence suggests they also draw on muscle model willpower (Vohs et al., 2008).

[23] Ainslie combines this with a deterministic view that the agent must pursue the greatest reward they are aware of. However, it is commensurability rather than determinism that is crucial here because commensurability ensures that the agent would have no reason to pursue a lesser reward – there is nothing a lesser reward could provide that the greatest expected reward could not.

body (and tools) and unintentional behaviour comes with the phenomenology of one's body (or tools) being out of control.

Ainslie defines succumbing to temptation as a case where a smaller sooner reward is chosen instead of a mutually exclusive larger later reward. This can only count as action if the agent is unaware that they have passed up a larger later reward. The agent is unaware of this because of hyperbolic discounting which makes the larger later reward appear smaller than it really is. The problem with this picture of temptation is that it cannot involve a phenomenology of *struggle* because there are never two synchronically available possibilities that cannot be clearly ranked based on expected reward (and that ranking translates directly into motivation to pursue the highest ranked).[24] Succumbing to temptation should therefore have exactly the same synchronic phenomenology as any fully controlled action. But it doesn't.

Ainslie's response to this concern is to assert that the folk-understanding of temptation phenomenology is wrong. Effectively, he attempts to redescribe the phenomenology of temptation as a purely cognitive clash between two conceptions of the choice – a synchronic conception as a discounted LL reward versus a non-discounted SS reward and a diachronic conception as bundles of LL rewards versus bundles of SS rewards.

> "The person will not experience this situation as the voyage past some siren or other, but as a simultaneous struggle between two ways of *conceiving* a choice" (Monterosso & Ainslie, 2009, p. 125).

His response requires two things of us. First, we must find another way to understand the phenomenology of temptation I have highlighted. Is it just epiphenomenal or is it actually connected with some other aspect of action? If the phenomenology we call temptation is actually arising from something different what is it? The folk can get things wrong but in this case I suspect these alternative explanations of 'temptation' phenomenology are dead-ends. In any case I won't pursue them here since I have other more weighty reasons to abandon Ainslie's project. Second, we need to try to understand how a choice between "two ways of conceiving a choice" could be an effortful struggle as Ainslie claims.

I admit there is *some* effort involved in forming different conceptions of a situation, certainly forming ones that are less habitual. For example, one might have to use effort to try and see the 3D

[24] Just to be clear, this ranking process is achieved by sub-agential systems not the agent themselves.

image hidden in an apparently 2D picture, or see an apartment as an investment rather than as a home. But in these cases, as in the purely cognitive choice of the correct answer to the maths problem, once we know the best way to conceive of a situation, then the struggle is over. We can then see it in the more beneficial way each time we need. If bundling choices together in test-case willpower was the best way of conceiving of a choice then why wouldn't we just always conceive of choices in that way and thus never feel tempted by (or succumb to) bundles of SS rewards? If rewards are commensurable then the less beneficial way of seeing the situation, without bundling choices together, should hold no appeal once we're familiar with the better way.[25]

Ainslie's theory implies that one or other conceptual framework (diachronic or synchronic) will be completely dominant at any one time. But our typical experience of temptation doesn't involve *complete* shifts in conceptual framework. For example, when I face a temptation, say a delicious dessert, eating the dessert doesn't entail forgetting my healthy eating plan. In fact, I may regret the dessert as I eat it so that my enjoyment of it is partially diminished because I am still aware of my healthy eating plan. Likewise, if I pursue my healthy eating plan I still might feel some regret because the dessert would have been delicious. I often remain aware of the synchronic aspect of my existence even as I pursue long-term goals and vice versa. I have two easily accessible conceptual frameworks for choice and self-conception that influence each other.[26] A normative view can accommodate this possibility because recognising the normatively endorsed option doesn't totally supress unruly desires, and desires don't necessarily blind the agent to their existing plans and policies.

Ainslie eliminates the conceptual space needed for a struggle between two options because, for him, there is nothing an SS reward can provide over an LL reward except that it arrives sooner. On such a picture it is impossible that an SS reward mutually exclusive of an LL reward could remain appealing (viscerally or otherwise) to an extended agent who is capable of bundling choices

[25] This is assuming that we are temporally extended beings. The struggle could be thought of as being between conceiving of oneself as a person time-slice and conceiving of oneself as a temporally extended person. If the agent was just a series of time-slices then, in contrast to the extended person, the SS reward would provide the best reward for each time-slice and bundling rewards would be useless. But we run into the same problem. Being a series of time-slices or extended is decided by what self-conception maximises rewards. If maximum reward comes from SS rewards then you are a series of time-slices and there's no point in bundling choices, but if time-slices are concerned by LL rewards then you happen to be extended and you will benefit from bundling. Once you have discovered whether you are extended or synchronic then you proceed as follows; there is no need to struggle.

[26] Perhaps I can only think from one perspective at a time so that one or the other dominates at any stage like a kind of gestalt switch, but it seems that even while I think from the perspective of one the other continues to influence me.

together.[27] Yet humans are extended agents who can distinguish temptations from what is best (by bundling or otherwise), they often remain tempted despite their judgment and they must struggle if they want to act on that judgment. Being unable to accommodate a specific aspect of phenomenology might be passed off as a fairly unimportant failing, however in this case it is the tip of a serious iceberg. Phenomenology is important because it grounds human practice. Our folk-understanding that temptations can be effortfully resisted grounds various practices including those of praise, blame, pride and disappointment.

Degrees of control in action

Degrees of self-governance on a normative account

Human agents don't just distinguish cases where effort is called for, we also judge the *degree* to which situations demand efforts of self-control and therefore the degree to which praise, blame and so on should apply. [28] To accommodate these folk practices we need some way to account for degrees of control. Ainslie's account does not provides this but normative accounts of agency can describe degrees of self-governance across both synchronic and diachronic time-frames. I begin with the account in synchronic conditions and then move on to degrees of control over time.

Degrees of self-governance in synchronic conditions can be explained with reference to the magnitude of the clash between endorsed networks of intentions and contrary desires. Some desires are stronger than others, last longer than others, or are more unpredictable than others and this varies by agent and context. If the strongest desire of the moment is contrary to normatively endorsed plans and policies then it counts as a temptation. The greater the motivational deficit that must be overcome to act normatively the stronger the temptation and the more effort that is required to successfully resist it. When the agent successfully overcomes a large deficit this indicates much effort dispensed and warrants greater pride (and praise); failure against strong temptation doesn't

[27] Or that one could later regret not taking it because one missed out on its incommensurable aspect.

[28] Admittedly we go beyond Bratman's account of self-governance once we bring inter-subjective praise and blame into the equation because this depends on how well the agent tracks social standards of the good. Bratman's account of self-governance only describes degrees of control in meeting one's own standards, which may or may not incorporate moral standards. Therefore it can only ground degrees of intensity in self-directed attitudes (and intersubjective attitudes of praise and blame insofar as we hold someone to their own standards). I continue to set the further standards of morality aside whereby the agent can be blamed for having plans and policies that contravene morals even if they perfectly meet norms of practical reason.

necessarily indicate a weak will and doesn't warrant so much self-blame (and social opprobrium). Equally it doesn't take much strength of will to overcome a small deficit and so no great pride (or praise) in such success is warranted. In contrast, to fail to overcome a weak temptation indicates a weaker will, and such situations warrant significant self-blame (and opprobrium).

The picture is further complicated by the fact that agents' commitments to intentions also vary in intensity. On the Bratmanian account, those intentions that support numerous inter-temporal connections are more strongly endorsed than those that support fewer. A self-governing policy to never entertain the desire to smoke might be more strongly endorsed than a simple intention to not smoke tonight because the former supports other plans and policies of good health, raising one's children, living a happy retirement with one's wife, et cetera. Therefore a temptation that threatens a more strongly endorsed intention *should* warrant more effortful resistance because failure here more seriously disrupts one's normatively organised network of plans and policies. Giving in to a temptation that only disrupts a peripheral intention is not so detrimental and so, although the agent should also resist those temptations, the normative strength of the 'should' is weaker. So maximal self-governance requires one to follow all one's normatively endorsed plans and policies. Achieving that is more or less difficult depending on the motivational deficits that have to be overcome. The agent can be overwhelmed by temptations but they retain a degree of self-governance to the extent that they protect their normatively endorsed intentions (particularly the more central ones).

Effortful deployment of finite muscle model willpower neatly fits into this picture. The agent can be proud of putting in the effort to resist temptation or ashamed of giving in too easily. Because muscle model willpower is finite, the more intense the temptation and the longer it remains available, the harder it is to resist. The more significant the intention being defended then the more effort (willpower) the agent will be prepared to spend to maintain it.[29] Furthermore, the agent might strategically (if usually implicitly) succumb to temptations that only threaten peripheral intentions in order to save willpower in case more significant intentions come under threat, for example, you might eat a chocolate biscuit instead of lighting a cigarette.

[29] It's worth emphasising that this strategic deployment of muscle model willpower depends on the agent being able to detect a characteristic phenomenology that alerts them to the threat a temptation poses and the strength and significance of that threat. The phenomenology of a desire that doesn't clash with an existing intention is quite different – in these cases one can just 'go with the flow.'

So far I have presented degrees of control in synchronic situations where the strengths of temptation and agent-endorsed norms are taken to already be set. Human life is of course not just one synchronic situation after another and degrees of self-governance are more complicated in diachronic focus. Diachronically the agent has some control over what temptations arise, what plans and policies they have and the various powers they have available to resist temptations. By carefully structuring their plans and policies agents can reduce the risks of temptation in advance. Temptations can be caused by adopting ends-ends inconsistent plans and policies, for example, you might glibly make promises to your family and to your work colleagues but then find that due to a lack of foresight you will have to break one or the other. Similarly, a morning exercise regime is threatened by going out to the pub until late the night before because that creates a strong desire to sleep in. Careful effort to avoid conflicting plans and policies will reduce temptation, drains on willpower, and losses of self-governance. Agents can also choose to follow plans and policies that clash less heavily with their typical temptations to improve their chances of successfully following those plans and policies. If, as I argue in the final section of the chapter, desires can be changed to better align with plans and policies, then the agent might slowly escalate the ambition of their plans and policies at a rate that avoids more intense temptations. The earlier, less ambitious plans and policies can then be considered means to the ends of later more ambitious plans and policies.

Agents may, therefore, be held accountable for succumbing to even the strongest temptations if they brought those temptations upon themselves or failed to develop a long-term strategy to counter them despite having ample time and resources to do so. An agent who carefully considers the consistency amongst his intentions therefore exhibits greater self-governance than an agent who doesn't.

In addition to correct use of one's skills for responding to synchronic and diachronic temptations there is the possibility of improving or neglecting the skills themselves. Some agents have better developed planning skills than others. Some agents have greater resources of muscle model willpower than others and/or use the limited resource more strategically. Because the agent can improve these techniques if they try, they can be blamed for lacking them (assuming they had sufficient opportunity) and praised for developing them. Succumbing to a temptation is more blameworthy if either you are culpable for underdeveloping your skills or if you developed them particularly well but couldn't be bothered to use them. Conversely, if you resist a temptation despite having limited skills (that you are not culpable for lacking) then that is worthy of high praise.

So a normative account offers a rich array of variables that influence the degree of control an agent has over their actions both at a time and over time. Synchronically, the agent can effortfully dispense more or less muscle-model willpower in the face of temptations of varying strength to defend intentions of varying degrees of endorsement. Diachronically, the agent can put more or less effort into developing means-ends coherence and ends-ends consistency. Diachronically, the agent can also put varying effort into developing the skills of achieving means-ends coherence and ends-ends consistency and developing their pool of willpower. This provides the basis for the degrees of pride, disappointment, praise, blame and so on. Let's now turn to Ainslie's account.

Degrees of control on Ainslie's account

Ainslie's account cannot accommodate degrees of control because action is defined as being moved by the greatest expected reward. Pursuit of anything less than the greatest expected reward is non-intentional behaviour. In the former case one acts with as much control as one could hope for and in the latter case one does not act at all. The agent will put in effort to overcome the extra-agential forces that stop him accessing the greatest expected reward but he will put in no more effort than the reward promises to compensate. As the costs of pursuing a reward increase, other rewards with better cost/benefit ratios will determine action instead. So there is a sense in which the agent engages in effortful activity but this effort is necessarily proportional to the expected reward. The agent would never put effort in to resist a dominant reward expectation because all other actions would provide less reward and so to pursue them would be irrational. There can be no variation in strength of will because that depends upon a genuine struggle between two possible paths of action that each have their own merit.

As we saw, Ainslie describes weakness of will as a case where the agent takes the SS reward instead of the mutually exclusive LL reward. Given that the largest expected reward determines action, the agent can only pursue the SS reward if they *fail to realise* it clashes with the LL reward or if they fail to realise that it is in fact a smaller reward. A synchronic struggle that could demand more or less effort on the agent's behalf just doesn't occur.

One might argue that, on Ainslie's view, the agent is still culpable for pursuing an SS reward when they *should have* noticed it clashed with the LL reward. There are degrees of control in the amount of effort they put into deliberation and thus what action now will provide the most overall reward.

The problem with this response is that it conflates weakness of will with recklessness. An agent is reckless if they make judgments without sufficiently careful deliberation, when they had the time and skill to do so. Recklessness is a different failing to weakness of will and should be treated differently; we lose valuable explanatory power by conflating them. Most glaringly, the conflation precludes the possibility of clear-headed cases of akrasia where the agent knows all too well that succumbing to the temptation is mutually exclusive of a reward that is better in some sense. An example is the addict who clearly sees the rewards of abstinence and the costs of drug-use but then deliberately takes the drugs anyway. These agents won't benefit from exhortation to better consider their decisions. One might argue that cases of clear-headed akrasia don't really exist. However, some of the cases of addiction I address in Chapter 5 suggest they do and, in any case, there are many examples of addicted behaviour that certainly doesn't appear reckless or aimed at maximising reward.

Ultimately Ainslie can distinguish action from unintentional behaviour but he makes no conceptual room for weakness of will as a failure of self-governance distinct from recklessness let alone degrees of self-governance in the face of those challenges.

The relationship between agents and their desires

In this section I argue that a normative account of agency can explain two kinds of long-term interaction between an agent and their desires: (1) the agent's desires permanently[30] change without the agent's intervention because of biological maturation and/or processes of cultural habituation. (2) The agent adopts plans and policies to govern their desires and this comes to permanently shape those desires. Ainslie's theory lacks explanatory power because it tries to reduce all these kinds of interaction into (1). I begin with the normative account of these two types of interaction before considering Ainslie's reductive account.

[30] Here I use the word 'permanently' in a relative sense to indicate something much more temporally stable (or much more slowly changing) than the oscillations we see caused by passing temptation. Such 'permanent' changes in desires or norms may not, of course, be absolutely permanent.

Natural long-term desire change

Desires often change over time without any input from the agent and sometimes the agent might make no effort to intervene in such processes. For example, a child might grow out of playing certain games or an adult might find that they have developed a taste for asparagus or jazz. This will typically result in some associated changes in plans and policies but those changes might be merely reactionary rather than an attempt to control the desires. That is, the agent might just form intentions that respond to the desires as they are, for example, asparagus gets put on the shopping list sometimes, tickets to jazz events are occasionally purchased. Such reactionary changes are typical when we judge that our natural changes in desire are non-threatening to our normative network. We don't usually need to exert much control over how much we like asparagus, for example, because our more central plans and projects are not undermined by strong asparagus temptations. But we cannot take such a neutral view to other natural changes in our desires.

Sometimes desires change for the worse and if you found that you were developing paedophilic desires or an overly strong desire for drugs or gambling then you may well want to inhibit or extinguish those desires.[31] Conversely certain desire changes might be seen as highly beneficial and rather than just react to those desires you might want to actively foster them. So if you started to enjoy jazz, rather than just go to some jazz concerts, you might deliberately seek out particular styles and artists to help broaden and deepen your enjoyment of jazz. Regarding recovery from addiction, natural processes of maturation might begin to reduce the problems an addiction causes. The agent could just hope that this happy trend continues; however, if the agent has suffered a significant negative impact from addiction, surely they would want to reinforce and accelerate the beneficial trend.[32] In summary, agents may take a *laissez faire* attitude to natural changes in their desires but often they will want to do more than this.

[31] Again, recall that on the Bratmanian account these desires are bad because they fail to fit with the existing set of normatively endorsed plans and policies and so threaten self-governance. It's a further question, that Bratman sets aside, as to whether desires track intersubjective standards of the good.

[32] Heyman's account, for example, describes all cases of recovery as the result of the environment and desires conspiring to promote global thinking; the agent is swept along in favourable currents.

Agent driven changes in desire

Bratman specifies self-governing policies as the means by which agents control the weight of desires in deliberation. Such policies are useful even if the longer-term frequency, strength and duration of a desire doesn't change in response to those policies. For example, a policy not to entertain the desire to smoke might be central in my network of plans or policies despite the desire to smoke still arising with the same strength and frequency as it always did. But what would be even better is if that desire to smoke started to fade in strength and frequency. I might then still have the policy but not have to use so much willpower to implement it.

It is our folk understanding that agents can, over time and with effort, change the intensity and frequency of their desires and even generate novel desires and extinguish existing desires through enacting intentions. We are born with certain desires, other desires come upon us or fade, and yet others we deliberately cultivate or extinguish. Not all desires are equally controllable however. Some desires may be impossible to fully extinguish (e.g. hunger and thirst) or impossible to develop (e.g. for some, a taste for Brussel sprouts), others may be impossible to stop from fading away entirely (e.g. enjoyment of childhood games). However there are many desires which can be extinguished (e.g. smoking) or that can be maintained when they would otherwise naturally fade away (e.g. enjoyment of a playing an instrument). The evidence for these intention-controlled processes is largely anecdotal and built on the shared, accumulated experience of the human race. Addicts who have abstained for longer periods of time tend to report lower intensities and frequencies of craving (Addolorato et al., 2002; Shi et al., 2009; Weddington et al., 1990). People who force exercise upon themselves quite often come to report that they now exercise more willingly. How does this long-term control of desires work?

One can encourage one's desires to flourish by *trying* to like things. For example I might remember liking soccer last season, judge that playing soccer this season would be best but find myself feeling apathetic towards it nonetheless. My self-governing policies can help ensure that I play soccer anyway by guiding me to give weight to my desire for aerobic fitness and team comradery while other self-governing policies guide me to give no weight to inclinations to be lazy. Once I begin to play again, the more visceral rewards of soccer become fresh in my mind and the motivation to play arrives more easily. Of course, this process of fostering a desire isn't always difficult because some desires consistently drive their own development from the outset.

Conversely desires can be limited or even extinguished through consistent application of self-governing policies not to indulge them and development of self-governing policies that encourage competing desires. If these efforts are consistent enough then the desire will tend to fade. Without experiencing the reward the body will 'forget' the cues and the reward that used to be associated with an action. So even though I loved playing soccer last season, after several months not playing all I see is a weekly commitment without the enjoyment. Of course playing soccer isn't a desire that one typically wants to be rid of. Other desires, particularly addictive desires are much harder to inhibit because the desire is strong, the associated cues are common and achieving any decent length of abstinence is therefore much more difficult. For some desires, even long-term abstinence might never totally extinguish the desire and there is some indication that addictive desires and cues for those desires are like this.

We can, therefore, add the skill of long-term desire control to the list of diachronic degrees of control that agents can exhibit. The agent is somewhat responsible for the long-term development of their desires. They can be blamed for letting desires contrary to their normatively endorsed plans and policies develop and for failing to develop motivational dispositions that ease self-governance.

Ainslie's account of long-term desire change

On Ainslie's view, category (1) explains all desire change in the longer-term. Expected reward is the result of extra-agential forces, i.e. sub-agential motivations and environmental affordances. Rewards change in *appearance* due to hyperbolic discounting but here I am interested in how the pre-discounted, base-value of rewards change. Ainslie takes it that, whatever delay-independent value rewards have, they are objective features of the world and, therefore, by applying the scientific method we can discover what they are, how strong they are and presumably understand how they change over time.

If we could get a complete understanding of desire dynamics in this objective way then this extra-agential category of desire change *would* be sufficient to explain action. However there are two problems with this approach. The first problem is that desires may not be objective phenomena and we might need to rely on first-person reports to find out what a person's desires are. This would doom the objective project from the outset but I won't pursue that problem here. The second problem is that we currently only have the sketchiest means of objectively determining what a

person's desires are and understanding how they changed by extra-agential forces. We know, for example, that bodily changes around puberty will typically result in some sexual desires although we don't currently know why sometimes such desires are heterosexual, homosexual or somewhere in-between, and, occasionally, such desires don't arise at all and people remain asexual. We know that the persistent use of drugs that flood the brain with dopamine will tend to increase desire for those drugs and make those desires extremely cue sensitive. But again, some people appear to develop much stronger desires through this process than others. As we look at more specific desires in individuals the challenge becomes even more difficult. Why does a particular person develop a desire for morris dancing and another doesn't? The agents might tell us their own explanations with reference to their plans and policies (or evaluative judgment) but in terms of an objective description we just don't know what's going on. We might assume there is *some* explanation tied to the agent's physical and social details but if we don't know which physical and social details are relevant and how they have their effect we are left with no explanation. The concern for Ainslie's view is that if he wants to refer to extra-agential forces to explain all desire change then he seems to be ignoring the ample folk-data on the subject. If our folk-understandings of long-term desire change are flawed there is currently nothing of substance to significantly contradict or replace them.

Conclusion

I began the chapter by outlining Bratman's planning account of agency. On his account the agent attempts to develop a network of plans and policies in accordance with norms of practical reason. When successful, those plans and policies ensure inter-temporal connections that define her diachronic identity and guide deliberation and action consistent with that identity. Self-governing policies are adopted specifically to improve inter-temporal connections and, therefore, they tend to have the most agential authority because they ensure the most inter-temporal connections. Temptations motivate intentions that would contravene the norms of practical reason, undermining the ends-ends consistency and diachronic stability of their existing network of plans and policies. The agent wields executive control over temptations through consistent application of the norms of practical reason including, when necessary, use of muscle model willpower.

Bratman's normative account provides more explanatory power than Ainslie's non-normative account in three broad ways. First, it can explain the phenomenology of a struggle against temptation because it allows for a synchronic clash between a more cognitive, abstract, globally integrated judgment and a more viscerally appealing temptation. The agent struggles because there is (initially) more synchronic desire for the temptation than the normatively endorsed option; only the effortful use of muscle model willpower can ensure the agent acts according to their normatively vetted plans and policies. Second, Bratman can explain the differing *degrees* of agential control, wielded both at the moment temptation strikes and across time. Synchronically the agent can put in more or less willpower to overcome temptation. Some temptations are stronger, last longer or are more unpredictable than others making control more difficult in those cases. Some intentions ensure more inter-temporal connections in one's network of plans and policies than others; failing to put in the effort to defend them is a more serious lapse of control than failing to defend more peripheral intentions. Diachronically, the more effort the agent puts into ensuring means-ends coherence and ends-ends consistency of her plans and policies the less likely she is to face temptations.

A further aspect of this diachronic control involves the third explanatory benefit of Bratman's account. Bratman can distinguish extra-agential changes in desire over time from agentially-driven changes in desire through plans and policies. The more consistently the agent applies norms of practical reason, particularly by creating policies that govern desires, the more they will find that their desires tend to fall in step with their plans and policies. The less consistently they organise and apply their norms the more likely they are to be dogged by temptation. So, greater self-governance isn't just achieved through use of willpower but by decreasing the need for that synchronic control by normatively organising plans and, as far as possible, training one's desires to fall in step with one's plans and policies.

Ainslie misses out on all this explanatory power because he reduces the agent to the mechanical play of reward expectations that only change according to biological and environmental processes. Because rewards are commensurable there is nothing to oppose to the greatest reward expectation and without two feasible options there cannot be a struggle. In any case there can be no rational basis to struggle against the greatest reward expectation. The phenomenology of temptation is, therefore, a mystery. When the agent acts on their strongest desire they exhibit as much control as they could hope for while anything less is unintentional movement. That leaves no way to capture the idea that an agent could be mostly, barely, almost, or far from governing their action. Although

folk beliefs in degrees of control grounds many self-reflective and intersubjective attitudes and practices, Ainslie has to say those beliefs are delusional. Finally, Ainslie insists that all changes in desire over time are extra-agential even though we are far from having a detailed objective theory of these processes; meanwhile he ignores the significant correlation between the agent's efforts to control their desires and subsequent desire change.[33]

Agency as it stands

Before concluding the chapter it will be helpful to summarise the picture of agency we have arrived at. In the first two chapters I introduced a variety of diachronic structures for action: global choice, prudential rules, singular diachronic plans, personal rules, pre-commitment and intentions. In Chapter 2 we saw that these structures are inevitably threatened by contrary desires (temptations) which cause judgement shift and akrasia. I argued that intentions stand out from these other structures for action because they can be endowed with an inherent diachronic consistency (potentially further protected by muscle model willpower). Furthermore, because intentions can take the form of rules, singular diachronic goals and global choices, intentions underpin the diachronic stability of all those structures. In addition, pre-commitments can support intentions over temptations by biasing the various costs and benefits in the intention's favour. Test-case willpower is unlikely to be a major player in diachronic control because it is vulnerable to judgment shift. The picture was therefore simplified – intentions provide the majority of diachronic stability; they include plans, policies and patterns of global choice. Intentions are supported by pre-commitments, muscle model willpower and, occasionally, test-case willpower.

In Chapter 3 we turned to norms of practical reason: diachronic consistency, means-ends coherence, and consistency between ends. We saw that if an agent follows these norms they will develop a network of plans and policies that create inter-temporal connections. That temporally extended network represents who the agent is; the plans and policies that create the most inter-temporal connections have the most agential authority. The agent self-governs when they act on

[33] Even if we admit that test-case willpower has some (relatively minimal) role to play in self-governance. Test-case willpower, at least as described by Ainslie, does not have any power to change the pre-discounted value of rewards. For Ainslie, the pre-discounted values of rewards set the parameters within which test-case willpower is deployed and judged for success or failure.

those plans and policies while pursuit of temptations involves a normative failure and lack of self-governance.

Self-governance has a diachronic and synchronic aspect. Diachronically, greater self-governance involves organising plans and policies so that they better accord with the norms of practical reason. Lack of ends-ends consistency and means-ends coherence create temptations. Furthermore, self-governing policies (and, I argue, plans) can adjust the strength that contrary desires tend to have. Finally if pre-commitments are to be used they also need to be set up in advance. Once temptation strikes, synchronic measures involve deploying muscle-model willpower to avoid reconsideration of existing intentions and, failing that, to avoid akrasia. Synchronically the agent exhibits greater self-governance the more muscle model willpower they use. The greater store of muscle model willpower they have the greater self-governance they can potentially display.

We're now ready to move on to Chapter 4 where I emphasise an important aspect of agency that Bratman neglects – the role of contingent aspects of self-concept. Much of what we target with plans and policies is a response to how we conceive of and integrate contingencies into our self-concept. Such considerations lead me to a narrative theory of agency.

Chapter 4: Narrative Agency

Introduction

I ended the last chapter supporting a variation on Bratmanian planning agency whereby agents follow norms of practical reason in order to develop and organise a hierarchy of intentions complemented by muscle model willpower and pre-commitments. In doing so they attempt to inhibit, extinguish, foster or create desires so that their desires tend to support their intentions. However, this account of agency has a blind-spot when it comes to the range of contingent factors that impact on agency. It encourages the view that agents are synonymous with their hierarchy of intentions while their contrary desires are something other, something extra-agential. But such a picture is at odds with how we typically experience our contrary desires. We count most contrary desires as our own even if we wish they were not; those desires can be as much a part of our self-concept as our intentions.[1] Furthermore, there are many other features we include in our self-concepts that are neither intentions nor desires. Self-concepts also include agents' interpretations of the unchangeable aspects of what they are (e.g. gender, race, certain bodily features), contingent constitutive features of their environment (e.g. where they were born and raised, who their parents are), the medley of accidents, windfalls, unexpected consequences that happen *to* them (e.g. surviving car accidents, meeting your future wife at the hospital), and future inevitabilities (e.g. puberty, menopause, death). I refer to these collectively as contingencies of self-concept.[2]

Like plans and policies, these other inputs to and interpretations of self-concept arguably create inter-temporal connections that help ensure the diachronic existence of self-governing agents. The planning account of agency therefore needs to be supplemented with this wider view of self-concept. In this chapter I argue that the evolving, inter-dependent amalgamation of an agent's plans, policies, and miscellany of self-interpretations is usefully thought of as a self-narrative.

[1] It is a much rarer situation to find our desires so alien that we feel they are not our own on any level. When such desires move us we are prone to feeling like bystanders to that behaviour. Frankfurt has described the behaviour of unwilling addicts as resigned bystanders to their addicted action (1978, p. 161). Such alien desires may feature in some failures of self-governance but, in general, addicted action, weakness of will and the struggles against them do not seem to involve an experience of such a strongly alien desire. In most cases of addiction the agent has previous embraced their addictive desire and so it has become far too familiar to be dismissed as alien.

[2] In what follows I will take an agent's self-concept to be his set of potentially articulable cognitive structures that feature himself, that's to say, his physical and mental self-description, plans, policies, desires, biography, and anticipated future. This immediately raises a complication regarding inarticulable elements of self. No doubt each person's self-conception also involves elements that just don't suit articulation (e.g. memories of spatial arrangements) and elements that others clearly see that should be articulated but that the agent does not (e.g. unconscious content). I assume these elements of self-concept exist but I won't be focussing on them.

Agents can self-narrate in ways that improve or undermine self-governance. My central claim, therefore, is that a complete account of agency should consider agents' self-narratives.

This is the last chapter in which I aim to develop a general theory of human agency; in this case, building self-narrative agency on to planning agency. By the end of this chapter, then, we will have considered three general theories of agency which each hope to explain addiction: a choice account informed by Heyman and Ainslie, a planning account informed by Bratman and Holton, and the narrative account developed here. The final two chapters are devoted to comparing how these theories perform when faced with a range of addiction phenomena. Chapter 5 compares the reward account with the planning account. Chapter 6 compares the planning account with the narrative account.

This chapter proceeds in two parts. In the first part, I argue that Bratmanian planning agency must be supplemented with a wider notion of self-concept. I begin with the observation that intentions, desires and contingent elements of self-concept all influence each other. Intentions need to be somewhat responsive to desires and contingencies and, therefore, desires and contingent elements of self-concept partially explain why the agent has the hierarchy of intentions that they do. This suggests that, not only can contingent features of self-concept be manipulated by plans and policies but plans and policies can be manipulated by the agent's interpretations of their contingent features. I analyse how self-concept affects self-governance. The result is a planning account of agency informed by self-concept. However, this account characterises self-concept as a relatively static and atomistic collection of intentions, desires, and contingencies.

In the second part of the chapter I improve on this view using the notion of narrative self-constitution. First I explain how a self-narrative connects one's intentions, desires and contingencies together so that they make sense in light of each other. The narrated elements are, therefore, not atomistic but holistically interconnected by the narrative weave; nor are they static because as the narrative changes so do they. Narrative self-constitution accounts claim that we control our lives by how we narrate them (within certain constraints). The final section of the chapter outlines the explanatory benefits of a narrative view over a Bratmanian view. Briefly, we can influence the power and character of our intentions, desires and other contingent features by how we narrate them. Furthermore, a narrative view better explains the detail and challenges involved in self-transformations, such as overcoming addiction. I develop these arguments further in Chapter 6. But to begin with, I make the case for the role of self-concept in agency.

Contingent elements of self-concept

In the previous chapter it was obvious that the agent's hierarchy of intentions form an important part of their self-concept because these represent the agent's active attempts to shape who they are becoming and how they are going about it.[3] There can be a temptation to believe that this network of intentions is all there is to an agent because everything else is beyond agential control, contingently helping or hindering the pursuit of those goals. However, the agent's network of plans and policies is not a uni-directional attempt to master contingencies but also a response to them.

In this first section I begin by illustrating the bi-directional interaction between intentions and desires before expanding on that to illustrate analogous interactions between intentions, desires and other contingent aspects of self-concept. This observation does not feature prominently in Bratman's account of planning agency but he can, nevertheless, easily accommodate it. A subsequent point cannot be so easily accommodated. Desires and other contingencies underdetermine their own interpretation and different self-interpretations create different self-concepts which affect the agent's ability to enact intentions. Self-governance is not, therefore, just a matter of developing networks of intentions while following norms of practical reason but also of developing and maintaining a self-concept that supports self-governance.

The interactions between intentions, desires and other contingent elements of self-concept

Desires influence which intentions the agent adopts and maintains, and which he can possibly adopt and maintain. The alcoholic might desperately wish he didn't have a strong desire to drink excessively yet he can't ignore the fact that the desire is a significant aspect of who he is. Indeed, being aware that the desire is *his* is a prerequisite for adopting self-control strategies such as adopting the policy, 'don't drink.'[4] On the other side of the coin, agents cannot adopt intentions that they consistently fail to garner the motivational support to enact. Agents needs to recognise such motivational shortcomings otherwise they are self-deceived (or lying) – they claim plans and

[3] Recall that I am setting aside the issue of coerced or inauthentic plans and policies.

[4] I don't deny the existence of dissociated desires where we cannot comprehend those desires as part of who we are. In schizophrenic cases, even a previously familiar desire might become alien. I assume there is a continuum from contrary desires that are clearly part of one's self-concept through to desires that are completely alien. My point is that our normative control of a contrary desire usually proceeds via considering it part of our self-concept. If a contrary desire is also dissociated from our self-concept this usually provides an additional challenge to control. Indeed some self-control techniques may proceed via dissociating oneself from the desire but to deliberately take such a step requires that one identify the desire as one's own in the first place (I consider these techniques in Chapter 4).

policies that they cannot or do not follow. If the alcoholic cannot follow the policy 'don't drink' because his desire to drink is too strong, then he needs to devise a more circuitous route to controlling that desire such as making a commitment to a treatment plan. So, in many cases, our intentions wait on our desires, some intentions only make sense given the need to control certain desires and all intentions depend on some contingent motivational basis.[5]

Yet desires are only one of the contingent aspects of self-concept. As mentioned above, there are a significant range of other contingent phenomena that feature in self-concept, such as, gender, body, ancestry, "found" communities,[6] nationality, accidents, windfalls, unexpected consequences, inevitabilities, et cetera. These contingencies, like desires, evade perfect normative control. Some things happen to you before you have developed the normative skills to control them (e.g. early childhood inter-personal interactions). Other things just aren't susceptible to normative influence (e.g. genetic conditions, who your parents are, puberty, death) or are extremely resistant (e.g. gender, height). Yet other things are normatively controllable in principle but typical limitations of cognitive and epistemic powers result in unforeseeable events (e.g. car accidents, meeting pleasant or unpleasant people).

Contingencies of self-concept influence both one's desires and intentions. Consider the following desire-contingency relations: The abused child may be averse to developing close relationships even though eventually developing such relationships are a necessary part of healing. A car accident resulting in paralysis results in a desire to walk again. Some people that you happen to meet you wish would stay, others you wish would leave. Consider the following intention-contingency relations: A car accident resulting in paralysis forces one to abandon walking as a means to one's ends and to abandon the plan to walk again. Our parents significantly shape our values when we are young and impressionable. Normatively developing the skills required to enjoy activities, such as hiking, kayaking, skiing, sailing, depends on having access to certain physical environments. Some cultures put certain plans, policies and self-descriptions out of reach of people who happen to be women, homosexuals, or of certain ethnicities. Contingencies, therefore, make

[5] Although intention formation often precedes the motivation to enact the intention. As described in the last chapter, new desires can be built through adopting a policy to try and foster those desires. Furthermore, general sources of motivation can be channelled more specifically with intentions. For example, a desire to look attractive could be co-opted to motivate exercise, body building, tanning, or plastic surgery et cetera.

[6] Friedman's (1992) term to distinguish the communities we find ourselves in from "communities of choice," those communities we voluntarily associate with as we develop our agency and self-concept. The distinction is not hard and fast, some communities are more "found" than others and some chosen communities are necessarily nested in "found" communities.

some intentions and desires impossible, other intentions and desires difficult to develop while leaving others freely available, and yet other intentions and desires almost inevitable or hard to shake.

Although contingencies partially escape normative control by definition, we do have some normative control over them. The agent's physical self-description is influenced by plans and policies to control his styles of dress, posture, coordination, exercise, diet, et cetera. As he grows up he can question and replace plans and policies that were promoted by his parents. He can change physical, cultural and social location to suit his changing network of intentions or to deliberately challenge some of his existing intentions. He can pursue or resist gender, race or ethnicity expectations. Normative failures and unexpected consequences can be ignored – 'I won't let that affect me' – learned from – 'I won't make that mistake again' – or even taken as starting points for completely different normative enterprises – 'That was the best mistake I ever made!' So, as in the above case with desires, understanding his contingencies as part of his self-concept is important for their normative control. Understanding those contingencies clues the agent in to which plans and policies will be helpful, enjoyable, difficult, necessary or possible for an agent like him in his situation. For example, 'I shouldn't bother trying to be a high-jumper with this physique,' 'being a middle-class, White male I have no excuse for failure,' 'They will never let me become a doctor but I'm going to try anyway,' et cetera.

So the agent's intentions, desires and other contingencies are irreducible to each other and they influence each other to some extent. This means that it is possible for self-governance to be strengthened or undermined by variations in how the agent interprets those contingencies in their self-concept. As an initial case take the paralytic car crash victim. For him, some plans, such as walking again, are put absolutely off limits by the accident; no matter what self-concept he develops he cannot walk. However, there are a range of plans and policies that are not necessarily off limits, e.g. driving again, playing with one's children, going to outdoor concerts. Different self-concepts might make these plans and policies seem more or less attainable. The self-concept, 'useless cripple,' might limit self-governance over and above the relevant objective limitations because useless cripples can't do much. This self-concept is fatalistic. Alternately the self-concept of 'medical miracle' might promote self-governance over and above the objective evidence as the

agent implements plans and policies that he (and others in his position) would not otherwise try.[7] Because agents have some control over the development of their self-concepts they can take advantage of these agential effects. I'll first consider how self-concept can inordinately undermine self-governance through fatalism and then consider how that effect can be mitigated and how self-concept can be manipulated to boost agency.

Detrimental self-concepts

The effect of a detrimental self-concept on agency is perhaps most familiar in cases of depression. A person suffering from depression might come to see himself as worthless despite the fact that there is no objective reason for that self-concept.[8] Friends and family profess to value the person and point out objective factors supporting his value to no avail. Despite the fact that a self-concept such as, 'worthless person,' is ungrounded, it informs his (lack of) action and begins to genuinely erode his intentional network. Friends despair at his negative attitudes and begin to avoid him, he loses his job through excessive absences, and so on.[9] The overly negative self-concept tends to shape action in line with that self-concept; it is self-fulfilling. The physiological basis of depression may unavoidably cause some negative self-reappraisal of self-concept, however I assume that some control of self-concept remains and letting one's self-concept become more negative aggravates and entrenches depression.

Fatalistic self-concepts are not only concomitant with mental disorders, in fact empirical work on what is called 'stereotype threat' suggests that we all have fatalistic aspects to our self-concepts that are unnecessarily detrimental to our agency. A stereotype threat is where a person's performance in a task is decreased by making salient the connection between their self-concept and a stereotype that would be expected to perform poorly in the task.

. A classic case of stereotype threat involves gender stereotypes and performance in a task where one attempts to correctly match a 3D shape with rotated versions of itself rather than similar 3D

[7] Self-concepts can also undermine self-governance through overconfidence. For example, if the car crash victim believed that he was super-human and so could walk again despite a severed spinal cord.

[8] Depressed people show a preference for negative feedback (Giesler et al., 1996) despite not liking that feedback (Swann et al., 1992).

[9] Depressed people's *style* of interaction also compounds the problem, e.g. excessive self-disclosure (Gibbons, 1987), hostile speech content (Gotlib & Robinson, 1982), lack of responsiveness (Bouhuys & van der Meulen, 1984), reduced eye contact (Dow & Craighead, 1987), negative facial displays (Schwartz et al., 1976).

shapes. Female performance in the 3D rotation task becomes worse than baseline when, just prior to attempting the task, the required skill is associated with traditionally masculine domains such as aviation engineering, nuclear propulsion engineering, navigation, et cetera. Conversely, if the task description associates the required skill with traditionally feminine domains, such as clothing design, interior decoration, flower arrangement, et cetera, men's performance is much decreased and women's improved (Sharps et al., 1994).[10] Furthermore, the more strongly the person identifies with the aspect of self-concept linked to the stereotype the stronger the stereotype threat effect. In a study on people from the Southern US, those who identified more strongly with that aspect of their self-concept had a greater decrease in cognitive performance when exposed to the stereotype that people from such places were of low intelligence (Aronson et al., 1999; Clark et al., 2011). Now, stereotype threats don't just undermine task performance, they undermine performance in *valued* tasks (i.e. intentions that support a relatively high number of cross-temporal connections). Experiments show that stereotype threat becomes stronger the more the agent values the task. Aronson et al. (1999, Experiment 2) exposed White, male, university students enrolled in the second semester of a rigorous year-long calculus course to stereotype threat (Asians are better at maths than Whites). They all suffered performance decreases, but the students who indicated they valued mathematical ability the most suffered a greater decrease in their performance than those who valued it less.[11] Finally, chronic exposure to stereotype threat has been shown to result in permanent dis-identification with whole skill areas, e.g. stereotyped African Americans disengaging from intellectual pursuits (Major et al., 1997; Osbourne, 1995). No doubt people with certain characteristics face more stereotype threat than others but any characteristic is vulnerable to some kind stereotype threat and, therefore, all people are vulnerable to these fatalistic attitudes.

In any case, it seems likely that we are vulnerable to taking negative generalisations of all types overly to heart, not just stereotypes. Vohs and Schooler (2008), for example, showed that we underestimate our self-efficacy if we read a text that portrays behaviour as solely the consequence

[10] For an interesting variation see Moé (2009) where the better performance by men in baseline, unprimed conditions was eliminated by priming participants with the statement, 'women perform better than men in this test, probably for genetic reasons.' This suggests that even in unprimed conditions there is an implicit stereotype threat about performance in such tasks.

[11] What about stereotype effects that improve performance? Are those effects non-normative? There is less work done on these effects but I assume they are typically non-normative. I expect that the men who were primed with a connection between traditionally masculine domains and 3D rotation would have been surprised that their performance was better than unprimed men. Improved performance was an unexpected windfall that they wouldn't have known about until the study results were made available. Subsequently, however, they might take advantage of the effect normatively by working to develop an environment that was rich in performance enhancing primes. I consider this approach below under 'agential response'.

of environmental and genetic factors. Participants who read the text had a decreased belief in free will and increased rate of cheating compared to controls who didn't read the text.[12] Analogously we can imagine that, say, English speakers trying to learn Chinese would see their efforts undermined when told that native English speakers find it easier to learn French than Chinese, or the aerobic fitness goals of smokers would be made more difficult if they are told that smokers have less aerobic capacity than non-smokers. Even people who rarely face stereotypes are routinely subject to generalisations that link traits in their self-concept to negative outcomes. So although stereotype threat is particularly unpleasant because it often trades on falsities, it appears that the truth of the generalisation is irrelevant to its effect, all that matters is whether the agent believes the generalisation applies to them. Therefore, most, if not all, of the aspects of our self-concepts leave us vulnerable to certain stereotype threats and other negative generalisations in some circumstances. Nelson suggests that aspects of our self-concept can have these effects because we sometimes take contingent aspects of our self-concept to provide an explanation for our actions.

> "Sometimes we don't have reasons. When we don't, a causal explanation is enough, it seems, to show us how a commitment fits in with our sense of who we are" (2001, p. 80).[13]

The stereotype threat evidence suggests that something even stronger is true. It seems that sometimes we act in accordance with what we take to be contingently true of ourselves even when it *clashes with* what we judge we have most reason to do, i.e. our valued plans and policies.

Controlling detrimental self-concepts

Detrimental self-concepts would not be particularly relevant to a philosophical account of agency if there was nothing we could do about it. However, people, whether they be depressed, negatively stereotyped or have just become wrongly convinced of their inability in a certain task, may have some success in fighting off these overly negative self-concepts.

[12] Cheating took the form of passively allowing the answers to be shown by a faulty computer program in one experiment or overpaying oneself for a cognitive task.

[13] Some research supports Nelson's view (Bandura, 1997; Cadinu et al., 2003). Other research indicates that, even if the subject takes the stereotype threat to be false, their performance is still undermined by their emotional response (Blascovich et al., 2001; Spencer et al., 1999, Experiment 3; Steele & Aronson, 1995). Although it is difficult to know whether some of that effect still stems from an implicit identification with the stereotype. Given the necessary role others play in building our self-concepts it may be difficult to completely dismiss the effects of their views even when we are sure they are false.

First, we can note that our plans and policies have some resilience. Just as a contrary desire will not necessarily undermine an intention, neither will a detrimental self-concept. In favourable conditions intentions may be followed despite even highly pessimistic self-concepts. In unfavourable conditions even small amounts of self-doubt might ruin a project. Therefore, we need not avoid self-concept-driven fatalism completely, but we need to minimise it so that our networks of intentions are sufficiently protected. How does one minimise such fatalism?

The simplest technique would be to avoid situations where demotivating generalisations are propagated. Unfortunately this is not always available because the environments necessary for the agent's normative pursuits expose her to stereotype threat. But there is evidence that certain cognitive responses to stereotype threat can protect from performance decrease. For example, the following prime preserved task performance in a maths test: "it's important to keep in mind that if you are feeling anxious while taking this test, this anxiety could be the result of these negative stereotypes that are widely known in society and have nothing to do with your actual ability to do well on the test" (Johns et al., 2005). One could also try to improve one's judgment of how generalisations should be applied to oneself, carefully distinguishing the truth of the generalisation in one's own case without adopting anything more.

These techniques provide some defence but what is more effective, where possible, is to change the *content* of one's self-concept. We aren't just at the mercy of others in developing self-concept, we can promote the kind of self-concept we want to have.[14] To do this we must remain within the boundaries of what we can believe of ourselves, what our social group will accept (on pain of being dismissed as self-deceived[15]), and what is physically possible.[16] As a result, changing contingent aspects such as gender, race, place of birth, et cetera, is usually impossible. However, within those boundaries, there are often several possible reinterpretations of oneself. For example, reinterpreting oneself as a sexual abuse survivor rather than a victim could help disconnect oneself from generalisations and stereotypes applied to victims. If one mainly grew up in New York but was

[14] Indeed there is an entire self-help industry based on this. My arguments suggest that there is some basis for believing such techniques of self-help will work although within limits.

[15] There is the possibility of changing the social boundaries themselves although this usually involves a lot of hard work, social conflict, and social isolation.

[16] The change to self-concept also has to sufficiently cohere with what the agent values. For example, the agent might believe they could have a sex-change, it might be accepted in the community, and medical science makes it possible but if they value their current gender highly they will not want to change it. I defer this issue until later in the chapter because I haven't yet developed the conceptual tools to explain how contingent aspects of self-concept can be valued. On the Bratmanian view only plans and policies are valued; a narrative account can accommodate the variation in value attached to contingencies.

born in small town Kansas one might interpret oneself as from New York rather than Kansas to avoid generalisations about people from small town, southern USA. This technique is not limited to being a defensive strategy against detrimental effects but can be used positively to try and boost intentions that are not under threat – I consider this below.

Beneficial effects of self-concept: 'fake it till you make it'

As with countering the effects of self-concept that undermine self-governance, agents can use self-concept to improve their self-governance in two ways. First, agents can seek out correlations linking their existing self-concept to success and thus capitalise on the inverse effect of stereotype threat. A person's performance in a task is improved by making salient the connection between self-concept and a stereotype that performs well in the task. For example, telling a man before a maths task that men perform better than women, or women worse than men, at maths improves male performance (Moé, 2009).[17] Of course these benefits come at the cost of reinforcing the stereotype threat to out-groups and they cannot help you if you happen to be on the wrong side of the stereotype, e.g. a heterosexual man attempting home decorating. If we assume that positive, non-stereotypical generalisations would also help, then one might be able to get around these issues by trying to reinforce more personalised positive correlations. However, given that the correlations have to be believed by the agent to have their effect, they might be difficult to engineer. Just paying someone to follow me around telling me that people named 'Doug' are always successful at whatever they happen to be doing is unlikely to help.

Rather than trade on existing or engineered correlations with your current self-concept, the second approach is to change your self-concept into one that has the positive correlations you want. Given the mutual interplay between intentions, desires and other contingent elements of self-concept one should be able to drive or support a plan, policy or desires through first changing self-concept. Obviously such 'fake it till you make it' strategies have the same limits as when the agent tries to change a detrimental self-concept. As the ambition of the self-concept grows so too does the risk that the new self-concept becomes unbelievable to the agent, unacceptable to others, or so ambitious that it is physically impossible. But within those limits there is some room to move. A

[17] This is called stereotype lift or stereotype susceptibility. Stereotype lift is where the negative stereotype is connected with an out-group, e.g. men do better when told that women do worse. Stereotype susceptibility is where one connects directly with the positive stereotype, e.g. men doing better when told men do better. Stereotype susceptibility has a greater effect than stereotype lift.

recent study of people trying to quit smoking appears to have caught this process in action (West, 2006). We can assume that all study participants were attempting to follow a policy not to smoke but when asked one week after quitting some identified themselves as 'ex-smokers' and some as 'smokers.'. Those that had the self-concept of 'ex-smoker' were much more likely to be abstinent in the long-term. The study showed a high (75%) relapse rate overall, but 50% of those participants who identified as 'ex-smokers' rather than 'smokers' a week after quitting were still abstinent six months later compared with *none* of those who identified as 'smokers.' This means that, of the successful quitters, all had adopted the self-concept, 'ex-smoker,' (presumably changing from the prior self-concept, 'smoker'). This is just an association but the evidence we have already seen for a connection between self-concept and intentions favours the following interpretation: The self-concept 'ex-smoker' helps people quit because ex-smokers in general have a lesser desire to smoke, do not form plans to obtain and smoke cigarettes, and find it easier to follow the policy, 'don't smoke' than smokers.[18] Presumably the successful abstainers in the smoking study did not deliberately adopt the self-concept in order to support their policy – that was a happy side effect of an implicit change in self-concept. However, there is nothing to stop people trying to deliberately change their self-concepts in order to benefit from the support to their normative agency. In fact Health Canada's Guide to Quitting encourages such a change with the advice: "Remember you are a non-smoker. You do not smoke. Make this your first and last conscious thought each day."

We can also speculate that the self-concept 'ex-smoker' promoted reinterpretation of the desire to smoke as something else (e.g. general irritability from tiredness, thirst or hunger), after all a non-smoker wouldn't feel like a cigarette. That speculation is encouraged by 'misattribution' experiments where, for example, pathologically shy people have learned to attribute their anxiety to something other than social situations and subsequently improved their social interaction (Brodt & Zimbardo, 1981). Of course, in the event that one stops smoking for good, or overcomes shyness, then the need to misattribute disappears.

Summary

Intentions, desires and other contingent elements of self-concept all interact with each other. Here I have focussed on how our self-interpretations of the contingent aspects of our lives influence our

[18] Velleman cites a case where such a change in self-concept seems to have worked (2002, pp. 99-100). I return to that case below.

self-governance. These interpretations influence what intentions and desires make sense to us but sometimes that influence generates an unfounded fatalism that undermines our self-governance. Stereotype threat and depression are clear causes of self-concept influenced fatalism but, if the above analysis is correct, then any creature with a self-concept and fallible judgment will be vulnerable to negative associations linked to their self-concept. Where possible, we can avoid such effects by avoiding damaging social environments and developing a power to judge to what extent generalisations apply to our own case. If fatalistic self-concepts develop, the agent needs to work to change that fatalistic content and we would expect this work to require considerable social scaffolding. On the other side of the coin, agents can seize the initiative by adopting the 'fake it till you make it' strategy whereby positive self-concepts create optimism that supports norms and counters contrary desires.

In the last chapter we saw that agents will achieve greater self-governance the better they follow the norms of practical reason, namely, sufficient but not excessive ends-ends consistency, means-ends coherence, and diachronic stability. It now seems that we need another practical norm: sufficiently optimistic but not overconfident self-concept. Self-governance will tend to be undermined if the agent is too pessimistic or too optimistic in the way they build their self-concept.

At this stage one might argue that this is a fairly minimal change to Bratmanian agency. Aspects of self-concept may be partially independent factors in agency but they are subordinate to intentions. From what has been said here, the agent tries to manipulate his self-concept so that it best serves his network of intentions. Ultimately the agent values their plans and policies not these other contingent elements of self-concept. But, interestingly, this does not seem to be the case. People appear to value certain aspects of their self-concept even when they undermine certain intentions and could be changed without breaking social boundaries. As we will see, a narrative account can explain why this is so. Contingent elements of self-concept, when incorporated in a narrative form, provide inter-temporal connections that help ensure the agent's temporally extended existence. Recall that this was the very basis on which Bratman claimed that intentions represent what we value. This also can lead us to an explanation of why changing self-concept can be more challenging than merely collecting or abandoning socially sanctioned descriptors. I now turn my attention to describing a narrative account of agency.

Narrative self-constitution and agency

In this section I present an account of narrative self-constitution. I begin by outlining the following key characteristics of narratives: They make sense of the temporal development of events (real or fictional); they are interpretations made by narrators and, as such, always reflect the perspective of the narrator; they specify a particular focus and context for cognition from a range of available possibilities, and; they have cogency, that is, they elicit affective responses relevant to the developments they describe. The scene is then set to explain how the agent constitutes (and reconstitutes) herself through an ongoing process of narrative self-interpretation and self-projection. She thereby makes her life intelligible from her own (socially embedded) perspective, specifies foci and contexts for her experience, and both creates and is effected by the cogency of her narrative.

What is a narrative?

Narratives represent, usually in words, the relationship between two or more events. They thus always capture an order of events across some temporal duration. But narratives differ from a mere chronological list of events because they are always from the perspective of a narrator and they always specify the causal relationships between events so that these are intelligible (Velleman, 2003). Train timetables, for example, fail to count as narratives for two reasons – they chronologically order events but don't explain their relationships and they are objective rather than a narrator's interpretation. Conversely, 'the man ran to catch the train because he was late for work,' does have an intelligible narrative form. Not only do we have the event of him running and the future event of him catching or missing the train, we are told *why* he is running which provides an intelligible connection between those events. The narrative also entails a perspective from which the scene is interpreted, i.e. somebody observing the man running.

Successful narration makes a scene intelligible by fitting it into a *narrative form* where meaningful connections between events are made apparent; narratives have to meet a normative standard of intelligibility. This norm is difficult to accurately define but we usually know pre-reflectively if it has or has not been met. If I say, "that man missed the train because his alarm didn't go off," then we understand. If I say, "the man missed the train because of a talking banana," then we don't understand. Why not? In the former case we have the necessary background knowledge – people

need to be on time for trains, they use alarms to wake up, the stereotypical narrative of being late for work is familiar in our culture. In the latter case we lack the necessary background information; there is no stereotypical narrative of how talking bananas make people late. We can imagine how it *might* happen but no single convincing possibility stands out from the others.[19]

However, the latter narrative fragment may be made intelligible if we can get the narrator to expand on the narrative to the point that we can connect it with our existing knowledge.

> "In successfully identifying and understanding what someone else is doing we always move towards placing a particular episode in the context of a set of narrative histories, histories both of the individuals concerned and of the setting in which they act and suffer"(MacIntyre, 1984, p. 211).

For example, if we are told that hallucinogens were secretly put in the man's breakfast, then we can understand that he hallucinated the banana and lacked the mental competence to get to the train. We already knew that getting to the train on time requires a certain mental competence, ingestion of hallucinogens might cause hallucinations of talking bananas, and hallucinating puts the required competence in danger. We needed that additional narrative connection to link that wider knowledge to the case at hand; we can then nest the narrative fragment within a wider narrative context.

The goal of those two narrative fragments was to explain causal connections, the cause of the man missing the train. But narratives may also aim to provide teleological explanations, e.g. to what end the man was taking the train, or bring out themes, e.g. the absurdity of workers' commitments to monotonous dead-end jobs in general. Whether narrators aim to reveal causes, teleology, or themes (or all of these) they do so by connecting the events of interest with a wider set of general narratives (either their own or those they assume of their listener). Those more general narratives ultimately form explanatory bedrock capturing the causal relationships, goals and themes that are considered intelligible. Being able to understand others and make oneself understood therefore depends on a shared background of general narrative structures (or the possibility of developing one).

[19] For someone from a hunter-gatherer culture the former narrative fragment may be as unintelligible as the latter (setting aside the fact it is in English) because they have no stereotypical narratives about the role of trains and alarms in daily life or even, perhaps, of being late.

The view of narrative I have developed here entails that narratives can be messy and banal, highly specified or vague, brief or epic.[20] Unlike literary narratives, they do not have to build up drama, tension and twists culminating in a conclusion that provides an aesthetic satisfaction. Nor do they need to be meticulously edited so that everything relevant and nothing extraneous is included. A narrator has successfully narrated a series of events when she creates meaningful connections between them that render causes, ends, and/or themes intelligible.

Effects of narrative: focus, context, cogency

Narrative has three major effects besides creating intelligible meaning – it creates a specific focus, a specific context, and can elicit emotion (i.e. has cogency). Any evolving state of affairs can by divided into many different events standing in many different relations to each other. A narrative picks out one subset of those events and relations, so narration is a selective process where some events and connections are emphasised and others ignored. The selective process depends on the perspective and purpose of the narrator (many possible narratives are possible from any perspective at any time). The station master might observe the man missing the train with a sense of satisfaction because the train has left exactly on time as a result of his hard work. But the engineer, observing the exact same scene, might hear that the engine needs his attention in the near future and realises he will have to work late tonight after the engine returns. The sound of the engine and its implications do not feature in the station master's narrative, and neither do the current time and the late man feature in the engineer's narrative; each narrative creates a different *focus*. Many narratives can be equally true of the same state of affairs and the 'best' narrative depends on how the narrator wants to focus their own attention (or the attention of others).

Each narrative creates a specific *context* for the elements that make it up; the context influences the significance of the elements. This is most obvious when the same element features in multiple narratives. For example in Roald Dahl's short story 'Lamb to the Slaughter' (1986) the police arrive at a woman's house to investigate the murder of her husband. Within that wider narrative context, another narrative context unfolds – the police enjoy a leg of lamb for dinner at the woman's house. The wife and the reader have the wider narrative context and so they know that that very same leg of lamb was the wife's murder weapon. The police only have the narrower narrative context so, for

[20] This definition of narrative fits with what Schechtman labels a 'moderate' view, i.e. more explanatory links than in a mere sequence of events and less coherence (of theme or plot) than a literary narrative (2007, pp. 159-160).

them, eating that leg of lamb is an innocent act of enjoying hospitality and reassuring the 'shocked' wife. In the wider narrative context enjoyed by the wife and the reader, the dinner involves the ironic destruction of the murder weapon by the very people who are searching for it. Each narrative context changes the elements and meaningful connections available. Humans become very good at simultaneously appreciating a variety of narrative contexts each tied to different perspectives.[21]

Narratives also elicit emotions; they have cogency. When engaging with a narrative, such as when watching a movie or reading a novel, the narrative draws us into imagining we are there. Imagining "…simulates to some degree the effect of an event we might have lived through…" (Wollheim, 1984, p. 81) including emotions elicited by the event (given its position in the narrative form unfolding). For example, when Frodo and Sam are climbing Mount Doom to try and destroy the ring in "The Lord of the Rings" (Tolkien, 1966) we feel something of their exhaustion and their fear of the evil Sauron seeing them. We also anticipate the potential fulfilment of success or disappointment of failure made highly significant by the current event's position in the epic narrative. We can imagine being there in an acentral position (i.e. a god's eye perspective) or we can imagine taking the perspective of one of the characters (centrally). Both perspectives have cogency. When imagining from a character's perspective we are almost always sympathetic rather than empathetic, that is, we emotionally *react to* their situation rather than experiencing that situation as if we *were* that character (Goldie, 2005, 2007; Wollheim, 1984, pp. 67-79). Empathetic imagining is more challenging because one has to momentarily ignore oneself and take on the self-concept of the person you are trying to be in order to imagine how *they* would react. Given the holism of our mental lives our ability to do this is limited (Goldie, 2000, 2007; Mackenzie, 2008). Nevertheless we can roughly imagine what it is like for Frodo to have to finally destroyed the ring given the deep, supernatural attachment he has developed for it. We can also imagine Sam's frustration; after all the effort to get to Mount Doom Frodo will not complete the final, most simple task of dropping the ring in the volcano.[22]

In summary, then, narratives make events intelligible by highlighting certain causal chains, goal-directed activity and/or themes by drawing meaningful connections between them and to an

[21] To appreciate dramatic irony we need to simultaneously consider multiple narrative contexts (Goldie, 2012). If we adopt only one narrative context at a time, e.g. take the perspective of just the wife or just the police, we cannot fully enjoy the scene. Simultaneous consideration of multiple narrative contexts, therefore, requires the agent to take an acentral, 'god's eye' perspective. Such an external perspective also happens to be important when we try to achieve objectivity in our self-narrations (Mackenzie, 2008). I pursue this point below.

[22] The possibility of taking acentral and central perspectives on one's own narrative is an important aspect to narrative self-constitution that I develop below.

existing narrative background. In doing so narrative specifies a focus and a context; it also tends to elicit an emotional response relevant to its content. Human agents grow to appreciate the individual narrative context of specific characters' perspectives and appreciate multiple narrative perspectives simultaneously from an acentral position. So what does all this have to do with self-constitution?

Narrative self-constitution

Narrative self-constitution views claim that we constitute (and reconstitute) ourselves through an ongoing process of self-narration. That is, we tell stories that make sense of and shape where we have come from and where we are going, who we have been and who we are in the process of becoming. Narrative self-constitution is compatible with there being non-narrative aspects to the self-conceptions of self-narrators, e.g. representations of one's body image, size and shape. It is also compatible with many non-narrators having a self-conception without narrative, e.g. young children, some animals, and those with severe dementia.

> "Autobiography … isn't life. It's a narrative structure that makes sense of life" (Nelson, 2001, p. 62).

However, if we cannot make sense of our life or of our potential futures then we cannot shape it and the quality of our lives will be severely reduced. This becomes obvious when we consider the connection between self-narration and agency in more detail.

Self-narratives involve both self-interpretation and self-projection, that is, self-narratives are not just descriptive but also *prescriptive*. The way the narrator's life unfolds depends on what they narrate. Take the following self-narrative: I grew up in a poor family in Kansas. The injustice I saw growing up led me to want to become a lawyer. Because we had no money I had to work hard enough at school to get a scholarship to attend law school. I succeeded and am now halfway through that training. When I finish I plan to return to Kansas to defend poor victims of injustice.[23]

[23] To simplify the picture at this stage I will focus on one narrative thread per agent. However agents do not live a single perfectly coherent narrative but create and weave together multiple, semi-independent narrative threads as required. We might have a series of career narratives, a series of intimate relationship narratives, a narrative of childhood, of fatherhood, and so on. The narrative self that results is never a single perfectly unified entity but rather a messy cluster of different narrative threads built to understand and guide different aspects of life. The self combines first personal embodied experience at a time and over time and third personal descriptions made by ourselves or others of our biography, body, relationships and social roles, etc. "The role of narrative self-conception … is to try to reconcile these different and sometimes conflicting dimensions of selfhood and to integrate the motives, values, and emotions

Here we have the narrator's description of who they take themselves to be and their projection of who they are in the process of becoming. Yet, I claim, that if they self-narrated differently then they would have a different life. To see this, consider three ways different self-narration can change a person's life.

First, in narrative projection the agent narrates an imagined future based on her current narrative self-interpretation. Most self-interpretations underdetermine the possible narrative continuations from that point and so it is up to the agent to imagine the narrative continuation she prefers and then enact it. The projection she does prefer is influenced by the cogency of the narrative possibilities she imagines. For example, the self-interpretation of being in law school could form the basis for a self-projection where one is made a partner in a law firm in Kansas, or works at the International Court of Justice in The Hague, or drops out to join the army, and so on. As the agent imagines each of these possibilities it becomes clear to her that the camaraderie and physicality of being in the army, though appealing, wouldn't offset the feelings of seeing justice done in court, and local justice would be more emotionally satisfying than international justice. The projection that the agent chooses sets them on the path to realising that outcome. This line of thought ultimately leads to the view that when we successfully enact our narratives, "we invent ourselves … but we really are the characters we invent" (Velleman, 2005, p. 206).[24] Of course successful narrative projection is not guaranteed, but it is significantly more likely to realise a valued future than if the agent didn't narrate a particular path to that future but just hoped that such a future would come about anyway.

Second, any state of affairs underdetermines how it should be narratively interpreted, so even when narration is purely post hoc and descriptive, the agent can choose which narrative interpretation of many best suits their experience. The law student could take herself to be enjoying the privilege of higher education or to be the victim of merciless lecturers who pile on the work. Later as a lawyer she might see herself as a win-at-all-costs crusader for justice or a political animal who knows how to get the best for herself by playing the rules of the system. Self-interpretation not only changes who one is at present but recursively puts limits on the narrative projections that make sense from

arising from them, into a relatively stable practical stance" (Mackenzie, 2008, p. 128). I return to this complexity in Chapter 6.

[24] Narrative self-projections are therefore very much like the intentions of planning agency considered in Chapters 2 and 3. The difference is that narratives explicitly include past and anticipated contingencies. In fact some narrative projections may be mainly reactions to (or anticipations of) what has happened (or will happen) *to* the agent. I point out the significance of this below.

that present. The student who saw herself as a victim, for example, might be more likely to develop the political animal narrative because that is how she can best look after herself in a social system she sees as oppressive.

Third, the interpretation and projection narrated creates a focus, context and cogency for one's life. As we control our narrative we control what we pay attention to and the context for our experience including which emotional responses are appropriate. The law student seeing herself on a mission to right injustice, is alert to instances of injustice, she seeks them out. However, if she becomes the political animal she may become relatively blind to injustice but more alert to chances for self-promotion. Within the narrative context of the crusader for justice, accusations of being unjust have an especially strong sting. If she develops the narrative context of the political animal, in contrast, such accusations can be a badge of pride, a symbol of power. So, if you want to change who you are, who you are becoming and the focus, context, and emotional character of your life, you can change your self- narrative. There are, of course, limits on which narratives can be self-constituting and I turn to them now.

Limitations on narrative self-constitution

There are limitations on which self-narratives are genuinely constitutive. Although I can come up with a story about how I am Napoleon or how I am going to become the president of the United States, no matter how I fill in the detail of those narratives they will not be self-constituting. Why not? To be self-constituting a narrative has to be true of me; it has to fit with reality. The constraints one places on narrative self-constitution, therefore, depend on how one thinks reality is constituted. Here I set out three sources of constraint, social, subjective, and objective. I think each has some independent force but others may believe one or more reduces to the others.[25] I will not try to settle the question of what fundamentally constitutes reality here. None of these constraints set an

[25] Although even if they are independent they overlap considerably, for example, other people will not endorse narratives they consider to be objectively impossible and neither will agents adopt such narratives.

absolute limit on what is self-constituting, rather self-narratives that better meet these standards are more self-constitutive than those that do not.

One source of constraint is social – individuals, institutions, and cultural expectations leave some narratives open while making others more difficult or impossible.[26] At a cultural level, some narratives are made available, others denied and some are even made compulsory.

> "We enter society … with one or more imputed characters – roles into which we have been drafted – and we have to learn what they are in order to be able to understand how others respond to us and how our responses to them are apt to be construed" (Nelson, 2001, p. 56).

For example, women are expected to choose from a different range of narrative possibilities than men. In the 18th Century it was basically impossible for a woman in Britain to develop the narrative projection of becoming a doctor; nobody would teach her or employ her.

The archetypal narratives that our society expects us to conform to do not just open up some possibilities while shutting down others, we are heavily reliant on them to develop *any* self-understanding. As MacIntyre observes:

> "It is through hearing stories about wicked stepmothers, lost children, good but misguided kings, wolves that suckle twin boys, youngest sons who receive no inheritance but must make their own way in the world and eldest sons who waste their inheritance on riotous living and go into exile to live with swine, that children learn or mislearn both what a child and what a parent is, what the cast of characters may be in the drama into which they have been born and what the ways of the world are. Deprive children of stories and you leave then unscripted, anxious stutterers in their action as in their words. Hence there is no way to give us an understanding of any society, including our own, except through the stock of stories which constitute its dramatic resources" (MacIntyre, 1984, p. 216).[27]

[26] Françoise Baylis (2011, 2012) argues that this is, in fact, the only constraint on narrative self-constitution. Schechtman (1996, p. 119ff) refers to a 'reality constraint' on narrative self-constitution whereby the narrator must largely agree with their social milieu on what counts as real.

[27] Jones makes a similar point, "…in the stories we tell each other about what it is like to have an emotion of a particular kind, stories shape our understanding of what is to count as (romantic) love, what lovers do, what they feel, and who may be properly loved by whom" (Jones, 2008, p. 270).

If one's narrative departs too dramatically from socially endorsed stock plots or archetypes then at best one is frequently misunderstood and, at worst, socially excluded, subject to violence, and left without even being able to understand oneself.

Our narratives do not just answer to these background cultural conditions but are also intimately co-authored by particular people.

> The content of one's self-narrative "…comprises those features of our lives and ourselves that we care about, there is also an extent to which our identities are constituted by the content of *other* people's narratives – the features of our lives and ourselves that *they* care most about. … Who we can be is often a matter of who others take us to be. Many practical identities require more than one person for their construction and maintenance" (Nelson, 2001, pp. 81-82, note 3).

For example, you can only live the narrative of a husband if someone will marry you. Then the self-narrative that you develop to understand what kind of husband you are is strongly influenced by what your partner co-authors.[28] So many self-interpretations and self-projections are only self-constitutive to the extent they are accepted by intimates, authorities, and society in general.[29]

In addition to these social constraints there are also subjective constraints. Primarily the agent must be capable of self-narration; Schechtman refers to this as the 'articulation constraint' (1996, p. 114ff). People might create a narrative for the life of an animal at the zoo but because the animal has not contributed to that narrative at all it is not a self-constituting narrative. The agent must have cognitive access to her narrative. Of course there are often true stories told about agents that the agent themselves are unaware of. Oedipus, for example, is a parricide, father to his sisters, husband to his mother even though he cannot articulate any of these things. These stories contribute to a *social* identity but, prior to the agent becoming aware of them, those stories do not feature in his

[28] Such inter-subjective effects are particularly relevant to addiction when we consider the effects of stigma and I return to this theme in the Chapter 6.

[29] This picture gets muddied when there is disagreement over which narratives are self-constitutive. Even if the self-narrator has no social support they are not necessarily deluded and, frequently enough, various social groups disagree on which narrative an individual is really living. For example, in the 1980's, some people considered Nelson Mandela a terrorist while others considered him a champion of human rights. In response to such cases, Nelson (2001) provides criteria for trying to settle the conflict. The narrative that is really self-constituting has stronger explanatory force, better correlates with the agent's action, and has more 'heft' (i.e. is woven around the things the agent values most). Baylis (2012) suggests that, rather than having to settle the conflict, we can just admit that conflicting views both contribute to a heterogeneous but no less self-constitutive narrative. I will not pursue this debate here. What we can take away from this, though, is that for narratives to be self-constituting, they typically need a reasonable degree of social support.

self-narrative and, therefore, only influence agency in ways the agent is largely unaware of. If the agent becomes aware of these stories they are then part of the material from which his evolving self-narrative is co-authored.

Being cognitively aware of one's self-narrative is a minimal condition for narrative self-constitution. Beyond that, the more the agent actively contributes to the co-authoring of her narrative the more self-constituting that narrative will tend to be.[30] If the agent reflectively structured her action according to a self-narrative that she had no hand in authoring then that narrative is barely self-constitutive.

One aspect of actively authoring one's narrative is to select life trajectories that are personally meaningful. The law student knows that returning to Kansas and fighting injustice is one of many socially acceptable narratives. It is another question as to whether that particular projection makes sense to her personally. If the narrative does not make subjective sense then the agent is being constituted by a narrative but it is not her own.[31] Another aspect of actively authoring one's narrative is to narrate in ways that make the best sense of one's subjective experience (past, present and anticipated). If our law student feels a surprising boredom when she imagines her future as a lawyer in Kansas she might wonder why. Her narrative was not supposed to become boring. Upon reflection, she realises that her dream to work in Kansas was misguided because she now knows that Kansas is not a hot-bed of exciting legal work; she had not updated her narrative projection. The (initially) inexplicable boredom alerted her to the fact that her narrative projection of an exciting legal career in Kansas is highly unlikely to come true. If she cannot believe that the narrative projection will be self-constitutive then it will no longer be motivating. She might then begin to consider a different projection, one that *would* provide the more exciting experience she had been dreaming of, say, joining a big law firm in New York. Had she not changed her projection and gone to work in Kansas she would have been plagued by an inexplicable boredom until she realised that the narrative that really constituted her was that of a small town lawyer dealing with petty complaints.

[30] Of course, the agent's contributions will still need to meet the other constraints on narrative self-constitution. Flagrant delusions are not made more self-constitutional just because they have been actively contributed by the agent.
[31] Sometimes the only narrative that makes subjective sense is excluded from the set of socially acceptable narratives. The agent then has to struggle to have their narrative socially accepted or has to live a relatively inauthentic life. Nelson (2001) describes some means by which agents can have their 'counterstories' heard by an unaccepting dominant culture. I do not focus on this kind of conflict since I am mainly interested in cases where addicts would be happy for their self-narrative to meet the socially accepted standards for recovery narratives.

Finally, if narrative projections are to be self-constitutive they need to meet an objective constraint; they must be causally possible. Consider, for example, Messner and Habeler who attempted the first ascent of Mount Everest without supplementary oxygen. This narrative projection met the subjective constraints on what would be self-constituting. They saw themselves as adventurous men and shared the view that climbing was best done without artificial aids. They had made several impressive ascents recently and this new, greater challenge made sense to them. Prior to the attempt society was highly sceptical, many claimed it could not be done. Had it been causally impossible their narratives would have been one of dying on the mountain or failing. However evidence of their success proved it could be done and convinced society so that their narratives became self-constitutive. In other words the constraints that individuals and society place on narrative constitution often wait on what we discover to be physically possible.

A self-narrative will be increasingly self-constitutive the more it is socially accepted, authored by the agent (so that their narrative trajectory and experience make sense), and causally possible.[32] The ability to take different perspectives on narratives is particularly relevant in narrative self-constitution because different perspectives are required to help the agent balance objectivity with imaginative self-creation (Mackenzie, 2008). When taking an acentral perspective on our own narrative we can judge the narrative's plausibility more objectively – how will this appear to others? Is this possible? When taking a central perspective, in contrast, the intensity of the narrative's cogency is more salient. That cogency can drive creative imagination better than the more detached acentral perspective but left unbridled it can create narratives that just cannot be self-constitutive. We, therefore, need to balance the perspectives we take in our own narrative self-constitution.

The recursive relationship between narrative interpretation and projection should be obvious. The better one's self-interpretation meets these constraints the better basis one has for a self-constitutive narrative projection from that point. The better one's narrative projection meets these constraints the more likely that that projection will be achieved and, therefore, provide the basis for self-constitutive self-interpretation.[33] Conversely, breaching these constraints in either interpretation or

[32] Of course it is possible to develop a self-narrative as an outcast or a social recluse but these narratives will be relatively thin in content without others to contribute to it and affirm or challenge the agent's self-ascriptions. I develop the idea of thin and fragile self-narratives as opposed to more thoroughly developed self-narratives in Chapter 6.

[33] As it happens, humans' epistemic limitations prevent this recursive process from ever running without interruption. Unexpected contingencies not readily connected to any existing narrative demand our attention and narrative projections rarely unfold exactly as envisaged. In these cases extra interpretive work is required and projections need

projection will tend to compound, resulting in non-self-constitutive narratives and failed efforts of agency. Importantly, despite these constraints, there are typically a variety of possible narrative interpretations and continuations to choose from.[34] Agency is therefore exerted through self-narration of who we were, who we are, and especially who we are becoming. Our narrative choices create a specific focus, context, and emotional character for our lives. Now it is time to look at how narrative self-constitution enhances the picture of agency developed in the first section of the chapter.

Explanatory benefit of narrative self-constitution

In this section I return to the view of agency developed at the beginning of the chapter where I supplemented Bratmanian planning agency with a more thorough consideration of the contingent elements of self-concept. Narrative self-constitution naturally accommodates a view where plans and policies are integrated with desires and other contingencies because this is exactly what self-narratives do. However, narrative self-constitution is not just a redescription of the points made at the beginning of the chapter; it also offers two broad explanatory benefits. First, when agents form narrative connections between their intentions, desires and other contingent aspects, that doesn't just describe the interaction between them but *shapes* those elements. Agency can therefore involve strategic self-narration in order to shape intentions, desires and contingencies to one's benefit.

Second, narrative self-constitution helps us understand why efforts of self-transformation succeed or fail.[35] I highlight the role narrative plays in self-transformation by considering a case of successful self-transformation and several cases of failed self-transformation. One case fails because the projected future narrative is insufficiently linked to the current self-interpretation. Other cases fail because the agent cannot form a sufficiently strong link from her present self-interpretation to her projected future. The latter cases involve what I call 'detrimental narrative momentum.' Detrimental narrative momentum explains how agents can become fatalistically trapped within negative self-narratives.

---

to be adjusted as required. The more ambitious our plans and the more unpredictable our environment, the more likely we will have to do extra narrative editing and creation.

[34] As we get older our possible narrative projections contract as we have less time to enact them but right up until our deaths we will have some different narrative possibilities in projection and in (re)interpretation

[35] The self-transformation I am particularly interested in, of course, is overcoming addiction. However I build the explanatory model here with other examples before turning to addiction specifically in Chapter 6.

Applying narrative form to intentions, desires and other contingencies

In the first section of the chapter I argued that, ideally, the agent's intentions, desires and contingencies are intelligible in light of each other. Intentions are developed to take advantage of, or mitigate, contingencies and desires while certain desires and contingencies arise because of one's intentions. In the second section of the chapter we saw that narrative self-constitution is the means by which we shape and render our lives intelligible. Clearly such narrative self-constitution must involve the narration of intentions, desires and other contingencies. Consider, for example, how the following two narratives link together intentions, desires and contingencies: "I have a strong desire to right injustice because I grew up seeing injustice all around me (contingency). It is because of that (contingent) past that I became a lawyer (plan realised), always seek to defend the poor,(policy), and aim to become a partner in the firm (plan in progress)," or, "I now love to smoke (desire), but I only started because it was cool (contingency), and I had to cough back cigarettes for ages before I really enjoyed it (plan realised)."

Two points follow from this relationship between narrative and the elements narrated. First, agency is influenced by which connections we explicitly incorporate in narrative and which possible connections we leave relatively unnarrated. Second, when desires, contingencies and intentions are organised in narrative form they are not atomistic; the narrative form or context doesn't just capture a relationship between them but changes their very *character*. In this I echo Schechtman when she says, "…the individual elements of a person's life gain their meaning – indeed *their very content* – from the broader context in which they occur" (2001, pp. 99-100, my italics). Therefore, we can deliberately adjust the character of certain elements in our self-concepts by how we narrate them.

Effects of narration versus non-narration

I will first address the effects of degrees of narration in regards to desires and contingencies before then addressing intentions. Many desires and other contingencies exist in some form without being narrated. For example, you might suddenly realise that you've been thirsty and that has been preventing you from thinking straight, or you realise people were only being nice to you because your Dad's a mafia boss. The thirst and your parentage were having effects without being part of your narrative structure. Narrating desires and contingencies brings them into cognitive focus and

makes them potential players or targets in structured planning. Desire can then be better inhibited, satiated or fostered, contingencies can be better mitigated or taken advantage of. Generally speaking, narrating desires and contingencies is advantageous. If the agent leaves desires and contingencies unnarrated then they cannot feature in structured plans. If thirst is preventing high quality thought but the agent fails to narratively connect that desire with their struggle, then they will be confused as to why they cannot think straight and won't know how to correct it.[36]

However, in some cases, the detrimental effects of certain desires and contingencies might be avoided by *excluding* them from narrative because narrating them gives them a dangerous prominence in conscious thought.[37] When an agent is trying to quit smoking, if their desire to smoke is not narrated at all then it cannot feature in plans to obtain cigarettes in which case such plans do not make sense and won't be formed. Velleman gives us an idea as to how excluding a desire from narrative might work in recounting the experience of his friend who gave up smoking.

> "The answer, he tells me, *was not to think of himself as a smoker*. … He imagined that he was not addicted – that he didn't like the taste of cigarettes, wasn't in the habit of smoking them, had no craving for them – and he then enacted what he was imagining, pretending to be the non-smoker that he wanted to be. And I suggest that this make-believe succeeded because it excluded the smoker's tastes, habits, and cravings from the story that he was enacting. That story lacked the narrative background that would have made it intelligible for him to buy, light, or smoke the next cigarette. … His motives for smoking were relegated to externally constraining his enactment of a non-smoker's story. Those motives had proved irresistible when they were available at center-stage to motivate the next episode in the story; but when they were written out of the plot and left to operate, as it were, *ex machina*, they were unable to deflect the story from its natural conclusion" (Velleman, 2002, pp. 99-100).

Perhaps this was part of the benefit the people who self-identified as 'ex-smokers' enjoyed in the study by West (above). Similarly, if one doesn't like a certain unchangeable feature of one's body

[36] More generally, it is thought that people who lack a well-defined self-concept, perhaps by being detached from their ancestry and culture, tend to suffer from an inability to form any stable plans. Agents who value various contingent aspects of themselves have a more stable concept of who they should try to become, e.g. Bennett (2002).

[37] One of the benefits of narrative is that it cuts out irrelevant information allowing cognitive resources to be focussed on important things. This means that less important desires and contingencies are ignored all the time, usually without causing a problem. We only tend to call desires or contingencies 'unconscious' when we think there is a detrimental effect from not paying attention to them, i.e. when something normative is on the line. Here I'm suggesting the opposite – sometimes making something more unconscious can be normatively beneficial.

it may be detrimental to think, 'that girl won't like me because my ears stick out.' Even if that is true, such narratives will tend to magnify the bodily feature in one's own mind making one detrimentally self-conscious and more easily embarrassed.

In order to know whether a desire or contingency should be excluded from narrative the agent will have to first narrate it to some extent so they can make a cognitively informed decision.[38] That does not undermine the point, however, because elements are not either completely narratively captured or not captured at all. Rather we variably reinforce narratives (and the significance of their constituents) by reiterating them or variations on them.[39]

> "The reality of a life lived in time is a perpetual weaving of fresh [narrative] threads which link events and lives – threads that are crossed and rewound, doubled and redoubled to thicken the web" (Lloyd, 1993, p. 144).

An element can be relatively excluded from self-narrative by minimising the number of narratives and narrative reiterations that it features in or more thoroughly narrated by involving it in multiple narratives or reiterations of similar narratives (where the narrative thread is 'doubled and redoubled'). So an agent might interpret every social embarrassment and rejection as being connected to having ears that stick out, only some, or none. Finding a balance can sometimes be difficult because overly minimal narration can result in denial and excessive narration can result in obsession. However the agent can exert some control over that balance to their benefit.

The case with intentions is slightly different. Recall that on Bratman's view agents need to develop a hierarchy of plans and policies to ensure ends-ends consistency. Plans and policies that provide the most inter-temporal connections, thus securing diachronic personal identity, should be towards the top of the hierarchy. The temporal extension of plan and policies provide those inter-temporal connections. Now, I assume that plans and policies require narrative structure.[40] Becoming a

[38] In many cases desires and contingencies are narrated without reflection anyway through various co-authoring processes. Those co-authoring processes may, of course, make it very difficult to exclude something from one's narrative. For example, it will be difficult to exclude the significance of one's ears sticking out in self-narrative if you are teased about it every day at school.

[39] I develop this view further in Chapter 6 where I argue that self-narratives are collections of multiply connected narrative threads.

[40] However, valuing something does not necessarily depend on a narratively structured intention. People suffering from severe dementia, for example, can continue to consistently value certain things even though they lack the narrative ability to construct plans in which those values can be realised, see Jaworska (1999). For example, a person may still value cooking but lack the narrative ability to know when to cook and how to coordinate a meal. If caregivers handle the wider diachronic coordination of the meal the dementia sufferer can still help and know they are helping at the time by, say, grating cheese when asked. So some values can exist without narrative but they depend more heavily on inter-subjective support than usual. The content of more complex values, however, do depend on self-narrative. For example,

lawyer, for instance, requires a complicated narrative projection outlining which steps are required and in which order. If someone did not have *any* such narrative projection they just do not have an intention to becoming a lawyer. Unlike desires and certain contingencies there is no sense in which you still have a plan or policy if you completely exclude it from self-narrative (although see footnote 40 for the limited exceptions). In that sense we can exert greater control over intentions than we can over desires and other contingencies. That said, one can still narrate an intention more or less thoroughly and this has certain effects. Developing coherent means to ends requires narration and we have already seen in Chapter 2 that this increases the chances that the end will be achieved. But narrative self-constitution goes beyond Bratmanian planning because one can also make narrative connections to an intention that go beyond settling the means. On the narrative self-constitution view those pragmatic connections are required but they are relatively sparse. One might reiterate the means to oneself many times (perhaps with some variation in the details). But one can also narratively connect intentions with a myriad of things only contingently related to the intention. For example, whenever the law student sees a case of injustice on television or in the newspaper they might narratively connect those cases to their intention to become a lawyer, 'that's exactly the kind of thing I can stop.' When making small talk at parties, of all the interests she could talk about she might focus on her law career. When trying on new glasses she might think, 'will these provide an image suitable in court?' This redoubling and more extensive inter-connecting of narrative threads better integrates the intention to become a lawyer into her self-narrative.

We achieve two things by making these further narrative connections. First, the relevant intention is kept near the forefront of our minds because it has been connected to otherwise unrelated contingencies. For example when the law student reflects on who she chatted to at the party and when she might see them again she might remember that she happened to chat about her career aspirations. The greater weight of narration woven around important intentions, therefore, tends to naturally focus our attention on those intentions at the expense of less well narrated intentions.[41] In

if a patient had a plan before dementia to see David Bowie in concert, the plan will be largely unachieved if the patient can no longer recognise Bowie or understand that this was one of his last ever concerts in a career stretching 40 years. They may still enjoy the familiar music which will resonate with them at non-narrative levels. This is a slight departure from the Bratmanian view developed in Chapter 3 where I argued that if an agent is to value something then he must commit to it with intentions. However, the Bratmanian point largely holds because without the structures of intentions the range of available values is highly limited.

[41] Thus we get the effect of a hierarchy without having to posit an excessively cognitive process where implicitly or explicitly all intentions are compared with an independent value-ranking.

other words, increased narration around an intention increases the density of inter-temporal connections associated with that intention beyond what is required for means-ends coherence. By doing so we give that intention a more central position in our diachronic identity; those intentions we narrate more thoroughly are more self-constitutive. We should, therefore, more thoroughly narrate those intentions we value the most.[42] The more extensive the narrative connections around an intention the more it will implicitly focus our cognition. We will find our minds turned to those plans and policies (thus further redoubling the narrative connections) pre-reflectively. This is beneficial as long as our plans and policies remain relatively constant but it creates an inertia that needs to be overcome if our plans and policies change. I make more of this below in regards to self-transformation and detrimental narrative momentum.

Second, creating more narrative connections with an intention can extend temporal inter-connections beyond the beginning and completion of the intention. For example, our law student might drop out and join the army but she might marry a person she met at law school. She might self-narrate, 'when I met my husband I was still really into law, we used to talk about my budding legal career all the time back then.' On Bratman's view, the inter-temporal connections of the legal career no longer ensure diachronic identity, only current intentions do, i.e. the marriage and the army career. On his view, if you change plans you cast off of that which partially ensured your diachronic identity. On the narrative self-constitution view the inter-temporal connections of the abandoned legal career still support diachronic identity if they continue to be linked to the evolving self-narrative. So diachronic identity is potentially much more stable on narrative self-constitution views than on Bratman's view; narratives can extend inter-temporal connections between different ongoing intentions, between ongoing intentions and past intentions, and between any intention and a range of contingent developments.[43]

[42] However, sometimes we over-narrate intentions that in a cool moment we would not value so highly and then those intentions tend to be overrepresented in our thought and experience. Something like this is happening in many cases of addiction and I pursue that point in Chapter 6.

[43] Furthermore, on the narrative self-constitution view, completed intentions can still be valued. The retired lawyer no longer values her legal career through practicing it but she might value it nonetheless. The narrative view can make sense of this because the narrative connections to the career can still be numerous even though none of them are means-ends connections.

Changing the narrative context of the elements narrated

More thorough self-narration not only provides an improved cognitive understanding and focus on the relationship between the elements within one's life (intentions, desires and other contingencies) but it also shapes those elements, changing their strength and character according to their position in the context of the narrative.

To show how this works I'll start with desire. Take a desire to smoke. If its development was part of a self-narrative to become cool then the desire comes to involve feelings of being cool and the satisfaction of achievement. If the self-narrative is of trying to quit smoking to become more healthy then the desire is an obstacle; its character then involves feelings of frustration and challenge. Therefore, if someone changes from the first self-narrative to the second, they reconfigure the experience of that desire to smoke from something associated with assured coolness to something associated with challenge. This may well be part of why it is difficult to change a previously accepted desire; the change in narrative context ruins what was a purely pleasant desire to fulfil.

Such effects of narrative context can result in many subtle shadings of a desire even where the intention and desires initially appear identical. Consider two different people who both aim to giving up smoking. One person has been convinced to quit by public health announcements while the other person has had to go through the ordeal of caring for her mother while she died of smoking-related lung cancer. In the latter case, part of the self-narrative might be, 'I want to quit smoking because it killed my mother and I don't want to end up doing that to my children.' The character of the desire to smoke is shaped by this context: 'it's *this* desire that lead to my mother's death and that has the potential to cause even more anguish.' When she considers giving in to the desire to smoke, the narrative connection to being beside her mother's deathbed and/or imagining telling her own children she has cancer elicits highly unpleasant emotions. These negative emotions not only make the alternatives preferable in contrast but they colour the desire so it doesn't seem so appealing; to fulfil that desire is to go down that path. These highly personal and significant memories and projections just do not occur in the self-narrative of the person convinced by public health announcements and so their desire to smoke is quite different (as is their plan to quit smoking).

We can see the same influence of narrative context on contingencies of self-concept. Becoming paralysed in a car crash is a traumatic incident but its effect is different for the eighty year-old war

veteran who already suffered from impaired mobility than it is for the aspiring athletics champion. The eighty year-old might narrate the accident as having cheated death, a stroke of luck providing a few more years to watch his grandchildren grow. For the athletics champion, however, the incident has damaged or destroyed many highly valued narrative strands [44] If he cannot narrate beyond this tragedy then his future will remain empty and his present stagnant and depressing – the contingency comes to dominate his life; the cause of all his failures and frustrations. But, if he can reconstruct his narrative future, then the crash may eventually become an event that elicits pride in a story of survival and resilience. The same contingency in the older man's narrative context is less likely to be the focus of such dramatic lows or potential highs.

The influence of narrative context on intentions is different because intentions are largely *constituted* by the narratives in which they are embedded. To change the narrative is therefore to change the intention so we cannot hold one steady while we vary the other. However, we can see the effect of narrative context on policies since policies retain a recognisable form across many narrative contexts. Consider the policy to, 'turn the other cheek.' This has a different character for someone who has been a devoted Christian all their lives than it does for the newly repentant violent offender or the atheistic pacifist. The violent offender might see the policy as particularly important given his upcoming parole hearing, the Christian sees it as a fully ingrained part of his existence and admittance to heaven, while the atheist sees it as a contribution to social cohesion. The violent offender may feel little pressure to follow the policy once granted parole because his wider self-narrative is that of a powerful man who you don't want to cross. On the other hand the Christian may follow the policy dogmatically given its entrenched position in his self-concept as a religious man.[45]

So if we treat intentions, desires and other contingencies atomistically then we will tend to see all desires to smoke as the same thing, all cases of paralysis as the same thing, all policies to turn the other cheek as the same thing. We therefore miss the fact that the character of each of these things is significantly influenced by the narrative context the agent creates for them. Nobody's desire to smoke, for example, is identical with anybody else's because it is partially characterised by the

[44] At least the context created by the old man's self-narrative more easily accommodates his acceptance of the event than the aspiring athlete's narrative context. The old man's narrative context still allows for the event to be narrated in more tragic ways, for example, as an event which forces him to spend his last years without dignity, completely dependent on others.

[45] Because the same element can be included in multiple narrative threads it can take on a multidimensional, even contradictory, character. The violent offender who trades on physical intimidation might also be a devout Christian and so have an ambivalent attitude towards the policy to 'turn the other cheek.'

way they connect it with their narrative context (past and projected). Therefore, the character and significance of desires, contingencies, sub-plans, and policies are subtly changing as self-narrative is built and edited but each can change dramatically if their narrative contexts are changed dramatically.

These effects of narrative context can be harnessed to support agency. One can try and favourably adjust the character of desires, policies and contingent elements of self-concept. For example, rather than leaving the desire to smoke unnarrated, one could give it a particularly unpleasant character by narratively connecting it with the worsening state of one's lungs and heart (aided by unpleasant pictures on cigarette packaging), letting down one's social soccer team through lack of fitness, future hospitalisation, et cetera. At the same time, pursuing the policy 'don't smoke' could be narratively linked with a self-projection of playing with grandchildren, being a highly valued member of the soccer team, being healthy, etc. In other words the emotions and motivations of other values can be brought to bear on a desire or contingency through selective self-narration.[46]

In summary self-narration provides two general forms of agential control not seen in planning agency. First, intentions, desire and contingencies can be more or less thoroughly self-narrated. The more narrative reiterations and inclusive connections that involve an element the more it influences cognitive focus and defines the context for thought. The agent can therefore exert some control over the extent to which particular intentions, desires and contingencies influence their thought and action. Second, the character of intentions, desires and contingencies can be favourably changed according to the narrative context they are given. To exert these forms of control, of course, the agent has to change their self-narrative and such changes are not always easy. The agent needs to narrate within the constraints on narrative self-constitution outlined above and they sometimes need to overcome the implicit cognitive habits built up by prior self-narration (i.e. narrative momentum). In what remains of the chapter, I consider the process of changing one's self-narrative by looking at a range of more dramatic self-transformations. Although such transformations might appear to be the execution of plans and so explicable on a Bratmanian view, there are crucial features to such transformations that only appear in a narrative lens.

[46] Kennett argues that aggregating various motivations is an important technique in self-control. Here I add that selective self-narration may support this process and make the aggregations somewhat pre-reflective.

Narrative self-transformation

The process of narrative self-constitution always involves some re-interpretation and re-projection as the agent responds to unexpected contingencies. For example, the bus is late so you have to rush into the office, or the restaurant is closed so you have to choose somewhere else. These are usually minor wrinkles in the projected story, the unavoidable results of normal epistemic limitations, and they can be taken care of without too much trouble. However, occasionally the required narrative re-interpretation and re-projection is much more demanding. Sometimes an unexpected contingency suddenly throws plans into disarray, such as becoming paralysed in a car crash or facing an unplanned pregnancy. In other cases a steady accumulation of contingencies almost imperceptibly shifts one's narrative trajectory in an undesirable direction that eventually the agent notices, for example, realising that one is addicted or that one's career has become depressing. In other cases the agent might be aware of the building evidence that their narrative projection is becoming unlikely but they stick to that projection out of increasingly desperate hope. At some point they may have to accept that the evidence makes hope unreasonable.[47] These situations force a choice between two (or more) incommensurable narrative continuations or require one to re-project the future afresh. The agent has to re-narrate who they are at a relatively fundamental level; this is self-transformation.

Considering self-transformation as a process of self-narration helps us understand how agents manage to self-transform and why they sometimes fail. To illustrate this I will consider one case of a successful narrative change and three particularly challenging cases where failure to change is a real possibility. The first challenge is created by inappropriately connecting a projection to one's self-interpretation. The second challenge is to create a plausible narrative projection when none seem available. The third challenge is where the detrimental momentum of one's existing narrative prevents one from building connections to a more highly valued alternative.[48]

[47] These processes involve self-deception when the agent remains hopeful despite overwhelming evidence that their narrative projection is unrealistic, or the agent remains committed to a self-narrative interpretation despite overwhelming evidence that it is inaccurate.

[48] In these examples I assume that the current self-narrative is sufficiently accurate, it is the creation of, and connection with the new, transformed self-projection that is in question. This seems to ignore the challenge to self-transformation created by a deluded self-interpretation. However, because of the recursive relationship between projection and interpretation, such a challenge can be seen as a variation on the 'misconnection' example below.

Successful self-transformation

To see how self-narration is involved in successful self-transformation, imagine a young, single Catholic woman who has unexpectedly fallen pregnant and the father has left her.[49] She is now torn between abortion and pursuit of her career on the one hand and having the child in accordance with her Catholic values on the other hand; she faces ends-ends inconsistency. On the Bratmanian picture, she needs to abandon or relegate one set of conflicting plans and policies, ideally so that the resulting hierarchy of intentions will maximise inter-temporal connections. But how does she go about making this decision?

A narrative self-constitution account provides the required detail. In such situations the agent considers the various narrative continuations she sees available; each narrative projection gives an approximation of the emotions, decisions, and interactions that will result from that path. In this case, one self-projection involves doing the morally right thing according to Catholicism and facing the demanding/rewarding role of a single mother but giving up her career and losing independence. The other involves a moral failure (in her eyes and that of her church) but the possibility of continuing her budding career while retaining the possibility of having children later. Imagining the abortion narrative she can consider the conversation with her parents, the interactions with doctors, the medical procedure, the gossip at the church, but also the freedom to continue a career and the related financial independence. Imagining the motherhood narrative she envisages the unrelenting demands of the baby, a different conversation with her parents, the pride and love that other mothers talk about, but a restriction in social connections with her peers, the increased difficulty in finding a husband, a distancing from her career and the risk of long-term financial dependence on her parents. Each imagined narrative projection helps her bring to mind exactly what she will be in for, practically and emotionally, given the choice to abort or raise the child.

This decision-making technique works well when the considered narrative projections meet the constraints of narrative self-constitution. They should be causally coherent, correctly judge the reactions of society and relevant individuals, and provide an understanding of one's anticipated experiences (and therefore one's projected actions). In this case the woman does this adequately well for both narrative projections. As it happens, when the woman imagines the discussion with her parents about having an abortion she is surprised to discover she is angry with her parents. Searching for a source of this anger she comes to believe that her parents foisted their religious

[49] I borrow this example from Mackenzie (2008, p. 130).

values on her and she begins to revise her Catholic commitments. The career narrative in contrast then becomes an especially poignant way of authoritatively signalling her own direction in life. In this case the process of imaginative projection has revealed which of the two possible narratives will be the better – one makes better sense of her emotional responses and provides a course of action that better represents her (changing) values.[50] As a result she has continued her self-narrative in a way that will provide more stable inter-temporal connections than had she chosen the path her parents preferred or if she had failed to resolve the ends-ends inconsistency.

In one sense, then, narrative self-constitution is just a way of filling in the detail of how humans achieve the norms of practical agency that Bratman highlights in his account of planning agency. However it goes beyond basic planning agency in that it makes sense of how agents can value contingent aspects of themselves. Contingencies can be woven into self-narrative so that they, like plans and policies, provide inter-temporal connections securing diachronic identity. An alternate narrative projection of our young woman can make this clear. She recognises that her Catholic values are contingent and that had she been raised by Muslim, Protestant or atheist parents than she would have had those other sets of values. She does not believe that the Catholic values are better for her than those other values; the Catholic values just happen to be the ones she has grown up with. However, they connect her to her ancestors and if she has the child she realises she would want it baptised. So she comes to think that, despite the contingent nature of her Catholic values, she values them highly enough that she will have to go through with having the child. My intuition in this case is that it is possible that the woman exhibits self-governance even though the value she attributes to contingent policies and rituals clashes with her carefully forged (less contingent) career plans.

A narrative view can accommodate the possibility that commitments to contingent aspects of self-concept exhibit self-governance because of the inter-temporal connections afforded by some contingencies when self-narrated. Bratman does not consider the supportive role of contingencies in diachronic self-concept; only plans or policies that derive from the agent's efforts result in self-governance. Therefore, for him, a commitment to a contingency at the cost of a personally developed plan or policy is necessarily a perverse means of undermining self-governance, a form of bad faith. But if my intuition about the Catholic woman is correct then this is not always the

[50] Of course, this process of narrative projection might not result in a clear cut decision and ultimately she might just have to plump for one. However, even in the cases where the choice remains incommensurable, narrative projection, done well, provides a more complete idea of what alternatives of yourself you are actually choosing between.

case when the contingency has been reflectively incorporated in self-narrative. On the other hand, narrative integration of contingent aspects of self-concept do not always support self-governance and I consider some of the negative effects in the examples below.

Self-transformation challenge 1: misconnection

In other attempts at self-transformation the agent does not manage to adequately connect their imagined future with their current situation. If we return to the young Catholic woman who is choosing between the abortion and keeping the child, she might begin to imagine a third possibility whereby she keeps the child *and* maintains her career. She imagines a conversation with her boss which ends with her position being guaranteed upon her return shortly after the birth. Babysitting could be made more affordable by often relying on her parents who, although not yet retired, might go part-time to help her out. However, this imagined future happens to be highly implausible. Her boss has no particular incentive to protect her job. Her parents are unlikely to acquiesce to those terms. The baby will be much more demanding on her time. As a result her attempts to live out this future will inevitably fail at some point – perhaps her boss or parents don't agree to her terms or perhaps they agree but then the demands of both career and baby are too much.[51]

Bratman would explain such cases of misconnection as failures to develop means-ends coherence in one's plans. The narrative self-constitution account can add detail on how exactly such agential failures occur; the agent imagines implausible narrative connections without realising that they are implausible. This might happen for two reasons. First, she finds herself trying to imagine details of unfamiliar scenarios. She has never raised a baby and so even her best efforts to imagine it are necessarily limited. That is not to say agents cannot benefit from their limited efforts but just that in unfamiliar territory they need to scale back the ambition of their projections; sometimes we need to hedge our bets. Second, the woman might have got carried away by the cogency of the narrative she was imagining. Her projected narrative had all the positive emotions of being a new mother, keeping her job, and maintaining her Catholic values. She avoided the negative emotions elicited by the moral transgression and the damage to her career. Buoyed along by these positive feelings

[51] As she attempts to live out this unrealistic projection it will become more and more disconnected with the current facts without her realizing. Perhaps her old job has already been filled by someone else. Perhaps even after raising the child for 5 years she still thinks she might walk back into a similar job when the job market has changed markedly. At that point her unrealistic projection has become a deluded self-interpretation. Any projection she bases on that deluded self-interpretation will then suffer from an inherent misconnection.

she failed to carefully consider whether this was a plausible narrative. As a result she failed to carefully consider how her parents and employer would really act and she overestimated her own abilities. She, therefore, switched from imagining her own life to imagining someone else's. As Mackenzie states, "I can imagine myself living all sorts of different possible lives, and I can imagine these lives from the inside, as it were. But in doing so, I may not be really imagining *myself*. Rather, I may just be engaging in a kind of make-believe, imagining different lives that have no real connection with me other than that I imagine them…" (Mackenzie, 2008, p. 124). To prevent this kind of misconnection between one's present and future self-narrative the agent needs to better assess her narrative projections. As mentioned above, this can be done by getting input from other people and/or by taking an acentral or external perspective on one's imagining. By stepping back from the internal perspective one can better reflect on the emotions elicited by the imagined scenario. Of course this external perspective is not truly external and so is vulnerable to being swept along in the fantasy itself but it is more objective than the internal perspective, see Mackenzie (2008) and Goldie (2012).

Self-transformation challenge 2: narrative stagnation

In another kind of situation the agent desperately wants to self-transform but they cannot see how to go about it. Take the case of the recently paralysed athletics champion. He is aware that because he cannot walk he must significantly change his life but it isn't obvious how to do it. Unlike the young Catholic woman there aren't two (or three) unavoidable options. There are at once too many options and no options. That's to say, there are many ways a young, wheelchair-bound person could live their life but none of those options stands out as an obvious option from the particular perspective of this young athlete. On the Bratmanian view the paralysed athlete should be able to form new plans more easily than he could before because the constraints of ends-ends consistency have relaxed. New plans no longer have to be fitted around demanding training programs. The agent has a relatively blank canvas. The narrative view, in contrast, gives us a way to understand why people find it difficult to make new plans in these situations.

The athlete knows who he was and how he got to the present but he can no longer imagine who he is becoming. Whereas the young Catholic suffered from a relatively unbridled imagination the paralysed athlete needs to tap in to some of his imaginative power. He may struggle to properly engage with the imaginative project because all the options seem poor in comparison with the

narrative trajectory that he had carefully fostered and identified with. If Mackenzie (2008) is right, the athlete might also benefit from trying to take a central perspective where imagination is less restricted than an external perspective. He might be being overly critical of the possibilities available. [52] In this case the paralysed athlete is still learning what it will feel like to live as a paralysed man, so he may be magnifying the feelings of frustration in his projections from a central perspective. Those feelings will reduce over time but early in rehabilitation that is hard to believe. The necessary narrative change could be fuelled by inspiring narratives of other people who were paralysed at a young age. These narratives provide precedents; they are some indication of what narrative projections are possible and how they can be plausibly attained. Of course he cannot imagine living out these other people's stories exactly because he has to connect a narrative projection to his own narrative self-interpretation. However, these other stories provide some building blocks that he can make use of.

The paralysed athlete also needs to begin to dissociate from many of his prior narrative threads in order to open his mind to other possibilities. Therefore the self-transformation task is not always just to connect with a plausible future; it can also require that one reinterpret oneself in the present so that other possible futures become plausible. The athlete therefore needs to change his self-narrative from that of a promising young athlete to that of a wheelchair-bound young man before the plans of a wheelchair-bound young man can make sense. New plans do not just have to be ends-ends consistent and means-ends coherent with other intentions, they have to be narratively connected to the rest of one's self-narrative. That process can be slow and painful given that the 'athlete' narrative was entrenched through years of activity and social interaction while the 'paralysed' narrative, although correct, seems superficial in comparison. An agent attempting to transform, therefore, needs social and practical situations where the process of laying new connections and redoubling them can take place. Without the chance to regularly self-narrate in

[52] There is evidence that trying to imagine oneself centrally in an alternate life can help develop relevant affective connections with that possible life. Galinsky et al. (2008), for example, had participants look at photos of either: a cheerleader, an elderly man, a professor, or an African American man. They were then asked to pretend to be one of these people and write about their typical day in the first-person (thus encouraging a central perspective). Control participants did the same but wrote about the person from a third-person (i.e. acentral) perspective. When the participants taking the central perspective subsequently rated their own traits the 'professors' felt smarter, the 'cheerleaders' felt more attractive and sexy, the 'elderly' weaker and more dependant, and the 'African Americans' more aggressive and athletic. Behaviour was also affected. Compared to controls, 'professors' had improved analytic skills and 'cheerleaders' had impaired analytic skills. 'African Americans' behaved more competitively in a game than 'elderly men'.

conversation or in response to practical pressures, suggested narrative continuations are likely to remain disconnected, abstract and alien.

Self-transformation challenge 3: narrative momentum

The final kind of case I will consider is where the agent has developed a well self-governed life that they now want to change but find they cannot do so. Imagine a successful lawyer who is on track to achieve his career-long dream of being made partner. Despite this, he has recently realised that he now values time with his family more than ever and being made partner is not compatible with the father and husband he wants to be. He cannot bring himself to reduce his time at work and effectively pull out of the competition to become partner. The lawyer fully understands *how* to reduce his time at work and increase his time at home but he cannot bring himself to implement those intentions.[53]

On the Bratmanian picture this kind of case is a mystery. There is nothing stopping him from changing his plans so if he will not change them then he cannot sincerely value a greater emphasis on his family life. A narrative account, however, can explain his action with what I call 'detrimental narrative momentum.'[54] Daniel Dennett famously described narrative momentum (although not with that term) as follows:

> "Our tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness, and our narrative selfhood, is their product, not their source" (Dennett, 1991, p. 418).[55]

Narratives can end up 'spinning us' because the complex of narrative threads that we have redoubled and woven together over time cannot be abandoned by fiat. We cannot just subtract those threads that no longer suit our values from those that do – they are thoroughly inter-connected. But

[53] At this stage I assume that these kinds of cases occur. In Chapter 6 I present evidence that, in addiction at least, such cases really do occur.

[54] Narrative momentum is beneficial when we are happy with the direction we are going in. The lawyer may be pleased with his path towards being made partner; he's worked hard to make this the most intelligible continuation of his narrative.

[55] Although Dennett tends to slide into a more extreme position claiming that we *never* spin our own narratives. Unlike Dennett, I think we spin our own tales at least some of the time. The agent therefore has some existence and control independent of particular narrative threads. Part of that control involves making sure they don't let the narrative momentum get out of control.

as long as those, now out-dated, constituents of our self-narrative remain in place they will continue to influence the focus and context of our thought and affective responses. [56]

In the case of the lawyer, it is plausible, I assume, that his longstanding plan to become partner has resulted in a variety of narrative threads that are completely enmeshed in his wider narrative. The causal effects of those threads go beyond their current value to him and those effects can undermine the agent's efforts to change them – this is detrimental narrative momentum.[57] Detrimental narrative momentum hampers the agent's efforts to change his life in at least two ways.

First, the focus created by the existing narrative conceals certain interpretations and projections while making others more salient. This effect can blinker the agent, preventing them from noticing opportunities for change around them while focussing detrimentally on events that tend to reinforce the apparent benefits of the current narrative. The lawyer, for example, might not notice that school holidays are coming up (despite the information being readily available) and so he does not think to schedule his own leave to match it. His narrative-driven attention biases lead him to aggrandise his minor efforts at being a better family man and downplay his shortcomings. He overplays the significance of spoiling the children with Christmas presents, assuming such gifts make-up for his absence when, in fact, their effects are transient. He underplays his wife and children's disappointment when he misses the school play. These biases tend to prevent and assuage his worry that he really needs to do more to live up to his value of being a family man. Second, the context created by the existing narrative can give events an emotive hue that tends to undermine change. When the lawyer's boss, aware of his family situation, suggests that he could reduce his work hours, the lawyer cannot help but feel depressed and angry at what would be a set-back to his chances of becoming a partner. This prompts him to work harder to show the boss that he is good enough to be partner even though in moments of greater reflection he knows the family man narrative will require he scale back his hours.

[56] Bratman's view can, in fact, explain some such momentum because the agent's intentions are interconnected. It can be hard to remove one plan when other plans and policies partially depend on it for diachronic stability. However, narrative momentum goes beyond these effects because narratives include that not only that network of intentions but also the connections between that network and the agent's wider self-concept. On a related point, Holton (2009) notes that we can be excessively stubborn in failing to reconsider certain intentions; this involves a kind of irrational momentum. Here I am suggesting that one reason we might be overly stubborn in certain cases is because of the way that intention is connected to the rest of our narrative. Self-narrative can tend to prevent rational reconsideration.

[57] This is one reason why some people trying to quit smoking in West's study (above) may have been unable to adopt the self-concept, 'ex-smoker.' Such an element of self-concept was incompatible with the detrimental self-narrative that had come to support their diachronic identity.

The lawyer's problem is analogous to that of the paralysed athlete; he cannot make the family man projection his *own* as opposed to just an intelligible possibility. The difference is that the lawyer's narrative of becoming partner is still an alluring possible future while the paralysed athlete's athletic future is impossible. To overcome this barrier the lawyer needs to develop connections with the new narrative and break off the affective connections to the old narrative. He can work towards this by consistently imagining "family-man" projections that elicit positive emotions, say, where he is playing with his children, his wife is smiling, they are going on holiday together, et cetera. Simultaneously he needs to stop day-dreaming about that glorious scenario he has played out in his mind for years where he is called into the boss's office, handed a scotch, and told that he's been made partner. He needs to stop posturing at the water cooler as if he is in line for a promotion and start boasting about the family holiday he is going to take.

Detrimental narrative momentum will tend to involve a somewhat fatalistic attitude to how one's story is unfolding but an agent in the grip of detrimental momentum is not necessarily passive; they will act in ways that their narrative demands. Neither is such an agent just 'going with the flow' – doing what feels best. The lawyer might complain that he has to stay in late again at the office and genuinely regret and dislike that he has to do so. However, his late night commitment is not coerced by others or a work culture but by aspects of his own self-narrative.

Conclusion

I began this chapter by pointing out that a variety of contingencies are just as much a part of the agent's self-concept as her plans and policies. We saw how intentions, desires and other contingent elements of self-concept are partially independent but need to be responsive to each other for effective agency. We also saw how contingent elements of self-concept that the agent associates with negative outcomes undermine agency while positive associations enhance agency. This interaction between self-concept and agency indicates that self-governance does not just involve normative commitment to a hierarchy of intentions but also development of a supportive self-concept.

In the second half of the chapter I outlined an account of narrative self-constitution. This theoretical move was motivated by the fact that self-narratives are inclusive of the agent's network of intentions as well as their interpretations of their desires and other contingencies. Narrative self-

constitution views claim that agents govern their lives by creating the narrative projections they want to live out and then enacting them. This goes beyond forming plans and policies because it involves the selective integration and interpretation of the variety of contingencies in their lives. But the agent cannot just self-narrate in any way they want. If self-narratives are to be self-constituting they must be inter-subjectively accepted and physically possible. The narrative trajectory chosen must also make sense to the agent and render their past and anticipated experience intelligible.

I then argued that narrative self-constitution nuances our understanding of the motivational influence of intentions, desires and other contingencies. The force and character of our intentions, desires and contingencies are influenced by how thoroughly we narrate them and what narrative context we put them in. Therefore, self-governance can be enhanced through strategic narration or undermined by detrimental self-narration. These effects and means of control do not appear on a Bratmanian account of agency.

I concluded the chapter by considering cases of self-transformation. When making self-transformative decisions, maintaining the continuity of significant contingent aspects of one's life may count as much as the continuity of certain plans and policies. Such decisions require realistic imagining of the available narrative paths otherwise the agent risks misconnection with their chosen future. To avoid misconnection in narrative projection the agent must temper an emotionally charged central perspective with the greater objectivity of an acentral perspective. Some self-transformations are especially challenging because much of one's existing narrative must be abandoned (narrative stagnation), others because narrative threads mutually exclusive of the transformation are so well-entrenched (detrimental narrative momentum). In the former case the agent has to build a new self-interpretation before any projection can make sense. In the latter case the agent has to replace the detrimental narrative threads despite them having been a central component of their self-concept for so long.

Plans and policies should be relatively easily created or abandoned when the agent's values change so it's not obvious why these self-transformations should be difficult. Self-narratives threads, in contrast, can be much more thoroughly inter-woven with other indispensable elements of self-concept, they can be psychologically entrenched more than practically required, and they pre-reflectively influence attention, the context for thought, and emotional responses. In contrast to the narrative self-constitution account, Bratman characterises self-transformation as a change in a

hierarchy of plans and policies. This not only glosses over the detail of the process but cannot account for the weight of contingent aspects of self-concept in decision-making, narrative stagnation, or detrimental narrative momentum.

We have now considered three general views of agency: a choice account promoted by Ainslie and Heyman, a normative planning account promoted by Holton and Bratman, and a narrative account. The rest of the thesis is devoted to comparing the power of these accounts to explain addiction. In Chapter 5 I will argue that the choice theorist's claim that agents necessarily do what they most want can only provide a rough explanation of some drug-using behaviour. This view cannot explain the paradigmatic cases of addiction where agents struggle to recover and often fail. The planning account is better placed here because it makes conceptual room for actions that break norms of practical agency and undermine self-governance. In Chapter 6 I return to the narrative account developed here and argue that it can better explain why addicts struggle to recover than the planning account. Many addicts do not just need to change plans and policies; they need to engage in a much more involved process of narrative self-transformation. The struggle with addiction frequently involves misconnected narrative projections, narrative stagnation, and detrimental narrative momentum.

Chapter 5: Choice accounts versus planning accounts

Introduction

In the first four chapters I assessed three different theories of agency with a focus on their ability to explain non-addicted agency. I argued that the normative planning account of agency supported by Bratman and Holton is superior to the reward maximisation (or choice) account of Ainslie and Heyman. My central argument was that planning agency provides a role for the agent in achieving self-governance while Ainslie and Heyman reduce agency to the result of a range of contingent factors. I then argued that the concept of narrative self-constitution helpfully builds on planning agency. The main advantages of the narrative view are that it accommodates variations in self-governance that depend on agents' interpretation of contingent aspects of their identities and experience and the way they weave their intentions together with these contingencies.

What I have yet to do is to make clear what these differences mean for our understanding of addiction, recovery, and relapse. With this in mind, I have two goals over these final two chapters. First, the challenge of explaining addiction provides new grounds for me to argue for the benefits of the planning account over the choice account, and for the narrative account over both of those positions.[1] Second, in comparing these accounts of agency we can improve our understanding of addiction and how it should best be treated. In this chapter I argue that the planning account can explain a greater array of addiction phenomena than the choice account and can better inform addiction treatment. In the next chapter I argue that the narrative account makes similar improvements over the planning account.[2]

This chapter proceeds as follows. I begin with a brief recapitulation of the choice account, setting out the cases of addiction, recovery and relapse it can best explain. Those cases are where people arguably maximise reward through drug-use. The choice account can explain why some drug-users have miserable lives despite maximising reward if they either genuinely lack access to greater rewards or if the 'toxic' nature of addictive rewards reduces their awareness of greater rewards. I then move on to consider the explanatory shortcomings of the choice account. First, it cannot make sense of the distress, ambivalence, and struggle that many addicts report synchronic to their drug-use. These addicts are well aware of preferable ways to live even as they use. Within the theoretical bounds of the choice account we are forced to describe such addicts as either deceptive or behaving

[1] Given that understanding addiction entails understanding its relationship with typical agency, one would expect that difficulties in explaining typical agency would entail difficulties in explaining addiction.

[2] Because the narrative account builds on the planning account it can help itself to the explanatory advantages the planning account has over the choice account.

unintentionally; neither explanation is adequate. In any case, this is a false dichotomy but to escape it we will have to abandon the choice account. Second, the choice account cannot explain why both recovered agents and clinicians refer to recovery as an *effortful* process. Third, anticipating the objection that these cases of struggling addicts are a tiny minority, I argue that these people are the paradigmatic cases of addiction.

In the second half of the chapter I consider the planning account. I begin by arguing that it can explain the struggle and distress associated with addiction with reference to effortful self-governance. It can also explain a variety of other addiction phenomena with reference to specific practical norms. First, the resigned fatalism of some addicts can be understood as a failure of means-ends planning to either *create* motivating goals or *connect* with known goals. Second, failures of means-ends planning can explain why addicts often claim to have goals that they chronically fail to achieve. Third, ambivalence can be characterised as a failure to find ends-ends consistency among one's intentions while the incommensurability of rewards can explain why that detrimental end-ends inconsistency is chronic. Fourth, the planning account provides a way of understanding why treatments that support these practical norms work.

I finish the chapter by considering a further explanatory challenge raised by a subset of the paradigmatic cases of addiction. Members of this subset not only know of and have a preferable alternate lifestyle available to them, as all paradigmatic cases do, but they also exhibit strong self-governance in domains other than drug-use. Why don't these people use their skills of self-governance, which one would expect to be domain-general, to recover? This behaviour remains inexplicable on the planning account and we have to turn to the narrative account in the next chapter to make any headway.

Choice accounts of addiction

Revisiting the choice account

As we saw in Chapters 1 and 2, Ainslie and Heyman both present versions of the 'choice' account of addiction; that is, they think that the choices involved in addicted action are best understood using the same motivational principles as for any other choices (Heyman, 2009, vii). The key motivational principle in action, i.e. voluntary behaviour, is that an "individual is constrained to choose the option with the greatest expected reward of all those she considers" (Monterosso &

Ainslie, 2009, p. 116). Or, in Heyman's words, "individuals always choose the better option. This is true by definition" (2013, p. 432). It is, therefore, impossible for the agent to intentionally pursue a lesser reward over a greater reward when she takes both to be available. This is the revealed preference view of classical economics (Samuelson, 1938); whatever the agent chooses reveals what they most prefer.[3] If an agent does not choose something it is because she does not prefer it or does not believe it to be available.[4] The result is a clear bifurcation – voluntary behaviour pursues maximal expected reward while everything else counts as involuntary behaviour.

Choice theorists take their view to stand in opposition to strong versions of the disease model of addiction which hold that addicts involuntarily use drugs due to damaged neurological mechanisms in the same way that someone suffers a heart-attack due to the mechanisms of cardiovascular disease. Choice theorists argue that habitual drug-use typically responds to the agent's knowledge of local consequences, e.g. the behaviour changes according to the price of the substance. On this view, that makes the behaviour voluntary and so it is not a disease. They reserve the term "disease" for involuntary processes defined as processes that are unresponsive to the agent's knowledge of local consequences.[5] In what follows I focus on the choice account's explanation of addiction and leave aside the question of whether addiction should be categorised as a "disease" or not.

The first hurdle for the choice account is to explain how it is that addicts are maximising their rewards, given that continued drug-use or relapse typically results in serious adverse consequences, such as the loss of health, significant relationships, respect, and career, losses of which the user is well aware.[6] Ainslie and Heyman attempt to deal with this issue by claiming that continued drug use is the pursuit of the maximum *perceived* reward. The drug-user, like all agents, is subject to certain systematic perceptual errors caused by the default tendency to discount the value of future

[3] I set aside the problem of interpreting exactly what a choice reveals about an agent's preferences. For example, is the agent helping an old woman or trying to win her confidence to swindle her?

[4] This position entails the further commitment that all rewards are commensurable and only distinguished by size and time of availability. If rewards were incommensurable then the agent would need an additional means of choosing between them that the choice account does not provide.

[5] It is obviously problematic to force all behaviour into either the voluntary and healthy box or involuntary and diseased box. First, we are happy to call a variety of conditions diseases even when the aetiology of those conditions clearly depends on voluntary behaviour, e.g. type II diabetes, smoking related lung cancer. Second, we need to be able distinguish degrees of control in action beyond voluntary and involuntary. If the agent's ability to respond to consequences is extremely limited it may count as voluntary but it exhibits less self-governance than if the agent could have responded to consequences in a more sophisticated fashion. The planning account can make these distinctions because self-governance comes in degrees. I develop this point below

[6] Indeed, one of the DSM IV criteria for substance dependence is, "the substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance."

rewards at a rate that approximates a hyperbolic curve. Left unbridled, such discounting distorts the reward size of distant rewards, making them appear smaller relative to more immediate rewards. On Ainslie's view, when these rewards are mutually exclusive the agent's preferences for one or the other chronically reverse over time and he never manages to access the larger later reward. Therefore, periodically, the addict believes drug-use will maximise reward even though each time after he uses he realises that he was wrong. Heyman describes the process slightly differently. On his view, the agent is not wrong to choose the smaller sooner (local) reward. Such actions maximise synchronic reward which, he claims, is a normal way to choose. It is just unfortunate that in cases of addictive rewards local choice will have unforeseen, highly destructive diachronic consequences.

As we saw in Chapter 2, if we had no way of countering hyperbolic discounting then we would all chronically overeat, drink too much, oversleep, and become addicted if exposed to addictive rewards. The reason we are not all like this, Heyman and Ainslie claim, is because of a range of self-control techniques: we follow prudential rules (Heyman[7]); choose whole series of options rather than one option at a time (Heyman[8]); use test-case willpower, i.e. treat the current choice as a precedent for future choices of the same kind (Ainslie); and adopt certain pre-commitments (Ainslie). To the extent that our behaviour is diachronically structured in these ways we maximise reward over a diachronic timeframe. This would appear to make room for self-governance – the more effectively the agent employs these techniques the better they govern themselves. However, as I argue below Ainslie and Heyman efface this potential self-governing role for the agent. The choice account faces a number of explanatory shortcomings but first I will outline its explanatory benefits.

Explanatory success of the choice account

The choice account has some success in its goal to explain all addiction in terms of agent choice (choices determined by sub-personal reward expectations and the contingent availability of

[7] Heyman may mean these rules to operate like diachronically stable intentions because part of their appeal, like intentions, is their low cognitive demand compared to global choice. Ainslie, however, is explicit that we always make our choices synchronically; we never rely on an earlier choice to guide us now, see Chapter 2 and Monterosso and Ainslie (2009, p. 122).

[8] Although, as we saw in Chapter 1, Heyman doubts our ability to do this very well because of the cognitive demand, hence the appeal of prudential rules which have a relatively low cognitive demand.

rewards). Taking Heyman's lead, we can roughly divide the drug-using population into two groups. The first group includes roughly 80% of those people who, at some point in their lives, are habitual drug-users. People in this group begin using drugs as young adults but quit by their mid to late thirties without treatment. The second group, roughly 20% of people who are habitual drug-users at some point in their lives, begin at the same age but do not stop in their thirties. Here I argue that the choice account can explain the drug-use trajectories of some people in both groups. But even in the cases it can explain, it appears to gloss over the likely struggle the agent faced.

The choice account can explain why people from both groups choose to start using drugs. Young adults find that drugs provide good "highs" and drug-use comes with rewards of social status and belonging. At the same time, competing rewards, such as raising a family and pursuing a career, are not yet available or do not appear significant.

First-person reports of early drug-use support this view. People frequently cite a positive experience and so the explanation that these people are maximising their reward seems plausible. A couple of our interviewees recount their experience:

> "When I was 20, 30, when I was 40, my drinking was good, I had good times on the drink…" (R24, our interviewee).

> "Heroin is an astonishing thing. I will never regret taking heroin. In fact those two years I took heroin are actually one of the best two years of my life. … you realise how much pain you carry around with you, all the time, and this stuff stops it. … I was … trying to write a novel at the time, and it made writing so easy, so it's just, so pleasurable. It made everything so pleasurable. It's just really good fun. Like alcohol, or the rest of it, they're just, they're rubbish by comparison" (Interviewee 101, our pilot study).

According to the choice account, the 80% of people who eventually quit do so because their reward landscape changes as they age. The "high" from drug-use decreases due to habituation, the rewards from social status and sense of belonging decrease while the burgeoning rewards of career and family begin to outcompete drug-use. Simultaneously, perhaps the skills to choose globally, set choices as precedents, form pre-commitment strategies, and adopt prudential rules develop with age. In other words, the choice theory claims that these people stop using drugs because they find other ways to maximise reward.

Heyman provides the following as an example of this voluntary change: Patty, a mother of two young girls and the president of the parent-teacher association (PTA) has had a cocaine habit for fifteen years and continues to sell cocaine. She narrowly avoids arrest for dealing and this makes her realise that selling cocaine is much more risky than she thought and so she considers the choice to stop selling. But if she stops selling cocaine then her reduced income will mean she also has to stop using cocaine if she is to feed her daughters. Other considerations are that her daughters have been embarrassed by the people coming to buy cocaine and that it won't be a good look if the president of the PTA is busted for dealing cocaine. She chooses to quit and Heyman takes that choice to be motivated by her expectation that the reward from being a good mother and PTA president will outstrip that of cocaine use (2009, pp. 60-61).[9] If we assume that all the people in the 80% who recover without treatment have an analogous story where they choose to quit and then do, then the choice account appears to have explained all these cases of addiction. Below, I question whether all these people really do recover so easily and, for those that do, whether those cases really count as cases of addiction. Those issues aside, what about the other 20%, the longer-term drug-users who do not quit?

The choice theorist claims that these people do not choose to stop using drugs because, for this minority, it remains the way to maximise reward. This is either because, for them, the reward of drug-use does not decrease or competing rewards do not arise. There is evidence for both explanations. First-person reports support the view that, for many, drug-use remains a rewarding aspect of life. These people strive to keep drug-use in their lives or choose to return to periods of use after periods of abstinence. They are aware of other rewards but just do not see them as being that valuable.

> "Most heroin users I know use heroin because they like it. Some of them are dependent, and they'll get sick if they don't use it, but all of them have given up heroin lots of times in the past, and yet they've wanted to use it again. You could say they've got a disease of addiction, but you could also say, 'No, they actually like using heroin'" (Rhys in Karasaki et al., 2013, p. 201).

As Rhys says, periods of abstinence and relapse may just be a natural variation in preference. This is exactly what the choice account claims – relapse is voluntary because it occurs when drug-use,

[9] Although there is an alternate explanation whereby her choice is motivated by a more holistic goal; she might be choosing who she wants to become. I consider this in Chapter 6, see also Kennett & McConnell (2013).

once again, becomes the way to maximise reward. This might be triggered by a sudden loss of competing rewards (e.g. losing one's job or spouse) or drug-use might, for some reason, be expected to provide a greater reward than it has in recent times.

What about those people in this 20% of longer-term drug-users who clearly live miserable lives? In these cases the choice account claims that better rewards are either not available or the agent cannot become (consistently) aware of them. Heyman suggests that these people suffer from comorbid mental illnesses that undermine the possibility of accessing rewards greater than drug-use (Heyman, 2009, p. 84). They might be less able to choose globally, adopt protective prudential rules, exercise test-case willpower, or set pre-commitments. However, none of this would entail that they are not voluntarily continuing to use drugs. On the choice account, a synchronic choice is just as voluntary as a diachronic choice. From the perspective of these people, life would be even worse without the, perhaps meagre, reward that drug-use still provides.

As with the first group of habitual drug-users, I also have some concerns about whether the actions of this second group of longer-term habitual drug-users has been adequately explained. No doubt some people happily continue using drugs throughout their lives. It is also likely that some people begrudgingly continue using drugs because they have nothing better and they would abstain as soon as better options presented themselves. However, I suspect that a large proportion of this second group are consistently failing to recover from addiction and so can hardly be described as choosing the life they lead. As we will see below, first hand reports of addicts and clinicians support my suspicion. Furthermore, these chronically struggling addicts are arguably the paradigmatic cases of addiction. In contrast to these cases, the cases that the choice account can explain seem less like addiction and more like fully controlled drug-use.

Explanatory limitations of the choice account

The major issue for the choice account is that it makes no conceptual space for effortful self-governance. The agent makes choices, such as to use drugs, abstain, et cetera, but those choices are determined by sub-personal and external contingencies. This leaves the choice account unable to explain a range of addiction phenomena. First, many addicts report an experience of conflict and struggle against addiction that depends on being aware of greater rewards even as they choose lesser rewards. By insisting that action must maximise perceived reward the choice theorist is

forced into a false dichotomy – either such addicts are lying (or self-deceived), or they are not acting intentionally. Neither explanation is satisfactory but we have to move beyond the choice account if we are to avoid the dichotomy. Second, the first-hand reports of recovered addicts and the reports of AOD (alcohol and other drugs) professionals attest that the addict must take responsibility for his recovery. Recovery is not a matter of just waiting for alternate rewards to become consistently available and visible; it requires effort. I explain why pre-commitments, test-case willpower, global choice, and prudential rules do not count as efforts of self-governance. Finally I consider a possible response on the choice theorist's behalf – people who chronically struggle with addiction are a tiny minority of addicts and so hardly a serious oversight. I argue that these chronically struggling addicts are the very cases we most need to understand; they only appear a minority when swamped with a range of cases that only trivially count as addiction.

I begin with the most glaring evidence that addicts face a struggle for self-governance – their first-hand reports. These reports clearly indicate that they experience a protracted and distressing struggle to overcome addiction.

> "I tried everything. I moved a thousand miles away from home to Chicago and a new environment. I studied art; I desperately endeavoured to create an interest in many things, in a new place among new people. Nothing worked. My drinking habits increased despite my struggle for control" (AA member in Alcoholics Anonymous World Services, 2001, pp. 269-270).
>
> "I don't really have any mental capacity for anything else because all of my mental energy, whatever mental energy that I actually have is going on to controlling, trying to control the drinking" (FAL-001, our interviewee).
>
> "I'd tried my hardest to control this problem – I was beginning to admit it was a problem – and I'd only lasted a matter of weeks. And I thought, 'I've done all I can. The doctors have done all they can, feeling a complete and utter desolation and despair" (Hugo in Addenbrooke, 2011, p. 59).

The choice account claims that these agents are maximising their reward (by definition). But reports of mental exhaustion, feeling desperate and utter desolation are an awkward fit with reward maximisation. These agents report struggling and failing, or barely managing to abstain with no energy for anything else. If they have decided on the way to maximise their reward it makes no

sense that they should have to *struggle* towards that goal given that every other option should be less appealing.

The choice account can explain distress and conflict if it stems from chronic preference reversal. Recall that preference reversal is caused by delay discounting. The agent unavoidably pursues the SS reward provided by drug-use because he lacks sufficient self-control skills to keep the true size of other LL rewards in sight. Throughout the process the agent is necessarily wholehearted; so when pursuing drug-use there is no thought that perhaps some other course of action would be better. Between attaining SS rewards the agent realises that pursuit of those SS rewards has undermined the LL rewards that they most prefer so only then does the agent feel regret and distress.

However, this description of distress interspersed with wholehearted, non-distressed pursuit of drug-use does not seem to accurately describe how addicts feel. They commonly remain distressed about their addiction even as they pursue drugs.

> "I desperately wanted to quit alcohol and drugs for many years. And I could continue to want to quit even as I was lifting the pipe to my mouth. What is going on inside someone who is having that experience is very hard to describe to people who don't have that experience, but if you take me seriously, I suppose it will be obvious to you that the self is divided or perhaps fragmented in that experience. Something is trying to make something else in the self stop; something is trying to make something else keep going…"[10] (Sartwell, 2008a).

Here Sartwell claims to feel distress and struggle *at the same time* as he pursues drug-use, not only in retrospective regret. The experience of acting while knowing that the action will fail to maximise reward should be impossible on the choice account. Such an experience should only occur in cases of involuntary behaviour.[11]

[10] Sartwell goes on to deny that this entails an essentialist position, "… putting it that way is not at all satisfactory because it makes it sound as though there is some unitary self that has been sledgehammered like glass, whereas I … hold that the coherent self, insofar as it is not just a delusion, or even insofar as it is a delusion, is an achievement reflecting a certain social/linguistic positionality" (ibid). I agree with him; the claim that the agent persists sufficiently to recognise their internal conflict does not entail that the agent has, or should have, an objective unity. It only entails that the agent is aware of their potential for a more unified existence.

[11] If the agent just is the result of the dominant sub-personal desires of the moment, as Ainslie claims, then the ambivalent self must present as two (significantly) different agents sequentially controlling the body rather than one

Flanagan also describes constant negative feelings about his drug-use and continuing to use in spite of those feelings and the knowledge that he was undermining LL rewards.

> "At that time … the anticipatory dread that I would go back out [drinking] was a constant companion. I had tried so many ways to stop and always failed. I had very little confidence. It was always just a matter of time. Knowing this felt worse than knowing I would die or even imagining that my own death would be very painful. Much worse. Dying is natural; enacting my spiritual death seemed necessary, inevitable, but thoroughly unnatural. Shameful" (Flanagan, 2011, p. 276). "I was … sometimes fighting off overwhelming craving for my drug, other times knowing, in some sense of 'know,' that my relationships with genuinely good people, my work, my life could, indeed, *would* be lost if I choose to use. And I'd still choose to use. Well, I'd use. That much was clear" (Flanagan, 2011, p. 277).

Sartwell is also absolutely clear that he was taking no reward from drug-use even as it undermines that which does provide reward:

> "…I hate alcohol, marijuana, cocaine, tobacco, methamphetamine, heroin. These stuffs or substances, … take away everything I have and everyone I love, every time. They are mindless, worthless, without value." "…the pleasure, in my experience, is fleeting and valueless, the dullness interminable, eventuating in excruciating pain and unredeemed death." (Sartwell, 2008b).

The association between drug-use and poor outcomes becomes so familiar that it is implausible that it is momentarily forgotten when the chance to use arises. As Flanagan says, he knew throughout, "in some sense," that, his drug-use would destroy what he valued. Similarly, Flanagan reports that his addicted desires persisted and tainted his enjoyment of LL rewards even while he recognised drug-use as a lesser reward and successfully resisted pursuing it.

> "I had the thought on the day of the birth of my firstborn that this was the most amazing day in my life for a host of reasons including the experience of that incredible precious

---

persisting agent experiencing a division within himself. Ainslie's model is, therefore, a closer fit for cases of dissociative identity disorder (DID). But DID is rare, affecting roughly 3% of addicted populations and 1.5% of the general population. Addicted populations do have a high prevalence of dissociative disorders in general, roughly 40% versus 5-10% in the general population (J. Johnson et al., 2006; Karadag et al., 2005; Ross et al., 1992). The experience of addiction and weak will in general seems distinct from these more extreme conditions for the very reason that the agent is aware of options that are in some sense better even as they act against them.

> feeling of unconditional love, *and* at the same time I thought that this event and those feelings were inconvenient because they were interfering with my drinking. Every addict in the room understood this; they had been there" (Flanagan, 2011, p. 273).[12]

It is not surprising that these agents have deeply ambivalent experiences because the detrimental diachronic trend of their drug-using behaviour is too obvious to be ignored. Again, this ambivalent conflict should not be possible on the choice account. Having identified a reward as being the greatest currently available, the agent should not reduce the size of that reward by simultaneously and pointlessly regretting mutually exclusive, lesser options.

Flanagan and Sartwell are clear counter-examples to the choice account. They are aware that there are better courses of action available even as they use and yet they continue to struggle with addiction. One might object that perhaps these agents lack some skill of global choice, test-case willpower, pre-commitment or prudential rule-following. However, all those skills are just meant to keep the greatest reward in view (or are fall-back options in anticipation that it will not be kept in view). On the choice account, keeping the greatest reward in view should be all it takes to pursue that reward but that just does not seem to be the case for some addicts.

The choice theorist has two ways to go when trying to explain Sartwell's and Flanagan's behaviour. First, the choice theorist could say they may have been too insensitive to the consequences. Heyman suggests that this would be his view: "... [if] only the threat of severe punishment brings drug use to a halt in addicts – then for practical purposes, drug use in addicts is involuntary" (Heyman, 2009, pp. 104-105). If we take this view, however, we are left with purely physiological models to explain these agents' addicted behaviour. Both agents' addicted behaviour appears too well integrated with higher cognitive processes to abandon person-level explanation so readily.[13] Furthermore, a satisfactory physiological explanation of behaviour like Sartwell's is still a long way off (if it can be found at all), so any person-level explanation, if it could be found, would be better than nothing.

[12] I assume here that Flanagan's involvement with the birth of the child was the pursuit of an LL reward, not an SS reward. That is, I assume he pursued it over time, perhaps the birth was planned, the relationship with the mother was maintained, he planned to be at the hospital, et cetera.

[13] See the introduction to the present work and Levy (2006, 2014) for a round-up of the evidence suggesting agential control in addiction. Of course, there may well be some cases where developing a sensitivity to the consequences is beyond the agent. It is just that this does not seem to be typical in addiction.

Second, the choice theorist might insist that Sartwell is actually fully in control of his drug-use and so must be lying or deluded when he claims to gain nothing from it. He might lie to himself and others to avoid the worst of the social opprobrium and reduce feelings of guilt. Deception of both kinds surely occur, but lives like Sartwell's during chronic drinking are so miserable it just isn't plausible that he could enjoy drinking more than the other paths open to him. As Sartwell says, "the pleasure, in my experience, is fleeting and valueless, the dullness interminable, eventuating in excruciating pain and unredeemed death." Sartwell narrowly avoids the unredeemed death, but in the face of interminable dullness and excruciating pain, *any* other path would be preferable, and as a well-educated man with some financial means Sartwell has other paths open to him. If the choice theorist is prepared to claim that even people in these miserable situations are maximising reward then the word 'reward' becomes trivial, merely a claim that the action is motivated in some way. Given that we already took their pursuit of drugs to qualify as action, we already knew that it was motivated in *some* way. Therefore, saying this behaviour maximises reward does no explanatory work; we still want to know *why* Sartwell does what he does (and so does he) (J. Kennett, 2013a{Kennett, 2013 #226)}.

The choice account traps itself in this false dichotomy by *defining* action as revealing preference. This definition was originally made by the economist Pareto. He was motivated to make this behaviourist move because it simplified economic theory by divorcing it from unobservable psychological processes.

> "Pareto's turn—the definition of utility as a quantity revealed by expressed preference equation—was an agreement on a convention for how to do economics, like the rules of tennis, or assuming away friction in physics" (Camerer, 2006, pp. 89-90).

Ainslie and Heyman adopt this convention but also back it up with reference to empirical data. This is the evidence for hyperbolic delay discounting, e.g. Ainslie (1975), Green et al. (2005), Green & Myerson (2004), Kirby (1997), and Mazur (2001). The problem with this is that empirical evidence only supports defeasible hypotheses not unfalsifiable principles. Even if the clear majority of people maximise reward in experimental conditions and outside the laboratory, this is not to say that people do it all the time. Sartwell appears to be an example that falsifies the hypothesis that all action is necessarily preference revealing or reward maximising. The claim that people act to maximise their reward may be a helpful rule of thumb, but to *define* action as the maximisation of

available reward forces us into a false dichotomy in cases of addiction like Sartwell's and prevents us getting any explanatory grip on the situation.

I now move from the first-hand experience of ongoing addiction to the experience of successful recovery. Here too we find reports that clash with the choice account. Recovered addicts rarely report just stopping once better options became more obvious. Rather they usually talk about a long, difficult struggle to regain control; many remain extremely vigilant lest they return to drug-use despite years of abstinence.[14] For example, Marc Lewis, a recovered addict, says:

> "In fact, most people beat addiction by working really hard at it. If only we could say the same about medical diseases!" (Lewis, 2012).

Empirical evidence suggests similar attitudes are widespread. A study of people who had met DSM-IV criteria for substance abuse or dependence for at least one year of their lives but had been abstinent for at least the last month were asked to define what recovery meant to them (Laudet, 2007, p. 249). Their definitions were coded for themes, the most common of which was "abstinence" (~40%) but there were a range of themes associated with taking responsibility for oneself: "a process of working on yourself" (11.2%), "self-improvement" (9%), "learning to live drug free" (8.3%), and "getting help" (5.1%). All these themes suggest that recovery involves a process that the person recovering must invest in and it is not merely a response to contingent circumstances.

AOD professionals and the general public tend to concur. Koski-Jännes et al. (2012) surveyed Finnish AOD professionals, clients, and the general population and found that all groups put a strong emphasis on the contribution of the addict to addiction onset and recovery.[15] As Koski-Jannes et al. note, this doesn't mean that the addict is seen to be fully responsible but just "...that the main responsibility for these problems lies with the individual, as preventing the problem or helping the person to solve it would be impossible without his or her contribution" (2012, p. 304).

At face value this seems to support the choice account which claims that the agent chooses to use or to abstain. That choice appears to be self-governed because it depends on whether the agent uses pre-commitments, test-case willpower, global choices, or uses prudential rules. But the problem

[14] It is possible to take this view without committing to a strong version of the moral model that claims drug-use is completely under the addict's control. This will become obvious when we consider degrees of self-governance in the planning account.

[15] Blomqvist (2004) reports a similar finding in Sweden.

for the choice account is that the agent's choice along with these techniques are all determined by something other than the agent. Both Ainslie and Heyman are averse to the role of an independent agent and by effacing the 'self' they leave no conceptual space for self-governance.

Ainslie is clear in his metaphysical commitment to a bottom-up picture where there are no executive inputs.

> "…If choice is determined in a marketplace of competing interests, 'she' [the agent] is just the resultant of their activities, and stable choice has to be achieved as it is in the kind of markets that don't have governors" (Ainslie, 2005, p. 642).

Therefore, the self-control techniques Ainslie refers to (pre-commitments and test-case willpower) are not wielded by the agent herself but by the dominant faction of sub-personal reward expectations. If such skills are not exhibited that is not a failure of self-governance but just indicative of a particularly competitive sub-agential environment where factions of reward expectations have yet to form with any stability.

Heyman avoids such clear metaphysical commitments yet he only focuses on extra-agential factors in addiction and recovery. He argues that addiction is caused by an 'addictive' *kind* of reward; these rewards are behaviourally 'toxic' and cause the agent to take a particularly synchronic perspective. Recovery and/or protection from this diachronic myopia comes in the form of prudential rules and global choice. Global choice is generally too cognitively demanding to overcome addiction, although it may happen to improve with age, hence the recovery rate seen in addicts in their late thirties. The availability and social enforcement of prudential rules depends on the social circles one happens to grow up in. Heyman does not say whether the agent can deliberately seek out or effortfully adopt beneficial prudential rules.[16] The overall picture of the choice account, then, is of an agent being swept along by contingencies: the sub-agential development of reward expectations, the availability of rewards (some of which might be toxic), the development of cognition capable of global choice, and the availability of prudential rules (some of which might protect them from addiction). Therefore, on the choice account, there is no way to distinguish people who effortfully overcome their addiction from those who just happen to stop using drugs because something better comes along. Everybody is assumed to be in the latter

[16] As we will see below, the planning account can help us understand why some rules are adopted and others are not with reference to how they fit within an existing network of intentions.

group and that clashes with the view of recovered addicts, AOD professionals and the general public.

At this point the choice theorist might object that, even if these cases of distressed addiction and effortful recovery occur, they are an insignificant minority. In response to this objection I put these cases in perspective, arguing that they are, in fact, the paradigmatic cases of addiction. The cases that the choice account can explain, in contrast, may fall into a broad category of addiction but are not particularly threatening to agent well-being.

Recall that roughly 80% of habitual drug-users stop using drugs without treatment by their late thirties. This does not entail that these people happily maximise reward throughout without skipping a beat. They may well suffer a period of distress about their drug-use and have to struggle to regain self-governance. It is hard to gauge what proportion of these people struggle and to what extent. Many people refuse to seek treatment for serious mental health issues and the stigma attached to addiction might encourage people to struggle alone (although ultimately successfully in these cases).

The other 20% of all habitual drug-users don't quit in their thirties. The choice account claims that these people all want to go on using drugs because it happens to maximise reward. However this is not easily reconciled with the fact that many of these people seek treatment. In fact it's common for long-term addicts to go through a succession of treatments over many years and still not recover. Heyman cites studies that indicate that addicts who engage in treatment are more likely to remain addicts in the long term and that drug-users who seek treatment have a much higher rate of comorbid mental disorder (~60%) than drug-users who don't seek treatment (~30%). Based on this evidence he suggests that,

> "…the persistence of addiction into middle age is due largely to the presence of additional psychiatric disorders. … Psychiatric impairment renders the drug experience relatively more valuable by undermining the ability to engage in and enjoy competing activities" (Heyman, 2009, p. 84).

But this explanation doesn't easily explain why all these people seek treatment for addiction. If drug-use is the way to maximise reward for the 60% of people in treatment with comorbid psychiatric disorders then why do they persist with treatment targeted at *reducing* drug-use? Perhaps all these patients and their clinicians are confusing drug-use with the real problem which

is the comorbid mental disorder. However it is not clear why choice theorists can lay claim to better understanding the problem caused by drug-use in these people's lives than patients and AOD professionals. In any case, 40% of people seeking treatment do not have comorbid psychiatric disorders so why are *they* seeking treatment that aims to reduce that drug-use?

Some of these people might be malingerers, maximising reward by taking drugs and getting some additional reward from treatment. Others periodically enter treatment to improve their short-term health while fully intending to continue drug-use later. Just as refusing to seek treatment does not entail the absence of a chronic distressing struggle with addiction neither does seeking treatment entail the presence of a chronic distressing struggle with addiction. However, it seems likely that the majority of people in treatment, whether they aim to abstain or reduce drug-use, are involved in a struggle to improve self-governance. Some manage to do so and eventually recover but others do not. This fits the reports of despairing struggle and effortful recovery described by addicts, ex-addicts, and clinicians described above. This explanation is unavailable to the choice theorist given their commitment to voluntary behaviour being reward maximising by definition.

It is hard to say exactly what proportion of the 80% who give up without treatment in their thirties and the 20% who don't give up in their thirties chronically struggle for self-governance and, at least for a time, knowingly pursue less than ideal rewards. But what does seem clear is that drug users who give up without a struggle or happily continue to use do not suffer a serious form of addiction. These people may represent the majority of habitual drug-users but they aren't *addicts* in anything but a trivial sense.[17] People who struggle and fail to control their drug-use are arguably the paradigmatic cases of addiction and, if so, the failure of the choice account to explain these cases is tantamount to a failure to explain addiction.

Summary

The choice account describes all choices as being the result of extra-agential forces. The resulting action explanation therefore effaces the experience and the role of the agent. Such distortions are not critical when agents appear to easily stop or start using drugs without an obvious struggle. Many cases of addiction, however, are characterised by a distressing struggle for greater control

[17] Such addicts are unlikely to meet many of the DSM-V criteria for substance use disorder (see the introduction to the thesis for those criteria).

over drug-use. The choice account cannot account for this struggle because it assumes that merely being consistently aware of greater available rewards should cause recovery. People struggling with addiction are consistently aware of rewards greater than drug-use yet still fail to recover. The choice theorist either has to accuse these agents of being deliberately deceptive (or self-deceived) or claim that their drug-use is non-intentional behaviour. The former approach ultimately trivialises 'reward' to mean merely 'motivated' while the latter approach does not fit with the sophisticated flexibility of addicted action. The addicts who struggle for control of drug-use are the people who most need help with *addiction* (rather than with socio-economic conditions) because merely having better alternatives available has not been sufficient. Therefore these are the addicts who we most need to understand but unfortunately the choice account cannot help us. I now turn to the planning account of agency to see what explanatory benefits it can provide.

Planning accounts of addiction

Revisiting planning accounts

Planning accounts differs from the choice account in three key ways. First, on planning accounts the agent cumulatively defines who she is by committing to plans and policies in accordance with the norms of practical reason. She governs herself when she acts in accordance with those normatively structured plans and policies. Importantly those commitments are underdetermined by what appears rewarding[18] and so the agent's act of commitment makes the difference. Second, the agent's normative commitments do not necessarily generate sufficient motivation; self-governance is not assured. From time to time the agent will find herself more motivated to pursue a goal inconsistent with her hierarchy of intentions even though she knows the more valued goal is available.[19] Third, the agent can overcome contrary desires by effortfully exerting her powers of agency. The greater the motivational gap between what the agent values and temptation the greater the effort required to achieve self-governance. That work is done synchronically by using muscle

[18] The agent's epistemic limitations render many potential goals incommensurable, e.g. becoming a professional footballer versus a concert pianist. However, if the agent is to achieve either one of those incommensurable goals she will have to commit to it over the others.

[19] As we saw in Chapter 2, Holton suggests that this problem may be aggravated in some cases of addiction by a neurological dissociation between what the agent 'wants' and what they 'like' or 'judge.' However he argues that this dissociation can, nevertheless, be overcome by the normal skills of self-governance – intentions and muscle model willpower. For present purposes we can assume that the motivational deficits that must be overcome for self-governance in the face of addiction are often large whatever exactly underpins their development.

model willpower and diachronically by better organising one's network of intentions so that they are means-ends coherent and ends-ends consistent. The greater the agent's planning and willpower skills the more easily she can avoid or overcome these motivational deficits. In contrast, a relative lack of these skills, or failure to exercise them, will see temptation more regularly undermining self-governance.

These key differences provide several ways to better explain the paradigm cases of addiction described above. Synchronic distress is caused by temptation overwhelming (or threatening to overwhelm) the agent's normatively endorsed intentions. In these cases the agent still acts but she lacks self-governance and undermines her diachronic identity. In addiction temptation is regularly challenging self-governance and so we would expect reports of distress both at the time of choosing to use drugs and during periods of greater control. Because self-governance can be effortfully maintained in these cases we would also expect the above reports of struggle in addiction and reports of successful long-term struggle in recovery.

Because the required struggle varies as a function of addictive desire strength, planning skill, and effort invested we would expect to see a range of severity in cases of addiction. At one end of the spectrum people recover without having to struggle much – action is relatively easily self-governed. At the other end of the spectrum, people fail to recover despite multiple cycles of treatment. Here self-governance frequently fails and we see cases of compelled action.[20] Compelled action occurs when the tempting desire is so strong, and/or the skills of self-governance are so lacking, that the amount of effort required for self-governance is beyond the agent. In the middle of the spectrum agents struggle for self-governance and manage to achieve it some of the time. They suffer occasional compulsion, where self-governance is clearly overwhelmed, and bouts of weak-will, where the necessary resources for self-governance were available but underutilised. Some of these people effortfully develop more consistent self-governance and move towards recovery; for others, efforts to improve self-governance wane, compulsion becomes more frequent and addiction more entrenched.[21]

As well as being able to explain the varying degrees of distress and struggle involved in addiction and recovery, the planning account can explain some cases of addiction in greater detail. The

[20] To be distinguished from compelled behavior that does not count as action, e.g. being paralyzed by fear. See Kennett (2001, pp. 155-159) for a more detail discussion of compelled action and how it shades into weakness of will.
[21] Of course, this is assuming that alternate, more highly valued lifestyles remain available to the agent.

planning account can explain the aspects of fatalism, unrealistic expectation, and chronic ambivalence we see in some cases of addiction.

Self-governance and means-ends coherence

Developing sufficient means-ends coherence in planning is a crucial element in self-governance for at least two reasons. First, hypothetical means-ends planning is necessary to *generate* possible goals; if one doesn't plan then one unnecessarily reduces the available options, leaving one relatively fatalistically attached to a limited set of goals. Second, if the agent imagines or is told of a possible goal, they have to be able to devise means to that end for it to be genuinely available to them. If they cannot devise such means then the goal is nothing more than a day dream.

Coherent means-ends planning is crucial for *generating* potentially rewarding goals. Some of the possible goals that agents take to be available are the result of hypothetical means-ends planning, for example, asking oneself, "What am I going to do this weekend?" The agent does not always begin with a rewarding goal already in mind and then begin to work out the means to that end (I deal with that case next); rather they start with their available means and work out what goals those means make available.[22] For example, one might wonder to oneself, 'I've got $500, two days free time, a car, some friends, a city full of entertainment and some national parks nearby; what can I do with that?' The better you are at means-ends planning, and the more effort you invest, the more rewarding the imagined goals can be. Unfortunately people suffering from addiction often no longer put in that effort.

> "I just don't bother having any aspirations or anything because I can't really do anything and I can't hold down a job because of my mental illnesses and my drug use. The same with study so anything ... if there was really anything that I did want to do, I wouldn't be able to do it anyway… I stay at home all day. I don't have any friends. I just sit at home and watch TV and listen to music and that all day" (FAL-001, our interviewee).
>
> "… the idea of me reducing methadone for a benefit that I can't really envisage anymore is hard…" (Interviewee 103, our pilot study).

[22] Typically means-ends planning will combine designing goals from means and working backwards from desirable goals to find means. I separate these two aspects of planning to show how addiction can undermine self-governance in both ways.

If you cannot, or do not, put effort into means-ends planning then you unnecessarily limit the rewards you believe possible to those more immediate, familiar rewards that require little planning (e.g. watching television). A lack of means-ends planning is therefore a form of fatalism. Of course many addicted people have lived with limited opportunities for a long time and experienced repeated failures to improve self-governance so we can understand their reluctance to bother planning. Nevertheless if they did put more effort into planning then they would better take advantage of opportunities if and when they arose. This is one of the more serious aspects of addiction - because it nips alternative sources of motivation in the bud goals that would be motivating and attainable are not even considered.

Planning does not have to proceed by first constructing goals from available means. Sometimes the agent thinks of a rewarding goal, or has one suggested to them, and *then* tries to develop means to that end. This is often the case for addicts who dream of an abstinent or more controlled life, or have the value of such a life impressed upon them by others, without yet knowing how to go about achieving that goal. If the agent can devise coherent means to the imagined end then that end is achievable should they commit to it. If the agent cannot devise coherent means to the end then they should realise that if they commit to that end they will be bound to fail; the goal is merely a nice dream. In other words, agents must follow the norm of means-ends coherence in planning if they are to self-govern.[23] People struggling with addiction often suffer from poor self-governance because they cannot follow this norm. For example, one of our interviewees, when asked if he had plans for the future said,

> "Not plans, things I'd like to happen, but not plans. … I'd like to meet someone and fall in love obviously. I'd like to finally have kids. I'd like to do all those normal things. I'd like all that normal stuff. But I can't quite see it happening" (Interviewee 101, our pilot study).

In this case Interviewee 101 seems to acknowledge that the things he would like to have happen are not going to happen because he has not or cannot create coherent means to those ends. If the agent has the potential to devise sufficient means but fails to do so, he over-readily abandons desirable goals and so fatalistically limits the scope of his self-governance. On the other hand, it is common in cases of addiction that people claim to have goals but perennially fail to enact them

[23] In Kantian terms, if you cannot will the means essential to an end then you do not truly will the end.

perhaps because they never sufficiently arrange the means. One of our interviewees recalls that his attitude to changing things in his life used to lack a focus on planning.

> "I used to be much flightier and much more fanciful and expect all of sudden (clicks fingers) things just to snap into place and that'd be amazing (chuckling), but sadly, that hasn't happened" (Interviewee 106, our pilot study).

So, through poor skill or effort in developing means-ends coherence the addict can limit their self-governance by: failing to see possible motivating goals, over-readily abandoning possible goals that they would find motivating, and pursuing goals that they will fail to achieve through limited planning.

Addicts suffering from poor self-governance would be expected to benefit from better means-ends planning and, in fact, many forms of treatment aim to improve means-ends planning. The SMART recovery system,[24] for example, provides a variety of tools one of which is a "Change Plan Worksheet" that helps to systematise the agent's planning with prompts like, "The steps I plan to take in changing are… I will know my plan is working if… Some things that could interfere with my plan are…" Other examples include detailed relapse prevention plans that help the addict notice the mental processes in the very early stages of relapse and act to regain control early. By formalising these means-ends planning processes around addiction it is hoped that the addict's recovery plans will be provided greater diachronic stability. Ideally the process will become habitual and generalise to other goals as they rebuild their life.

The planning account can explain some cases of addiction as failures of self-governance due to poor skill or low effort in means-ends planning; it distinguishes these cases from those who can plan well but genuinely lack opportunities. Of those agents struggling with self-governance it can further distinguish those who have become fatalistic and fail to pursue goals that are within their means from those who are chronically frustrated by pursuing goals that they fail to support with sufficient means. The planning account can also explain cases of recovery where self-governance is improved by improving skills of means-ends planning or helping the agent rediscover that their efforts of means-ends planning will be worthwhile.

[24] SMART stands for Self-Management and Recovery Training – the SMART recovery group operates across he US providing materials and group meetings to support recovery. It attempts to work from a scientific rather than spiritual basis providing an alternative to 12-Step programs.

The choice account, in contrast, reduces all these forms of habitual drug-use to the one category – those who habitually use do so because they are aware of no better rewards. The causal relationship between reward expectations and action is one-way; the reward landscape is fixed *prior* to anything the agent does and it determines action. As a result, the choice account is unable to describe self-governance. The choice account's agent will only bother planning if the expected reward is great enough to make the required planning effort seem worthwhile.[25] Choice theorists assume that the value of all rewards and the costs in accessing them are set externally, the agent has no control over those. But, as we have seen, agents will not know which rewards are available, what size they may have, and what costs they are likely to entail *until* they construct some hypothetical plans. The self-governing agent effectively plays a role in constructing the rewards that are available to him by investing effort in planning before he knows if and how it will be compensated for. Dynamic variation in reward size and expectability *as a function of agential planning effort* are impossible when one takes these to be the objective pre-established underpinnings of motivation.

The choice theorist could argue that social pressure initially encourages the agent to put a certain amount of speculative effort into planning, prior to knowing the effort will be compensated. Because this effort would pay off on average, the agent would then engage in the degree of hypothetical planning that had been sufficiently compensated in the past. But the problem with this is that the planning effort would always be determined by the history of how much reward such a policy yielded on average. This is problematic for developing unique diachronic goals, such as recovering from addiction, because the average benefit of planning in general is not necessarily a good guide to whether one should put in more planning effort in this unique case. Neither is it obvious how the agent might use their past experience to calculate a probability that their planning efforts might be compensated for in this unique case. In any case, the choice theorist's account clashes with what Interviewee 103 says above, "… the idea of me reducing methadone for a benefit that I can't really envisage anymore is hard…" The most natural interpretation of this statement is that 'hard' refers to a large amount of effort. He knows he *can* envisage a rewarding sober future for himself, but because it takes so much effort, he is relatively disconnected from that future. The

[25] Although, as we saw in Chapter 2, Ainslie rules out planning because he believes that all decisions are made in the moment, not in advance. Heyman doesn't mention where he stands on this. The problem is that when decisions are made in advance for planning the success of the plan depends on the agent committing to those decisions. This commitment entails sometimes acting on the decision despite a contrary change in greatest reward expectation. But if the agent commits to plans despite the incidence of alternate greater reward expectations then we have gone most of the way to adopting Bratman's view where the agent's commitments define who they are and what counts as self-governed action rather than the oscillation of expected reward.

choice account leaves no conceptual room for the agent to put in a little more planning effort in hope of regaining self-governance or to be a little lazy and fatalistically surrender some self-governance.

Self-governance and ends-ends consistency

There are a range of addiction-related phenomena that we can better understand if we take the planning theorist's view that ends-ends consistency is an achievement of self-governance. First, we can understand how it is that addicts can struggle with synchronic ambivalence – they simultaneously hold plans or policies that are relatively ends-ends inconsistent. Second, in some cases of addiction people chronically fail to resolve their ambivalence despite being consistently aware of its damaging effect on their lives. Third, many recovered addicts continue to place a value on drug-use despite knowing that they should not go back to using. Fourth, some treatment approaches aim to resolve chronic ambivalence and this appears to help some addicts recover. Finally, the need for ends-ends consistency among one's intentions can help us understand why agents cannot just adopt protective prudential rules. The choice account can only offer limited understanding of these phenomena.

On the planning account, the better one meets the norm of ends-ends consistency in one's goals the better one self-governs. Self-governance is undermined by failing to follow this norm because ends-ends inconsistency tends to reduce the chance of achieving either end. The agent may not realise immediately that they are being ends-ends inconsistent but once they do, they experience ambivalence. Ambivalence is the result of valuing two inconsistent goals, recognising that, but being unable to prioritise one over the other.

Ambivalence is made possible on the planning account by the agent's epistemic limitations and the fact that each end has its own unique characteristics. Epistemic limitations prevent the agent from being certain about which ends will be the most rewarding – yet they typically have to commit in advance to achieve those goals. Unique characteristics entail that the agent is rarely comparing like with like; the rewards of one end will never entirely replace another. Sky-diving, heroin use, becoming a father, retiring, going to Spain, and so on, all provide rewarding experiences that are fundamentally distinct from each other. No number of great trips to Spain will provide the kind of reward associated with becoming a father and vice versa. Incommensurability entails that reward

cannot be maximised in any straightforward way. Therefore agents can understandably find a decision difficult when choosing between mutually exclusive but irreplaceable rewards.

This picture is complicated by the fact that ends are not completely mutually exclusive or completely compatible. All agents maintain plans and policies that come into occasional conflict, e.g. obligations at work encroach on obligations to one's family. These conflicts can usually be managed by planning and forming a normative hierarchy. One can prioritise, e.g. putting family before work, and plan carefully, e.g. fitting one's workload around the family meal, taking the kids to soccer, etc. Of course events sometimes conspire to create unavoidable conflicts so that the agent fails in one commitment or the other, e.g. unexpected work crucial to keeping one's job happens to clash with a child's birthday party. The more the agent is torn between two valued ends the less she can properly meet both standards and the more she suffers from ambivalence. When suffering from ambivalence agents can vow to plan better and therefore hope to maintain the same ends. Habitual drug-use is infamous for being difficult to balance with other ends. Nevertheless some people succeed through significant organisational effort and are sometimes labelled "high functioning" addicts. Our interviewee, Interviewee 104, for example, maintained a heroin addiction for 20 years, holding down jobs throughout and never having trouble with the police:

> "…there were lots of things I was responsible for, I mean work, Sam, family. It [heroin use] doesn't take away your responsibilities, it sometimes makes meeting those responsibilities very tricky, time-wise. Because the ugliness of the addiction makes the addiction time, the thing like, you know whenever that time is that you have to get it for the day, that ... you've got to make that time up, you've got to find that time to do that. And that, that's an ugly situation because invariably, you know if you're time conscious, you've got to lie to make ... to find that time to go and do those things are hell bent in your head to do" (Interviewee 104, our pilot study).

Such 'high-functioning' addicts do not obviously suffer more ambivalence than some non-addicts who, say, balance high-flying careers with parenting.[26] So not all addicts suffer from poor self-governance caused by ambivalence but many addicts do.

[26] But given the required planning effort, it is common for such addicts to eventually tire of maintaining drug-use; the value they put on drug-use decreases. They then either manage to recover by abandoning drug-use as an end or they try to abandon it as an end and find they cannot. In the former case they appear to maximise their rewards throughout and so roughly fit the choice account. The latter case, however, is a challenge for both the choice and normative account

To see the explanatory benefits of the planning account we need to consider cases of addiction where chronic ambivalence is part of the problem. Bernadette (in Addenbrooke, 2011) is a classic example of an ambivalent alcoholic. Bernadette phones an AA helpline a few times concerned at her drinking and they encourage her to come to meetings but she would stop drinking for a few days and so think there was no need to attend. She starts to recognise this pattern, "I want to stop, but I can't stop it. I can stop it for days but then I'm back on it" (Addenbrooke, 2011, p. 32). This leads her to get treatment but she doesn't sincerely engage with it and ultimately fails to follow the path of recovery:

> "…in a way got a bit of hope from [the AA meetings]. But then I stopped going and when I started drinking again I was referred to a residential centre. They took me in there. Their treatment is based on the Minnesota method. It's the most intense thing I've ever been through. But I bullshitted my way through that – I've got to admit it now. I was very clever, because I didn't open up. I knew exactly what they wanted out of me, and I gave it to them. I did dry out, and I stayed dry for another three months, till I went to my home town, then I picked up a drink. This was two Christmases ago. I know in myself, I can stop drinking if I go back to AA, but… It's like AA says, you have to come to your rock bottom. Now luckily, so far I've stayed off rock bottom. … But I have caused riots with people, and I'm a nasty piece of bloody work when I've been drinking. I don't like that side of me, but as yet I haven't hit my rock bottom. It's bound to come I know it is. … It's like being two people really, being an addict, isn't it? One side of you wants the addict side, and the other side wants the side that isn't. … You've got one side of you that knows what's right and what's wrong, and you do know, everybody knows, underneath it all, and yet you've got the side of you that behaves badly" (Addenbrooke, 2011, pp. 33-34).

Bernadette has tried for years to maintain her alcohol use with her other values but her drinking consistently undermines her other values; she cannot sufficiently improve her planning to manage the conflict.

The planning account can explain Bernadette's half-hearted, insincere engagement in treatment as being the result of her continuing to value drinking throughout. She seems to actually value drinking rather than experience it as a contrary desire because she says, "I can stop drinking if I go

---

because such addicts continue to enact drug-using intentions despite no longer valuing drug-use. These people present a challenge to the planning account; I return to this issue below and expand on it in the next chapter.

back to AA, but… It's like AA says, you have to come to your rock bottom. Now luckily, so far I've stayed off rock bottom." So she acknowledges an easily accessible path to abstinence (albeit a difficult path to stay on), one she could initiate when not under temptation, but she will not take the first steps on that path presumably in order to continue drinking.

The choice account might hope to explain Bernadette's behaviour in terms of chronic preference reversal – her inconsistency is the result of sequentially failing to see the real size of the abstinence reward. Bernadette's actions are therefore sequentially inconsistent but each is aimed at maximising perceived reward at the time. But being insincere in treatment is an implausible way to maximise reward; she should either engage with treatment wholeheartedly or not bother at all.[27] Similarly, even though she is in a period of drinking at the time of the interview, she worries about being a 'nasty piece of work' and inevitably sliding towards rock bottom. Presumably she would enjoy her drinking more if she gave up these other concerns. So the choice account cannot explain why she would reduce the reward of drinking by dwelling on the costs and risks entailed by her choice. The planning account can accommodate these ambivalent feelings by postulating that she simultaneously values inconsistent ends: drinking, meeting certain social standards, and avoiding rock bottom.

The choice account also has to assume that Bernadette is ignorant of the diachronic trajectory her life has been taking throughout her preference reversal. But because the pattern of inconsistency has been going so long it could not be more obvious that reward maximisation requires dropping one of the mutually exclusive ends. Whether Bernadette was drinking or abstaining her recent action would be framed by a background awareness of her inconsistency. Given the obvious costs of this inconsistency, in order to maximise reward she should be prepared to do whatever it takes to resolve this chronic preference reversal, e.g. adopt highly inflexible pre-commitments. But she does not do what would obviously better maximise reward, her ambivalence just goes on.

The planning account can give us a way of understanding this chronic failure of ends-ends consistency. Because rewards are incommensurable it is possible that drinking provides Bernadette with something she values highly that cannot be fully replaced by any other end. Even the perfectly abstinent life would lack that something. So even though she might agree that all-things-considered, not drinking is preferable to drinking, she might still be unable to face life without that

[27] There may, of course, have been various other tempting rewards while in treatment that disrupted her goal of recovery, such as a feeling of power at manipulating the social workers or avoiding having to talk about painful truths.

specific reward. The incommensurable aspects to rewards make such choices *difficult* choices.[28][29] Choices are easy when rewards are commensurable; people don't agonise over choosing between two pieces of chocolate and one piece of chocolate or two sky-dives and one sky-dive. The choice account considers all rewards to be commensurable and so cannot distinguish difficult choices from easy choices – all choices should be easy. The choice theorist, therefore, just cannot understand why people like Bernadette hesitate in ambivalence.

The planning account can also explain why so many *recovered* addicts continue to value drug-use (albeit demoted sufficiently in their hierarchy of intentions that they continue to abstain). This does not make sense on the choice theory since a reward acknowledged to provide less reward should be of no interest. One of our interviewees, for example, is clear that drug-use remains more than just the occasional target of a desire contrary to self-governing intentions:

> "…don't get me wrong, I love using mate. If I could use successfully I would. I'd still be using. I love using. I just don't like the shit that comes with it" (R50, our interviewee).

The continued value placed on drug-use makes sense on the planning account because the unique, incommensurable character of drug-use remains. Of course it is not advantageous to dwell on the incommensurable goals you have demoted because relapse is probably more likely than for someone who completely repudiates drug-use. However, if our interviewee maintains his abstinence then the valued character of drug-use should begin to fade from memory and new values should come to occupy his attention. This pining or regret for the drug-use one can no longer allow oneself makes less sense on the choice account. Why waste one's time thinking of lesser rewards when you could be actively pursuing greater rewards? The choice theorist might claim that our interviewee is now allowing himself the smaller, compatible reward of reminiscing about drug-use. But it seems more plausible that the agent experiences this as a loss rather than as a reward.

Given that paradigmatic cases of addiction involve ambivalent agents, it is not surprising that we see forms of treatment that aim to improve ends-ends consistency (both in general and in attempts

[28] Although there is still something unconvincing about the planning account response. Can the unique value of drinking really be so good that it is worth continuing with a miserable ambivalent life? I return to the issue in the next chapter where we see that ambivalence can become entrenched in the agent's self-narrative so that to give up one of the conflicting values would be to give up part of who one is.

[29] Research by Wang et al.(2010) suggests another factor might also be in play. They found that the size of the reward given up when making a choice was proportional to a subsequent reduction in willpower. Because ambivalent agents are often faced with choices where one highly valued end or another is given up, their willpower may be sapped more quickly than less ambivalent agents.

to demote the value placed on drug-use in particular). This is an obvious approach on the planning account since ambivalence undermines self-governance. The private rehabilitation network "We Do Recover,"[30] for example, claim that they aim to quickly identify degrees of ambivalence and work to exclude drug-use from the addict's ends through counselling.

> "Addicted and alcoholic patients usually enter treatment with varying degrees of ambivalence about their dedication to treatment and long term recovery. It's important that motivation and ambivalence be explored early in the rehab process. ... Ambivalence is a conflicting craving to do, and not to do the very same thing. ... Moving towards a point of acceptance where the alcoholic is acutely aware of their illness and accepts the changes they need to make to enter stable recovery, can be a difficult journey. It's a transition best achieved with professional addictions counselling" (We Do Recover, 2010).

As an example of training skills of ends-ends consistency in general, Treatment Planning MATRS[31] (Stilen et al., 2007, Module 3) aims to help addicts identify and prioritise problems. The module suggests Maslow's Hierarchy of Needs as a model for prioritisation. The idea is to develop a stable base from which to address problems further down the priority list. This should help addicts who are being overwhelmed by the sheer number of problems they face and who are failing to solve problems by poorly prioritising them.

One of our interviewees mentions her improvement in ends-ends consistency and how this has helped stabilise her recovery:

> "Yeah. I work out what I need … what needs to be done for me first. Like if someone rings me up from my family and needs to go somewhere and needs a lift or whatever else, if I've got an appointment, well, no, that comes first. I don't let anything come between my drug counselling anymore. I get my priorities straighter and when my children need me, I'm there" (FHE-045, our interviewee).

Notice that she claims that her improvement in ends-ends consistency is the result of an effortful exercise of self-governance skill. She *works out* what she needs; she sees it as her responsibility to get her 'priorities straighter.' Previously she did not put in so much effort to ends-ends consistency,

[30] A private network of rehabilitation clinics in the UK, South Africa and Thailand.
[31] A treatment initiative stemming from cooperation between the National Institute on Drug Abuse (NIDA) and the Substance Abuse and Mental Health Services Administration's Center for Substance Abuse Treatment (SAMHSA/CSAT).

as a result her organisational skills were out of practice or undeveloped and that lack of self-governance contributed to her ongoing addiction. The choice account, in contrast, makes no room for agential effort to overcome ambivalence. If the agent's sub-personal motivations continue to cause preference reversal despite the agent being aware of their history of preference reversal, then that ambivalent action reveals the agent's, admittedly unusual, preferences. We might judge that such a life could not really be rewarding and so treat the agent paternalistically. We could change the reward landscape making drugs harder to access or other rewards greater until we broke the cycle of preference reversal. But just as with using the approach to improve means-ends coherence we would expect ends-ends inconsistency to return as soon as we removed the artificial reward adjustments.

Finally, the need for ends-ends consistency among one's intentions can help us understand why just knowing of beneficial prudential rules (i.e. policies) does not necessarily result in recovery. First, the agent might not be easily able to adopt a protective rule because it is currently ends-ends inconsistent with their existing network of plans and policies. For example, even if one knows that certain religious prudential rules protect against addiction, one might be unable to adopt those intentions because they will create ends-ends inconsistencies. This would be the case if, say, the agent has a central self-governing policy to only form beliefs on the basis of convincing evidence. Second, an ambivalent agent might adopt a prudential rule but find that it is only ends-ends consistent with their intentions related to abstention and not their intentions related to drug-use. In these cases the agent's ongoing ambivalence will tend to undermine the beneficial effects of the prudential rule.

In summary, because the planning account acknowledges concurrent conflicting values it can describe ambivalent addicts who unnecessarily undermine the rewards of both recovery *and* of drug-use even as they know they are doing it. They struggle chronically to make the difficult decision to drop either of their inconsistent ends because each end continues to promise an incommensurable value that cannot be replaced. This ambivalent struggle may go on for years and characterises many cases of addiction.[32] This also explains why those who make the hard decision to abstain continue to pine after what they have lost. Training a normative skill of overcoming ambivalence should be superfluous if the choice account is correct yet AOD professionals and their

[32] Although, as I investigate in the next chapter, it's likely that in many cases of ambivalent addiction the agent is struggling not so much with organising plans and policies but with more fundamental issues of who they are. In these cases a narrative focus can be helpful.

clients clearly value it as a treatment modality. Awareness of protective prudential rules does not necessarily help because they are ends-ends inconsistent with central self-governing intentions or because they do nothing to resolve ambivalence.

Because the choice account defines all action as wholehearted, it cannot make sense of the half-hearted efforts at recovery and drug-use that characterise ambivalent addicts. Awareness of this synchronic clash in rewards should resolve in favour of the most rewarding end but clearly, in many cases of addiction, it does not. Over the longer-term, awareness of one's chronic pursuit of mutually exclusive ends should result in more extreme attempts by sub-personal systems to maximise reward, such as by adopting strong pre-commitments. Yet many ambivalent addicts fail to resolve their ends-ends inconsistency when it clearly appears necessary to maximise reward.

Issues for the planning account

Despite clear improvements over the choice account, the planning account still suffers from some explanatory shortcomings. First, people with highly developed planning skills can continue to suffer from addiction despite those skills. Second, an improvement in planning skills does not always seem to be sufficient to recover from addiction. To illustrate these explanatory shortcomings I draw on the cases of Interviewee 104 and Flanagan.

Flanagan disvalued his drinking at the time and continues to disvalue it now that he is no longer drinking:

> "I wanted not to use. I expressed to myself, my loved ones, and mental health professionals a sincere desire not to use, and I used." (Flanagan, 2011, p. 276).

Interviewee 104 continues to have occasional relapses and he used more heavily in the past. He disvalues his past and present use:

> "…I mean this whole drug thing has been one big nightmare, really. I mean there's a lot of ... there's every aspect of it that I just wish would've never happened." (Interviewee 104, our pilot study).

Their negative evaluations of their drug-use implies that they knew of other more highly valued ways they would prefer to live. Indeed, both had access to a more highly valued lifestyle; neither was particularly limited by socio-economic circumstances. Both men are in contact with people

who love them. Flanagan is well educated and has managed to hold academic jobs involving research, high quality writing, and teaching. Interviewee 104 continues to struggle with addiction but still holds down a job and owns property. So why did these agents not pursue the lives they valued more highly?

A planning theorist could argue, just as the choice theorist could, that these men actually value drug-use that highly; they are lying to us or self-deceived when they claim to value other lifestyles more highly. However, this argument is unconvincing in these cases because their evaluations of their drug-use are settled, the costs they suffer from drug-use are obvious, and there is little benefit in lying. A more plausible explanation for the planning theorist is to say that Flanagan didn't, and Interviewee 104 does not, have the skills of self-governance required. In addicted action temptation overwhelms their self-governance through judgment shift and/or a dissociation of their motivational systems from their judging systems. They fail to form the necessary intentions to stabilise their behaviour through periods of temptation. Perhaps they form some intentions toward recovery but with insufficient means-ends coherence or ends-ends consistency. Similarly, they fail to devise or implement pre-commitment strategies. They use their muscle model willpower to resist drug-use but, without the protection of intentions or pre-commitments, that willpower is inevitably exhausted by the constancy of the temptation.

Although this explanation might be true it does not sit so easily with these cases. Interviewee 104 continues to exhibit planning skills concurrently with his drug use. He reports,

> "[drug-use is] just time-consuming, you know as far as the methadone programs, the in-house patients, … I might not necessarily have to use three times a day, but I would still use ... have to use once a day. So I could control it to a degree, but I couldn't just say no, which used to shit me … And in some stupid way I was able to budget it. But you know it still had effects, and I still spent money that I probably could've used in other ways, but I never actually lost everything through it. I lost enough, but yeah, I wish sometimes I could've understood how I could've kept it to a minimum, and not just stopped, you know like why, but anyway … I suppose there's a lot of other things that take up my day other than just drugs, whereas some people I know who are in this situation, it has their ... their focus 24/7. Whereas for me, there's just too many other things, like I love my son, I like work, I've never been unemployed, I've bought property…"

So, Interviewee 104 must have been exercising good means-ends coherence in planning to hold down jobs and develop a property portfolio. He also had to work hard to achieve ends-ends consistency between drug-use, financial commitments and family responsibilities. His problem was that he did not actually value drug-use even as he integrated it with his other ends. He would have preferred to attribute less time to drug-use and more time to his family but he could not make that change consistently. Yet on the planning account he doesn't display any clear impairment of agency; because his plans and policies of drug-use are relatively ends-ends consistent and means-ends coherent. His drug-use intentions obey the norms of practical reason so they have agential authority and he self-governs when he enacts them. This is counter-intuitive; if the agent consistently reports that he disvalues an aspect of his life, it seems he lacks self-governance to some degree, no matter how well integrated that aspect is. In the next chapter I argue that the narrative account can explain why agents like Interviewee 104 lack self-governance and how they can regain it.

Flanagan's case is slightly different. He also had a period where his drinking was somewhat balanced with the rest of his life, however, ultimately his drinking came to completely dominate his life. Why were his sophisticated planning skills insufficiently domain-general to control his drinking? Upon seeing his values increasingly threatened by drinking he should have been motivated to set up the intentions and pre-commitments that would protect those values. Of course, the temptation to drink may have been so much more powerful than desires in other domains that his domain-general planning skills were insufficient to counter it. However, given that his control over drinking decreased over time, we have to add that his desire to drink must have either increased over time or progressively undermined his planning skills in that domain. The increasing dissociation of motivational systems from judging systems could explain this so the planning theory has a plausible explanation for the *onset* of Flanagan's addiction. However, it is Flanagan's recovery that is harder to explain using the resources of the planning account.

Flanagan recovers and so he must have found a way to counter the temptation to drink.[33] As it happens he *did* learn a helpful domain-specific planning skill. He learned that he had to focus on avoiding the first drink. He claims he had been trying for years to design policies that would allow

[33] Alternatively, the desire to drink may have naturally faded away but Flanagan's effortful struggle for recovery speaks against that.

moderate drinking and that had been the problem (2011, 290). However, tellingly, Flanagan doesn't claim that this new planning knowledge alone was enough for his recovery or recovery in general.

> "You are motivated to stop. You think, all things considered, that it is best for you and others that you stop. But you are not *yet* at the point where you can reliably negotiate the wee zone of control between yourself and the first drink or drug. Normally there are two favoured options. Sit with fellow addicts, talk and let them help you not use and/or go into rehab where the same thing will happen, that is, a social group will help you overcome you own seeming powerlessness (in rehab there will be drugs usually benzos to get you through – benzos are used even to get you off benzos) (Flanagan, 2011, p. 290).

Flanagan emphasises that merely knowing that he had to avoid the first drink was not enough to abstain. He suggests that you also need to sit and talk with other people to help overcome your apparent powerlessness. This is somewhat mysterious on the planning account because, for a skilled and motivated planner like Flanagan, it seems that this knowledge would be sufficient to then construct intentions and pre-commitments to abstain. Although Bratman does not emphasise it, perhaps social interactions can also improve planning. However, as I argue in the next chapter, the narrative account suggests that something more holistic is happening in these social interactions. The agent is not just learning to plan better but they are reconfiguring significant aspects of their self-concept. These reconfigurations of self-concept are likely to influence recovery in addition to any changes in planning skill.

Conclusion

We are alerted to the inadequacies of the choice account when we notice the significant numbers of addicts who repeatedly seek treatment and complain of experiencing a distressing conflict. These people are not maximising their available reward. The scale of their failure is such that it is implausible that they periodically forget those greater rewards and use drugs wholeheartedly. Addicts' first-hand accounts bear this out. They report knowing of better alternatives, feelings of distress, and struggling for self-governance even as they voluntarily pursue drug-use. Furthermore, recovered addicts and clinicians describe recovery as a prolonged struggle not, as the choice account would have it, a cool, calculated adjustment to changed circumstances. The possibility of effortful struggle cannot be accommodated by the choice account which claims that all

motivational conflicts are resolved at the sub-personal level. The choice account, therefore, has nothing helpful to say about these chronic, distressing cases of addiction. This is problematic because these people need the most help and put the greatest burden on society. They are arguably the paradigmatic cases of addiction and certainly the cases that we most need to understand.

The planning account is better geared to explain paradigmatic addiction because it makes room for effortful self-governance. Diachronically, effort is required to develop and enact a network of intentions that are means-ends coherent and ends-ends consistent. Synchronically, the agent should try to maintain their normatively endorsed intentions in the face of temptation by using muscle-model willpower. If the agent's efforts fail then they will be more likely to enact an intention that breaks with these norms and thus exhibit a lesser degree of self-governance. Paradigmatic cases of addiction are a result of a lack of skill and/or effort in these aspects of self-governance. The planning account can also distinguish a variety of more specific failures of self-governance when we look more closely at the skills of developing means-ends planning and ends-ends consistency.

A lack of skill or effort in means-ends planning can have three effects. First, an agent who does not construct hypothetical ends from their means will limit their awareness of potential motivating goals. Second, they may be aware of appealing goals but incorrectly judge that these goals are beyond their means. Poor means-ends planning therefore entails a form of fatalism; the agent unnecessarily limits themselves to basic routines they know will work. Third, the agent might falsely believe that a goal is within their means when it is not, or, at least, will not be unless more extensive means are constructed. As we saw, reports from addicts that suggest that all these failures of self-governance are relevant in various cases of addiction.

Agents may also fail to develop sufficient ends-ends consistency in their intentions; this can explain a range of phenomena related to ambivalence. First, if an addict simultaneously holds inconsistent plans and policies then they will frequently feel ambivalent – simultaneously valuing and disvaluing their actions. Second, ends-ends inconsistency will tend to undermine the achievement of all the conflicting ends. But we can understand why the agent might chronically fail to resolve the conflict because each end provides a certain kind of reward that no other ends provides. For the same reason we can understand why many recovered addicts continue to place a value on drug-use despite knowing that they should not go back to using. Fourth, because ambivalence can be symptomatic of a struggle to achieve self-governance it makes sense that some treatment approaches aim to help resolve chronic ambivalence. Fifth, the need for ends-ends consistency can

explain why mere awareness of protective prudential rules rarely helps recovery on its own. The agent may find the rule ends-ends inconsistent with their existing intentions or the rule may be ineffective given the incumbent ambivalence.

However, as I discussed at the end of the chapter there are a sub-set of paradigm cases of addiction that pose a challenge to the planning account. These people want to stop using drugs, have better alternatives open to them, *and* appear to have the necessary skills of self-governance. So why do people like Interviewee 104 still fail to recover? The planning account has to claim that people like Interviewee 104 are self-governing when he clearly lacks some self-governance – he wants to abstain from drugs but cannot. Similarly people like Flanagan insist on further treatment even when they appear to have attained the necessary planning skill to recover. In the final chapter I argue that consideration of the agent's self-narrative can help us understand such cases and it can also add depth to our understanding of the cases that the planning account can explain.

Chapter 6: Planning accounts versus the narrative account

Introduction

In this, the final chapter, I illustrate how the narrative account of agency builds on our understanding of the agent's role in addiction and recovery. Sometimes agents have developed self-narratives that entrench addiction and hinder recovery. In these cases addicted agents cannot just reform their intentions to regain self-governance; they must rework their self-narratives. My aim here is to detail exactly how self-narratives can entrench addiction and how self-governance can be recovered through re-narration.

In the last chapter we saw that the planning account improves on the choice account in understanding addiction and recovery. The choice account tries to explain all behaviour as the result of extra-agential factors, such as social pressure, available rewards, and sub-agential cognitive processes. These factors are certainly important but, by focussing on them exclusively, the choice theorist eliminates the possibility of self-governance. The planning account, on the other hand, makes room for self-governance. Greater self-governance is achieved by successful agential efforts in pursuing norms of practical reason. This enables the planning account to describe recovery from addiction as a (re)development of self-governance in which the agent plays a crucial role. We finished the chapter with a problem, however. Many addicts have developed and continue to enact a network of intentions targeting drug-use that meet norms of practical reason. However, they now want to replace those intentions with something else but find that they cannot. This creates two related problems of the planning account. The planning account describes these addicts as self-governing when they appear to *lack* self-governance. Furthermore, the planning account cannot say what more these people need to recover from addiction nor why it is they do occasionally manage to recover.

The narrative account can explain these cases of ongoing addiction and the timing of their occasional recoveries. Self-narrative effects can entrench drug-use intentions and prevent the development of recovery-directed intentions. Self-governed action is not, therefore, straightforwardly defined by creation and pursuit of intentions. The agent's evaluations can guide self-governance even when those evaluations are not (yet) grounded in a network of intentions. Self-governance requires that agents enact their positively evaluated, aspirational future. To do so agents needs to develop new self-narrative threads connecting aspects of their existing narrative with that future.

The chapter proceeds as follows. In the first section, I begin by reviewing the main points from Chapter 4 showing how the narrative account builds on the planning account. I then build on this account by describing the 'multiple thread' view of self-narrative. According to this view, agents do not create a single, all-encompassing self-narrative but a number of overlapping semi-independent narrative threads of varying temporal length. This level of detail is relevant because the thread structures of addicts' self-narratives tend to undermine self-governance while recovery involves developing neglected narrative threads and building new narrative threads.

In the second section I present my central arguments for preferring a narrative account of agency over a planning account when trying to understand addiction. My arguments are built on examples of addicts who either recover or fail to recover because of their self-narration not because of their normative planning. With these examples I distinguish several narrative-specific effects. Existing self-narratives can entrench narrative projections of addiction and shape the narrative's constituents so that they undermine recovery. Reinterpretation of self-narrative can reverse these two effects, promoting recovery. Self-narratives focussed almost exclusively on drug-use, treatment and recovery will only ground fragile self-governance. Greater self-governance will be achieved through the development of robust narratives with multiple narrative foci. Ends-ends inconsistent behaviour can become chronically entrenched where the narrative threads developed around positive drug experiences are kept largely independent from those developed around negative drug experiences. The first step in overcoming ends-ends inconsistency is forming narrative links between these semi-independent threads. Because self-narratives are co-authored, recovery is encouraged by interacting with co-authors sympathetic to the cause.

Finally I analyse some of the implications of the narrative account for the treatment of addiction. I argue that all clinicians would benefit from narrative awareness. Narrative awareness can help develop the therapeutic relationship and better match treatment modalities to the client's nascent recovery narrative. There is also reason to believe that many clients would benefit from treatment that explicitly focuses on the content of their self-narrative. For some clients it may even be essential. I consider one form of such treatment, 'Narrative Therapy.'

The distinction between narrative and planning agency revisited

Before I show how the narrative account can help us understand addiction it is worth recapitulating how narrative agency builds on Bratmanian planning agency. On the planning account the agent's self-concept beyond their plans and policies is barely considered. Agents just need to have a sufficiently accurate understanding of their contingent circumstances to enable them to meet the norms of practical reason in intention formation and, thereby, to be self-governing. But this minimal consideration of contingent circumstances ignores three important factors. First, agents can usually interpret themselves and their situation in one of many ways, and each interpretation can be sufficiently accurate to enable successful planning and self-governance. But the way they *do* interpret often suggests a specific planning response and, therefore, a different path to self-governance than had they interpreted in other ways. Second, the intentional network agents develop affects subsequent interpretation; therefore intention formation and self-interpretation cumulatively influence each other.[1] Third, given the intimate relationship between intentions and self-interpretation, people do not typically consider their evolving self-interpretations on the one hand and their evolving planning responses on the other, rather they consider them simultaneously.

The narrative account claims that self-narratives are the form in which agents make sense of the evolving relationships between their intentions and contingencies. They strive to develop self-narratives that capture connections they find salient among their plans, policies, biography, body, expectations, accidents, windfalls, et cetera. The narrative agent still forms and follows intentions but those intentions are nested in self-narratives where they are connected with self-interpretations of various contingencies. When intentions are being followed by a narrative agent they are always shaped by the narrative context the agent has created.[2]

In Chapter 4 I discussed the explanatory benefits that narrative agency adds to planning agency in general. They can be summarised as follows: 1. Self-narrative shapes the character and motivational effects of its constituent intentions, desires and contingencies. Agency is, therefore, not just influenced by these constituents but by the way they are narratively inter-connected. 2. Successfully changing one's life requires new self-narrative projection but that projection must

[1] And so I agree with Taylor when he says, "To know who you are is to be orientated in moral space, a space in which questions arise about what is good or bad, what is worth doing and what not, what has meaning and importance for you and what is trivial and secondary" (1989, 28). Self-interpretations are normatively loaded.

[2] With the caveat that non-narrative agents such as advanced dementia patients can have values of limited sophistication as discussed in Chapter 4. Obviously narrative context doesn't shape these values.

make sense given the self-narrative to date; more radical changes in projection require reinterpretation of one's narrative past. This requires more than means-ends coherence of intentions because self-narratives are amalgams of intentions and contingencies. 3. The agent might find their existing self-narrative so well interwoven with contingent facts about themselves (enforced by inter-subjective relations) that they struggle to change it even though it has come to clash with their evaluations. In other words, existing self-narratives have their own momentum; certain stories tend to turn out certain ways. Agents can therefore find themselves enacting narratives they no longer value. 4. In an additional point only hinted at in Chapter 4, I adopt the view that self-narratives are not single stories but many partially overlapping, semi-independent narratives featuring the same protagonist. We don't typically try to force all the events in our lives into a single narrative form.[3] Instead we create many semi-independent narrative threads, some short lived, some that take up most of our lives; many overlap with each other, while others have loose ends that we never get around to completing.

> "The reality of a life lived in time is a perpetual weaving of fresh threads which link events and lives – threads that are crossed and rewound, doubled and redoubled to thicken the web" (Lloyd, 1993, p. 144).

> "…The stories that constitute [identities] often appear to be a hodgepodge of narrative fragments, some giving meaning to very localised experiences, others forming a kind of umbrella tale that pulls together a number of local stories but possibly leaves an equal number out of account. In bits of some people's stories, the plot intertwines with the stories of others but goes nowhere in particular, and there might or might not be connections among the narrative strands that constitute the histories of our relationships to the various things we care about. The story that is constituted around a deeply significant event may have an iterative chronology, looping back repeatedly through time as the event is characterised from different angles" (Nelson, 2001, p. 76).

We create these narrative threads whenever there are events in our lives, past and expected, brief or lengthy, that we want to understand. The protagonist and narrator are almost always the same

[3] Perhaps this would be possible, even desirable, if one particular open-ended plan was so important during one's entire life. But such lives are rare if they exist at all. Usually diverse values seem worthy of our attention and we develop multiple narrative threads to understand and realise them.

person throughout[4] although their character develops through time. The resulting collection of narrative threads makes a multifaceted life story.

Such a view of self-narrative allows us to make a range of distinctions. 'Thin' self-narration is where the agent creates relatively few connecting threads among events leaving much unexplained. Conversely, 'thick' self-narration is where the agent creates a relatively large number of narrative threads among a series of events, explaining their causes, implications and themes in detail. 'Narrow' self-narratives may be thickly or thinly narrated but most of the threads are focussed on one plot or theme. 'Wide' self-narratives have a diversity of semi-independent foci running in parallel – some may be more thickly narrated than others, some may be more independent than others. For ease of presentation I describe self-narratives that are relatively wide and contain more thickly narrated threads as 'robust' self-narratives. Thin and/or narrow self-narratives tend to undermine self-governance while robust self-narratives tend to support self-governance.

The multiple thread view of narratives furnishes our understanding of narrative momentum with greater detail. Narrow self-narratives of addicts tend to undermine self-governance while the development of more robust self-narratives supports self-governance. Chronic ends-ends inconsistent behaviour can be entrenched by two (or more) narrative foci that the agent develops in a semi-independent way. By narrating them separately the agent fails to face their ends-ends inconsistency or resolve it. I elaborate on these points as they become relevant below.

Narrative effects in addiction and recovery

In this section I argue for a variety of narrative-specific influences on addiction and recovery. 1. Existing self-narratives limit the plausible projections available to the agent. Self-narrative reinterpretations allow more highly valued narrative continuations to become more obvious and plausibly attainable. 2. Self-narrative changes can adjust the motivational effects of the narrative's constituents, so recovery can be supported by adjusting the strength of certain intentions and affective responses. 3. Narrow self-narratives focussed almost exclusively on drug-use, treatment and recovery will only ground fragile self-governance. Self-governance improves exponentially through the development of more robust narratives with multiple narrative threads. 4. Ends-ends

[4] Except in cases of DID, where the narrator/protagonist unit changes over time, and severe schizophrenia when the narrator may occasionally differ from the protagonist.

inconsistent behaviour can become chronically entrenched through the separation of the positive experiences of drug-use into a narrative focus semi-independent from the negative experiences of drug-use. 5. Self-narratives are co-authored and so the addict can benefit from efforts to interact with co-authors who encourage self-narrative development and avoid co-authors who entrench addiction self-narratives.

Many of these narrative-specific effects are often present to some degree but they are not always strong enough to make the difference as to whether someone recovers or not.[5] My purpose here, however, is to detail the explanatory advantages of the narrative account over the planning account. To illustrate these advantages I draw on rarer cases where narrative effects clearly make the difference between recovery and continued addiction.

Narrative reinterpretation to enable better projection

The first narrative effect I consider is the way an existing self-narrative limits the self-narrative projections that make sense. Because the agent can reinterpret their existing self-narrative, within limits, they can adjust the range of narrative projections that make sense but this is not always easy.

If we consider Sartwell's[6] case again we can see how the content of his existing self-narrative threads developed around addiction are extremely difficult to plausibly connect with a future of recovery.

> "Every time I have raised a bottle to my lips, I have felt free, and I have felt compelled. I made a decision, and the decision felt inevitable. I could have done otherwise, and I did what I had to do, *what my identity and history demanded*. Indeed, every time I raised a bottle to my lips, I kept faith with my father and brothers and my wife, my love; I shared their life and death. I kept faith with what we are, and I betrayed us. …" (Sartwell, 2008b, my italics).

[5] Indeed, because addiction becomes so holistically integrated with a person's life, recovery will rarely turn on a single causal factor and therefore treatment will rarely warrant a single focus. If narrative-specific effects are present to any degree that makes them relevant in treatment since the client will benefit from an accumulation of any factors that promote recovery. I consider the significance to treatment of narrative effects of all strengths in the last section of the chapter.

[6] Sartwell does eventually enter recovery but, because he only provides a first-hand account of his addiction and not his recovery, I only use his report to illustrate continued addiction.

Sartwell's narrative includes the facts that his brothers, father and grandfather all died through addiction or drug-related causes. Their deaths are particularly salient because Sartwell can see that his struggle to control drug-use is failing just as theirs did. To expect that his narrative would go any other way would appear to ignore or misinterpret the facts and their likely causal outcomes. Interestingly, he also speaks against the very possibility of recovery from addiction for anyone.

> "Addiction, I tell you, isn't an epic tale of redemption …It's dying by choking on your own vomit. It's common as excrement and as profound: reeking, valueless, purposeless, pointless, meaningless" (Sartwell, 2008b).

This belief goes against the objective evidence; at least some people recover. Given the power of self-narrative to focus one's attention his current self-narrative might blind him to the evidence for recovery or make him suspect that such recovered addicts weren't really ever properly addicted. Perhaps there's a sense in which he is deliberately giving himself an excuse to stick with the fatalistic narrative. But this doesn't make any sense given his clear statements that he disvalues drug-use and the obvious misery drug-use creates.

If we take Sartwell at his word, he is going out of his way to continue to use drugs because that is what his self-narrative demands. I suggest that he is motivated to enact this 'valueless' narrative because, even if it is detrimental, at least it provides an understanding of who he is and who he is becoming. Intentions directed at recovery appear hopeless from Sartwell's self-narrative perspective; to adopt them he would have to ignore who he is. But if he were just to change his self-narrative or ignore it, he would be deliberately undermining a credible understanding of who he is. Reference to this fatalistic narrative allows us to make a kind of sense of Sartwell's relapses. They may not provide pleasure, relieve pain, or have any value to him but they do accord with his self-narrative, his sense of who he is and what he can expect in life. Therefore it seems Sartwell's self-narrative is an independent motivating factor here, not an additional hidden reward or plan that ensures drug-use trumps his other values.

The next case I will consider is that of Isabel (in Addenbrooke, 2011). Her case illustrates how reinterpretation of a detrimental self-narrative can help support the recovery process. Isabel's mother died when she was eleven years old and, unknown to her at the time, her father had another secret family. As an adult she became addicted to opiates and she had two children. She first entered detox when her partner died. She describes that time as follows:

> "There were lots of parts of my drug use that had become uncomfortable in terms of there never really being enough drugs for me and the fact that I wasn't finding methadone particularly satisfying. I had quite a lot of physical problems at the time, with abscesses and not being able to get veins… I went into the hospital here to detox and I just remember it as being a something and nothing affair. I guess because my heart wasn't in it. There were lots of attempts to try and link me up with a social support network, but I think I just wasn't ready. I remember the day I left the hospital going straight to someone who'd picked up that day and I remember having a hit of heroin and methadone…" (Isabel in Addenbrooke, 2011, pp. 66-67).

Later she was in a new relationship and had begun drinking heavily. Her father had just died and at that time she found out about his other family. She says, "I was aware by now that I wasn't living as I wanted to, but I felt incapable of doing anything about it" (ibid, 67). She entered a new treatment program with her boyfriend, however, six weeks into treatment she received a sizable cheque (her inheritance) and so they left treatment and spent it on prescription benzodiazepines.

> "… I had this money, a raving habit and a feeling of desperation… We'd tell this wonderful story, that if they just gave us what we needed, then we would be the most wonderful parents. We both ended up with these enormously great big scripts and it didn't stop there – I was trying everybody else's big scripts as well. My existence was literally getting all these scripts and we still couldn't get enough benzodiazepines, so I was still doing all these private doctors, and the NHS doctors – I gave false names as a temporary patient, and that required some organisation, I can tell you, for somebody whose mind was befuddled" (Isabel in Addenbrooke, 2011, p. 69).

At the end of this extended benzodiazepine binge Isabel goes into treatment for what happens to be the final time. By this stage she had developed a fatalistic self-narrative analogous to Sartwell's. She says,

> "I was a pound over twenty stone. I could hardly walk. It was terrible. … I had all these sores and abscesses all over my legs… The methadone was so concentrated it used to create these burns that would get infected … I felt I was never going to be any good. … I'd really lost myself, I can't really begin to describe – I'd gone from, in my early twenties, this person that everybody had so much hope in, the good person, the star, the amiable one, the problem solver [to this]" (Isabel in Addenbrooke, 2011, p. 69).

Isabel's self-narrative is that of a steadily worsening, repeatedly relapsing addict who will never be any good. Such a self-narrative makes only a few unpleasant projections seem plausible. People who have struggled like this for so long rarely recover; they struggle on to eventually die by overdose or succumb to other drug-related health problems.

Isabel's recovery is preceded by a significant change in her self-narrative, not a change in her plans and policies (or her awareness of available rewards). Specifically she changes her narrative interpretation of her past. Here she describes the early stages of this successful treatment:

> "She [the social worker] did see something in me, and I felt that was really positive. But the other good thing about her was that she kind of explained things to me. In the past people made off the cuff remarks but nobody explained to me that broken attachments earlier in my life affected how I operated today. So I might have feelings of loss now that would be magnified because of feelings of loss earlier. … Key workers before had said, 'Do this, do that,' and I would tend to play the game. I would be the perfect patient. But now I was able to show the other side of me, that isn't the lovely, easygoing, compliant person. I was able to just be me" (Isabel in Addenbrooke, 2011, p. 70).

The key worker helped Isabel by considering Isabel's personal history and then suggesting a narrative reinterpretation tailored to her specific case. Isabel is informed that feelings of loss now might be magnified by feelings of loss earlier in her life. This information is relevant to her life because her mother died when she was 11 years old and she subsequently suffered other significant losses with the death of the father of her children and her own father. She uses this information to reinterpret her narrative history of drug-use from one of self-indulgent disgrace to an extended struggle to cope with the loss of her mother at a young age. This reinterpretation suggests that her addiction can be controlled by coming to terms with the original loss. This suggests planning solutions to addiction that weren't apparent before, such as finding better means of dealing with feelings of loss in general.

This narrative reinterpretation of the past helps set the foundation for a projection of recovery. It makes more sense that someone who understands and manages their feelings of loss can recover than a person who is regularly overcome by inexplicably strong feelings of loss. A plausible path to recovery is revealed – if she can deal with her emotional issues she can control her drug-use. This begins to change recovery from merely being something she evaluates positively into a genuinely plausible narrative trajectory for *her*.

The narrative account predicts that changes to self-narration of one's past will be important when making significant changes to projected narrative because the narrative has to make sense throughout. The planning account, in contrast, ignores the need for any reinterpretation of the past; intentions may have to be formed in light of occasional strong feelings of loss but those feelings are taken as a fixed aspect of the contingent context for planning. Of course Isabel could have adopted a policy to try and dampen those feelings of loss and that may well have helped but it was not until she changed her self-narrative that she could see that such a policy might be helpful. Without careful consideration of Isabel's self-narrative, social workers tried to get her to adopt generalised policies to little effect. Isabel complains that prior key workers would present generalised policies for recovery, "do this, do that." Even though Isabel could follow those policies briefly by "playing the game," it is not surprising that such an approach ultimately failed because the policies were ultimately incompatible with her self-narrative as it stood.

It is also worth noting that the narrative limitations existing narrative places on narrative projections can help entrench beneficial narrative trajectories. Isabel develops a narrative of redemption, of the drug addict who becomes a 'big hit' drugs worker. This narrative is more satisfying than just being a big hit drugs worker because it is a story of success against the odds from an extremely challenging past. As Isabel says, "I love coming back from behind." To relapse now would ruin the redemption narrative that she has "fallen in love with" and turn it into the tragic story of just another hopeless addict, but that change becomes more and more implausible as the redemption narrative develops. Isabel's new self-narrative therefore increasingly develops a beneficial momentum. It would take a significant change in contingency, self-interpretation, and inter-subjective environment to derail that recovery narrative. Although, this recovery narrative itself channels Isabel's motivations in certain ways, those ways now align with her evaluations and are relatively beneficial and empowering – she enjoys good health, financial security, more emotional stability, her future is more certain and thus allows for diverse plans, etc.

Changes to self-narrative adjust the influence of the narrative's constituents

Isabel's reinterpretation of her self-narrative arguably also changes the motivational effects of certain elements in her self-narrative in ways that improve her self-governance. Isabel claims that

she was motivated to recover by recognising a significant parallel between her own narrative and that of her daughter:

> "The real underpinning, the thing that was preying on my mind, was that my daughter was about to be ten, and I was scared that I would die and she'd end up without a mum – like I had" (Isabel in Addenbrooke, 2011, p. 69).

Now, I assume that Isabel had always had some commitment to plans and policies aimed at being a good mother. However, it is plausible that her recent narrative reinterpretation reshaped and *strengthened* the motivational effect of those intentions. Recall that her narrative reinterpretation has made the link between the death of her own mother and her subsequent addiction exceedingly clear. As a result she can now see that her own death would not just be traumatic for her daughter but it could saddle her daughter with the same inability to deal with loss. So this new narrative context strengthens the policy to stay alive for her daughter because to fail in that policy would not just cause a one-off trauma, it could consign her daughter to lifelong misery. Furthermore Isabel is all too well acquainted with the misery of addicted life so anticipating this future for her daughter is likely to be more motivating than anticipating a less familiar future that she could only imagine rather abstractly.

This narrative restructuring should also help protect Isabel against judgment shift. Recall from Chapter 2 that judgment shift is less likely the larger the gap in value between the normatively endorsed intention and the temptation (Karniol & Miller, 1983). By restructuring her self-narrative Isabel has widened the gap between the value she places on her policy to protect her daughter and the value she places on drug-use. She should, therefore, benefit from a decreased chance of judgment shift in regards to this policy. Of course she might still act akratically but, if we assume that akratic action typically requires a particularly intense temptation, then her narrative restructuring has provided a net gain in self-governance.

There is also reason to think that the context created by her new self-narrative changes the character of her feelings of loss. The earlier self-narrative allowed the instances of loss to balloon out of proportion. This may have overly motivated the need to soothe those pains with drugs. The new self-narrative explains the strength of present feelings by linking them to her mother's death. The original loss was indeed serious but that loss has long since happened and was not the fault of an eleven year-old girl. This helps her realise that a current feeling of loss is not actually as devastating

as it seems and so she can begin to bring the feeling in check; it is not so serious that it warrants self-medication.

The reverse of this effect of narrative context might be part of what prevents Sartwell from recovering. He says that, "…every act of love, every home place, every hint of peace or happiness…" is a premonition of future cycles of inevitable relapse only ending in death (2008b). Feelings of love, home, peace and happiness (and their associated targets) would normally be expected to motivate recovery, a lifestyle where such feelings will be more abundant. However, the motivating effects of those feelings are likely to be severely reduced for Sartwell because he has narratively connected them with his inevitable doom. If he could replace that narrative trajectory with a more positive one, the way Isabel has, then those feelings would no longer be tainted by the promise of suffering and death. There is, therefore, reason to think that the self-narratives of some addicts less than ideally configure their constituents, thus limiting the effect of potential sources of recovery-promoting motivation.

The planning account ignores the contextual effects of self-narrative on motivation; it thereby limits the agent's means of strengthening intentions to further intentions, pre-commitments, and muscle model willpower. However, if the agent does not take advantage of controlling self-narrative context she unnecessarily puts herself at greater risk of judgment shift and under-powered intentions. This in turn puts greater demands on planning efforts and muscle model willpower and creates the need for more restrictive pre-commitments.

Cumulative effects of a narrow self-narrative on self-governance

As I outlined above, the multiple thread view of self-narratives provides a means of describing variations in the structure of narratives. Robust self-narratives include a wide range of semi-independent foci, such as career, marriage, friendships, or hobbies. In such self-narratives several of these foci are also thickly narrated, that is, the agent develops many narrative connections relevant to these foci. As a result they understand many of the causal relationships and themes relevant to those foci. More fragile self-narratives, on the other hand, are characterised by a narrow range of foci and/or thin self-narration of those foci. Robust self-narratives support self-governance while fragile self-narratives leave it vulnerable, for reasons I make clear below.

Many addicts who have been struggling with addiction for years tend to have a thickly narrated self-narrative but one that is narrowly focused on their addiction career. In some of these cases the agent had originally developed a self-narrative with a variety of narrative foci but drug-use escalated at the expense of other foci and so the self-narrative narrowed. Sartwell indicates that this had happened in his case and alcoholism had become his "whole life" (Detritus). Other addicts have never managed to build up a diversity of foci in their self-narrative because heavy drug-use and other comorbidities start early in life. In such cases, the agent may be largely unaware of anything else worth valuing. This was the case for one of our interviewees who says,

> "I can't remember before I used. It's my whole life. Drugs have been my whole life. I remember at five sitting there drinking wine and eating mull cookies, so that's all I really remember is drug use" (R38, our interviewee).

So, many addicts have narrow self-narratives although they arrive there through a variety of routes.[7]

We can also note that, for most agents, a recovered lifestyle is a lifestyle that includes a variety of values and thus a variety of narrative foci.[8] This is not just because a variety of values appeal but as a matter of necessity – the agent needs to occupy their time. Sitting around just thinking about not using drugs will tend to be counter-productive. Minette, a recovered addict, retrospectively describes her experience of developing a more robust self-narrative as a much more challenging step than just abstaining:

> "When you first go into recovery you don't have anything other than the fact that you don't want to use anymore, but that's not enough. The fact that you don't want to use isn't going to sustain you. You need to have all the other things that make 'not using' viable: a good relationship, a home, something to do with your time, - *that's the single biggest area of concern*. What are you going to do with the twenty-four hours of the day? … All you know

[7] Another route to a narrow self-narrative is where drug-use is used as a technique of self-avoidance. "Addiction is in large part an avoidance of the self. It has its roots in self-hatred. … Drug addicts are people with emotional problems around self-esteem. … In many cases they are in search of an identity, a sense of self that is whole and entire" (Ben in Addenbrooke, 2011, pp. 173-174). In cases like Ben's, more robust self-narration would provide the self-understanding he needs but he anticipates that understanding as painful and so avoids it. He avoids that pain through drug-use but as a result the only self-narrative he develops is focussed on drug-use. It is, therefore, no wonder that Ben sees recovery as a search for his real identity, the self-narrative he would have had had he not engaged in this elaborate diversion.

[8] Some may be happy to replace one monomania (drug-use) with another but I assume this is unusual. Self-governance would also remain at risk in such recoveries for the reasons outlined below.

> is that you have to have a different life to the one you had. And I think that takes a good five years. … It's not done in six months" (in Addenbrooke, 2011, p. 108, my italics).

Minette makes it clear that recovery cannot just be about changing a narrative of addiction into a narrative of abstinence; it is about turning a narrative with a single focus into one with a diversity of different foci. Therefore, the addict does not just have to develop one narrative link between their existing self-narrative and recovery; they need to develop several such narrative links, one for each of the foci they aim to develop. This gives us an idea of the magnitude of the problem in addiction – the more narrow your narrative has become (or has always been) the more work you will have to do to recover.[9]

However there is reason to think that the effort required in this recovery work can gather or lose momentum. This is because narrative foci can support each other. The fewer narrative foci you have the more work you have to do to stabilise the development of any new foci. As more foci are narrated the process becomes easier as the existing foci support the new development. I begin with cases where addicts struggle to recover because they are limited by their narrow self-narrative. I then consider a case where successful recovery appears to have been supported by the development of multiple narrative foci.

There are two reasons why narrow self-narratives compound the difficulties of recovery. First, because self-narratives set the context for our experience and cognitive focus, narrow self-narratives tend to support narrow thinking. Addenbrooke notes that when talking with the long-term drug user, Lee, it can be hard to talk about anything but addiction:

> "It is difficult for Dr Rathod to lead him [Lee] away from details about his present drug use and treatment. This type of communication resembles the conversations of many long-term drug users who can hardly be drawn away from descriptions and arguments about the amounts, strengths, effects and availability of various drugs. These details can preoccupy them to the exclusion of virtually everything else" (Addenbrooke, 2011, p. 143).

[9] Empirical work by Best et al. (2012) supports this idea. This team investigated subjects who reported once being dependent on heroin or alcohol but who now considered themselves to be in recovery and had been abstinent for at least a year. There were statistically significantly associations between the frequency of days in which participants claimed to be engaged in meaningful activities and better quality of life, less anxiety, fewer physical health symptoms, better self-esteem and better self-efficacy (Best, 2012, 338). If we assume that relapse is less likely in these cases then it looks like time spent in meaningful activities supports recovery. Of course this data does not specify what range of meaningful activities the agent is engaged in. But I assume that a greater frequency of meaningful activity days would tend to correlate with a greater variety of activities.

Lee finds himself in a mental rut; he needs to imaginatively develop new foci but his current narrative restricts his focus. Without having *any* existing non-drug-using narrative threads to build on, all other potential foci will tend to seem alien. Even if Lee does begin to develop new narrative foci, initially he risks inaccuracy in those projections. The further from one's experience one imagines the more likely the imagined narrative either lacks cogency or involves misleading cogency. If cogency is lacking then the imagined future remains abstract and relatively unmotivating. Similarly, if the cogency makes the imagined future seem unrealistically unrewarding then the agent will be unnecessarily averse to enacting recovery projections. If the cogency of the imagined future overstates how rewarding it will be, that could lead to disappointment and relapse. This tentative phase where disappointment is likely will need to be worked through for each of the variety of new narrative foci Lee needs to develop if he is ever to escape his rut. However, it should become cumulatively easier as he builds up his self-narrative – Isabel's case illustrates this below. Second, having a narrow focus to self-narrative leaves the agent prone to wild oscillations in self-confidence. If one's narrative is monopolised by the struggle with addiction then a lapse appears depressingly significant. This isn't just one aspect of life going badly, this is one's "*whole life*" going badly. Success, of course, will also appear highly significant but few successful recoveries proceed without some lapses and the agent must be able to weather those lapses without succumbing to full relapse.

To see how multiple narrative foci can help recovery, again consider Isabel. Isabel developed several semi-independent narrative threads. Along with her social work and her study related to becoming a better social worker Isabel developed narrative threads around her relationships with her partner and the practice teacher (and her daughter as mentioned earlier). But the crucial point is that those narrative threads weren't just good for their own sake – they each helped stabilise the other threads involved in the recovery.

> "My sanity was the fact that, no matter what, I could still study; I could still do my job. You know, apart from my partner. I could do my job and I do it well. And, I had a lovely practice teacher. Wherever I go I carve out a mother. It's obviously unconscious transference, but I do. It's not a conscious thing. I had a wonderful relationship with this practice teacher, who was just like another mum. So I enjoyed that. They were the kind of things that kept me sane through that period" (Isabel in Addenbrooke, 2011, p. 71).

Isabel recognised that her recovery was fragile during this time, especially the drinking culture of some of her co-workers. She attributes her 'sanity' during this time to these multiple foci and this makes sense if we think that the diverse narrative context they create helps balance out the stresses and failures that occur from time to time in each thread. For example, troubles in the relationship with her partner or her daughter might be balanced by success at work and vice versa.

Compared to Lee and lifelong addicts like interviewee R38, Isabel was in a better position to develop her recovery narrative because she had a number of pre-existing, if poorly developed, narrative threads that she could draw on. She did not have to develop new narrative foci from scratch. Isabel connects her nascent plan of actually becoming a social worker with her old dream of becoming a social worker. Her history of playing roles for social workers gave her a good understanding of the dynamic between patient and social worker and what the job entailed. Perhaps most importantly, she had learnt by watching her mother solve other people's problems, and using her own mother as a role model was a way to keep her mother in her life in some form; a way that would have made her mother proud. She connects the strong relationship with her practice teacher with these deep roots in her self-narrative; this woman is something of a mother figure for her, which in some sense makes up for the one she lost. All these threads linked Isabel's plan with her wider self-narrative and so she was unlikely to feel alienated from that plan;[10] she had managed to make it a plausible continuation of her narrative. We can also see how these different foci support each other making their further development easier. For Isabel, the remembrance of her mother resonates through several narrative foci thereby helping to develop them: her relationship with her key worker, her work in helping people, the focal point of loss which helped her overcome her addiction, her relationship with her own daughter. As this theme around her mother develops it adds support to each new focal point that can draw on it. Equally the power of that theme to support these foci is increased by each individual focal point that develops it.

On the planning account, this dynamic of momentum in self-governance can be partially explained in terms of planning skills. As we saw in Chapter 5, coherent planning is required to construct valuable ends just as much as to discover paths to valuable ends one already knows of. Being familiar with a wider range of effective plans across different areas would presumably aid planning in other areas. Meanwhile, some intentions are adopted to target a specific end but then happen to diachronically support other intentions targeting other ends. So developing a more network of

[10] I discuss Christman's (2009) view that self-governance should be non-alienating below.

intentions targeting diverse ends can provide cumulative advantages to self-governance across the board. But this focus on planning doesn't fully capture the dynamic in some of these cases of addiction. Many of the new narrative foci that Isabel develops in recovery are not built on prior plans or knowledge of plans. She connects with prior threads in her self-narrative that had no obvious planning associated with them, e.g. the memory of her mother. Furthermore, the overarching theme around her mother that both reinforces and is reinforced by a variety of new narrative focal points is not easily reduced to a plan or policy. The prior plans she *does* reconnect with are not just redeployed 'as is' but are partially redesigned in the developing narrative focus. For example the prior plan to manipulate the therapeutic relationship to evade recovery is re-narrated as something like, 'managing the therapeutic relationship to support the client's preferred health outcomes.'

Ambivalent addicts revisited

The multiple thread view of self-narrative also provides an improved explanation of chronically ambivalent addicts. Recall that the planning account explains ambivalence as a commitment to ends-ends inconsistent intentions. This conflict is chronic because each intention provides an incommensurable reward and the agent struggles to make the hard decision to give up one or the other. This explanation is strained in cases of chronic addiction where the damage done through hesitation is extreme. It seems that, even though some of the benefits of drug-use are irreplaceable, abandoning drug-use would be a relatively small price to pay given the massive costs of ongoing ends-ends inconsistency. Recall the case of Bernadette. She says,

> "You can keep a modicum of respectability [with alcohol rather than heroin] … But you can't really, because *in the end you are going to show yourself up anyway*. I mean, I'm barred from every pub round here" (Addenbrooke, 2011, p. 31). "But I have caused riots with people, and I'm a nasty piece of bloody work when I've been drinking. I don't like that side of me, but as yet I haven't hit my rock bottom. *It's bound to come I know it is*" (ibid, 33, my italics). "It's like being two people really, being an addict, isn't it? One side of you wants the addict side, and the other side wants the side that isn't" (Addenbrooke, 2011, p. 34).

What is remarkable about such cases is that the agent recognises that their ends-ends inconsistent action is putting them on course for serious self-destruction. But despite this recognition they remain fatalistic about their future rather than trying to resolve the inconsistency in their values.

One explanation for the chronicity of ends-ends inconsistent behaviour in addiction comes from work by Shaffer, a proponent of motivational enhancement therapy. He speculates that, when established positive experiences of drug-use begin to conflict with negative ones, addicts experience ambivalence. These ambivalent feelings are painful and as a defence mechanism addicts 'split' their experiences. "When the split between positive and negative consequences becomes rigid, clients tend to flee into health with little capacity to sustain such respite or, alternatively, retreat into a more exclusive bond with the object of their desires" (Shaffer & Simoneau, 2001). In this way the ends-ends inconsistency becomes entrenched and the agent cannot overcome it until they learn to face and work through feelings of ambivalence.

The narrative account can build on Shaffer's explanation. The 'split' could be created by selective self-narration, which would explain how it develops and why it is entrenched. One narrative focus develops around the positive effects of drug-use. This thread is particularly well developed early in drug-use but whenever the agent notices a connection between drug-use and positive outcomes they contribute narrative threads to this focus. As negative experiences begin to accumulate around drug-use they begin to narrate these too but in a separate set of narrative threads. This negative narrative focus develops over time and so ambivalent addicts end up with two relatively thickly narrated foci around drug-use.

Such self-narration can account for a kind of psychological 'split' because of the focussing and contextual effects of self-narrative. When they recall parts of the positive drug narrative their focus is directed accordingly and any drug-use projections tend to have positive cogency. When they recall the negative drug-use narrative the effects are the opposite. Because the narrative foci are split, so is their cognition and they struggle to properly consider the pros *and* the cons in any one decision.

The agent can usually switch between two or more self-narrative threads when he chooses.[11] When making decisions he is aware of the existence of the other way of seeing drug-use but its content and the causal connections it would make salient remain largely excluded from the process. This

[11] The cognitive biases caused by the narrative split are not as strong as those proposed in Ainslie's hyperbolic discounting where the SS reward totally blinds the agent to the LL reward.

would explain why these ambivalent agents remain aware of the problem without being able to resolve it. It is one thing to recognise the inconsistency, another to realise what has to be done to overcome it and yet another to actually do what is required.

The narrative account adds an additional reason as to why agents might chronically struggle to overcome their inconsistent behaviour. Because both narrative foci have become well entrenched in self-narrative, to give up one or the other is to give up a significant part of who one is. Perhaps it is these high stakes that ground the ambivalent feelings and explain why they are experienced as being so painful. If the narrative account is right, the possibility of overcoming ends-ends inconsistency will become increasingly difficult the longer the agent fails to resolve it because the two inconsistent narrative threads only become more entrenched in self-narrative over time. The narrative account, therefore, makes the chronic nature of these addicts' end-ends inconsistent behaviour more plausible than on the planning account. They are not just averse to giving up the irreplaceable value of drug-use, they are averse to giving up a whole focal point in their self-narrative – a focal point that their selective narration has ensured appears wholly positive.

The narrative account can also add to Shaffer's suggestion of how chronic inconsistency can be overcome. Shaffer suggests that motivational enhancement encourages the agent to face their ambivalent feelings and develop a tolerance for them. The idea is that only once the resistance to ambivalence is reduced will the agent be able to properly consider the pros and cons and thereby make a settled decision (Shaffer, 1997; Shaffer & Simoneau, 2001). Subsequently therapy can aim to increase the motivation for recovery (if that is what the agent wants and needs).

The narrative account describes some detail in the underpinning psychological processes. The agent's ends-ends inconsistency will only be overcome if he forms narrative connections between the two independent narrative threads he has developed around drug-use. Creating these connections makes both the pros and cons of drug-use salient simultaneously and it necessarily causes feelings of ambivalence. This is why reducing resistance to feelings of ambivalence works; it helps the agent form the necessary narrative connections. The increase in connections between narrative foci underpins the ability to simultaneously consider both the positive and negative aspects of drug-use and thereby make more settled decisions.

Co-authoring effects

As we saw in Chapter 4 our self-narratives are co-authored in discourse and action where people implicitly and explicitly express views about each other. Subsequently, "we are never more (and sometimes less) than the co-authors of our own narratives" (MacIntyre, 1984, p. 216). On balance it is probably a beneficial disposition to take on others' narrations of us because others report from a perspective that we can only poorly emulate.[12]

Such co-authoring is typically influenced by master narratives. Each society has an evolving collection of master (or archetypal) narratives, some of which apply to each of us. These master narratives represent the accumulated body of cultural knowledge. We therefore tend to draw on master narratives when we self-narrate and narrate each other because these narratives generally provide a short-cut to knowledge.

> The "…stock plots and readily recognisable character types of master narratives characterise groups of people in certain ways, thereby cultivating and maintaining norms for the behaviour of the people who belong to those groups, and weighting the ways others will or won't tend to see them" (Nelson, 2001, p. 106).

We cannot just change our society's master narratives, evade master narratives we don't like, or ignore others' specific views on who we are. If we were to shut out the sources of co-authoring whenever we didn't like them then we would tend to suffer from ignorant, deluded, self-flattering narratives. That would ultimately undermine our agency by putting us out of touch with reality and excluding us from cooperative activities. To distance our self-narratives from master narratives is to distance ourselves from established ways of understanding the world. Therefore, to self-narrate without help from master narratives demands the additional effort of developing novel understandings. Self-narratives that conflict with master narratives will deviate from others' expectations and so they are more likely to be ignored or misconstrued; more effort is required to have such self-narratives heard, understood, and accepted. Therefore self-narratives that are consistent with the narratives others tell of you have more momentum than those that contradict them.

[12] I assume that taking seriously others' narratives of us is an important part of developing an effective external perspective on our self-narratives. As we saw in Chapter 4 being able to take an external perspective on one's self-narrative is important in judging whether the narrative is sufficiently realistic.

Unfortunately the range of master narratives applicable to addicts tend to make negative outcomes more plausible than positive ones. Here is a selection of the master narratives society makes available to addicts: Addicts have a brain disease so they cannot control their behaviour; junkies are untrustworthy, lazy, unhealthy and beyond help; alcoholics won't recover until they hit rock bottom; addicts must abstain to recover because, for them, controlled use is impossible. These master narratives contain grains of truth, or are true in certain cases, but they will also support unnecessarily detrimental self-narration and co-authoring in many addicts. Addicts who want to recover will need to overcome the momentum in their self-narratives caused by detrimental master narratives.

To get an idea of the effects of co-authoring on self-narrative momentum consider the views of Interviewee 101 (from our pilot study). He says,

> "I'm conscious of the way you get treated in the pharmacy, the way you're put in a separate queue and so … I'm conscious of the fact, of the way people obviously regard what I am. … Sometimes I go outside and I really feel hated, I really feel like people are thinking God who's that guy who never talks to anybody, who is that guy, that freak, what's he? ... I mean when people make their lists of most hated people I mean junkies or ex-junkies is way up near the top I'd say. So you know you're dealing with that all the time" (Interviewee 101, our pilot study).

Interviewee 101 would not be encouraged to socialise with people who hated him and considered him a freak. Because that co-authoring influences his self-narrative, he will tend to assume that he is a hateable freak, independent of any particular person's views, and thus avoid social contact. The self-narrative of being a socially isolated freak tends to be self-fulfilling. This self-narrative has a strong momentum because it reduces his social opportunities to have alternative self-narratives heard and supported (or challenged and developed). Interviewee 101 confirms this, saying,

> "… I don't see any other people so I don't know if I'm forgiving or compassionate or (laughs) conciliatory 'cause I don't get any experience…" (Interviewee 101, our pilot study).

These comments were in response to being asked to rank values[13] rather than any specific attempt to develop a more compassionate and conciliatory self-narrative. However, it is not hard to imagine that this social isolation would also limit any changes in self-narrative Interviewee 101 would want to initiate.

Isabel had also been facing detrimental co-authoring effects which became obvious when she saw her old drug-using acquaintances.

> "I'd changed and moved on, and people were angry about it. It was like, 'She goes up there and she floats in here now with all these plans, and we're on the floor.' Twenty years of this chaos, but my family would never talk to me. This was part of my problem in the first place, no one had ever been as upfront as to say to that. That's what I realise now" (Isabel in Addenbrooke, 2011, p. 70).

In hindsight Isabel realises that her old social environment limited the changes she could make in her self-narrative. Her acquaintances looked down on recovery while her family had left the 'chaos' of her life unchallenged. Her new social environment involves other professionals who are more open to co-authoring different narratives with Isabel as a trainee and then as a colleague. Ultimately, Isabel's colleagues come to co-author her self-narrative as a talented colleague while her old acquaintances have little to no co-authoring input in her narrative. Isabel's daughter now co-author's Isabel's self-narrative as a responsible (rather than absent) mother.

So, we can see from these cases that our self-narratives can gain momentum towards detrimental outcomes when others tend to co-author such outcomes or refuse to help us co-author changes. People trying to develop a recovery narrative will be more likely to succeed if they are in contact with supportive co-authors who draw on positive master narratives.

The planning account doesn't focus on the inter-subjective aspects of agency. Bratman describes the development of networks of normatively organised intentions from the subjective point of view. Of course the account could relatively easily accommodate a picture where we develop our networks of intentions in collaboration with others. Certain people might encourage intentions that are consistent with self-governance and others might encourage intentions that break norms of practical reason undermining self-governance. However, it is the nature of the planning account to

[13] The first part of the interview is structured and includes a Schwartz Value Survey. These comments are a reflection on those values and come from the subsequent, in-depth part of the interview.

focus solely on intentions while assuming contingent aspects of self-concept to be fixed. The narrative account goes further by considering how others influence our interpretations of our contingent features and how others specify the expected futures for people with such contingencies. This affects the intentions we adopt in ways that go well beyond what is needed for diachronic stability, ends-ends consistency and means-ends coherence. As we have seen, certain self-interpretations limit the future one can plausibly aspire to, ruling out the consideration of certain intentions despite them being compatible with norms of practical reason.

Self-governance redescribed

One implication of the narrative account is that we have to redescribe the planning theorist's view of agential authority and, therefore, what counts as self-governance. To do this in detail would require more space than I have available here but I will sketch the outline of the issue drawing on Christman's (2009) work and suggest a potential solution.

Recall from Chapter 3 that Bratman is motivated to attribute agential authority to normatively organised intentions (that stem from evaluative judgments) over evaluative judgments alone. Evaluative judgments are ruled out for two reasons. First, they exhibit less diachronic stability than intentions as they are prone to judgment shift. Second, they cannot resolve choices between highly valued but incommensurable options while commitment to intentions can. When multiple intentions are normatively organised they tend to support each other providing even greater diachronic stability. Such networks of intentions guide agents' actions and the weighting they tend to give to desires and emotions so it makes sense to claim that those intentions represent agents' personal identities. Normatively organised intentions have agential authority and, therefore, the better one acts in accordance with them the better one self-governs. To act on evaluations (or desires) that contradict these intentions would break norms of practical reason and exhibit a lack of self-governance.

Cases like Isabel and Sartwell are problematic to the planning account because these agents appear to fail to self-govern despite continuing to act in accordance with drug-use intentions that are relatively means-ends coherent and ends-ends consistent. These intentions may be relatively short-sighted and sparse compared to those of the average agent but they form a network nonetheless. At least, these agents do not exhibit any *other* network of intentions that is more diachronically stable,

ends-ends consistent and means-ends coherent and that would, therefore, have a better claim to agential authority. On the Bratmanian picture, if Sartwell's self-governance is threatened by anything, it is by his nascent ends-ends inconsistent intentions to recover. Of course, if he no longer values his drug-using lifestyle at all then he should abandon those plans and policies and then he would be free to self-govern in other ways. But he does not drop those plans and policies.

As I have argued above, the narrative account provides a number of ways to understand how the agent's self-narrative can entrench a network of intentions and self-interpretations. If those entrenching effects are strong then, if the agent changes their evaluations, they will find themselves in opposition to their own intentions. That opposition will last until they overcome the narrative momentum or realign their evaluations with the established narrative. During such a period of opposition why should we think that Isabel's and Sartwell's evaluations of their drug-using lives represent where they really stand rather than their destructive, but relatively well-developed, addiction narratives? In these cases of addiction it seems intuitively correct that agents' evaluations have agential authority. They really do want to recover. Their lives are clearly miserable and there is no obvious motive to lie about their evaluations.[14] If we want to say that sometimes evaluations have agential authority over established intentions after all, we need a way of distinguishing which evaluations have authority over existing intentions or narratives and which do not.

Christman (2009, Chapter 6) describes one way we can make this distinction. He argues that to properly self-govern it is not enough to competently enact a narrative; in addition, one should not feel deeply alienated from that narrative.[15] He describes alienation as an evaluative and affective state that motivates the agent to repudiate, resist and alter the effects of the alienated content.[16]

[14] We might add that these agents appear capable of living better lives so their evaluations could have an influence on practical agency. In Bratman's terms it is, therefore, *possible* that they could actively value that better life rather than merely evaluate it positively (Bratman, 2007, p. 248). In comparison, I might evaluate the life of an astronaut highly without that valuation ever having an input into my practical agency. I cannot actively value such a life because I cannot enact it. Presumably such non-practical evaluations have little agential authority for Bratman, certainly less than committed intentions. However, these non-practical evaluations can tell us a lot about another agent; we are interested in the 'inner' lives of people even when there is little clear practical import.

[15] Developing the condition of non-alienation is a response to the problems that result from claiming that self-governance requires that an agent must *identify* with her intentions. The identification requirement is too strong. We may identify with ego-ideals without ever achieving them and that shouldn't entail that we are failing to self-govern. Furthermore we are ambivalent about many things in our self-narratives without rejecting them. If identification is weakened to some attitude such as 'acknowledgement,' then it becomes too weak and agents like Sartwell count as fully self-governing.

[16] Christman claims that, when evaluating aspects of self-narrative for alienation, we should consider how they developed (2009, pp. 137-138). However he acknowledges that one does not *necessarily* feel alienated from traits that one cannot remember developing or that were the result of coercion. Furthermore, given a large enough change in evaluative stance, it would be possible to become alienated from a narrative that one exhibited high self-governance

Isabel and Sartwell appear to be alienated from their addiction narratives. I assume that a chronic failure to resist or alter the alienating threads of one's self-narrative can result in a more fatalistic attitude towards it, such as we see in Sartwell's case. But, for the reasons Bratman outlines, not just any evaluation can be attributed agential authority. So on what grounds can we say that Sartwell's negative evaluation of his addiction narrative has more agential authority than that alien narrative (which he continues to enact)?

A crucial feature for attributing agential authority to an evaluation is the diachronic stability of that evaluation. "We must … require that the reflection definitive of autonomy be such that it would yield the same result if repeated over a variety of conditions" (Christman, 2009, p. 152).[17] In many cases of drug-addiction the negative evaluation of the drug-using narrative exhibits this diachronic stability. Sartwell, for example, evaluates his drug-use negatively even as he uses drugs.

But why would we be sure that such diachronically stable evaluations are the agent's own? The agent could just be parroting evaluations that others have encouraged him to make. The basis for the agential authority of these diachronically stable evaluations, I suggest, is that they come from the agent's 'external perspective' as discussed in Chapter 4. Recall that taking the external perspective is a way to make judgments while being detached from any particular narrative thread. Such a perspective avoids potentially biasing emotional contexts of specific narrative threads but has the disadvantage of lacking the kinds of imaginative play those emotions drive (Mackenzie, 2008). In the external perspective the agent is not dissociated from all self-narrative effects but more or less equally affected by their entire array of self-narrative threads.[18] If this is correct, then the better developed the agent's external perspective is the better they will be able to form diachronically stable evaluations of their particular narrative threads. The external perspective is developed by creating a wide variety of narrative threads. Agents with relatively narrow self-narratives will struggle to judge particular narrative threads without that judgment being excessively influenced by the cogency of the thread judged; they have an underdeveloped external perspective. For that reason Lee (above) may have become virtually unable to make a stable reflective evaluation of his addiction because he has no perspective except from within that

in creating. Cases of alienation from addiction narratives are typically between these two extremes; the agent has been somewhat coerced by a substance and social milieu but he has also colluded in the narrative development.

[17] Exactly how stable evaluations need to be in these cases is difficult to define. However, when agents negatively evaluate their ongoing addicted behaviour their evaluations are often consistent for months or even years and so I assume the conditions of diachronic stability are met.

[18] Presumably the more thickly narrated threads will have a more dominant influence than thinly narrated ones.

narrative thread. Similarly, ambivalent agents like Bernadette might alternate between an addiction narrative thread and a recovery narrative thread; whichever narrative thread Bernadette cognitively inhabits drives negative judgments of the other thread. Without developing a more stable external perspective she cannot form a diachronically stable evaluation of either thread. Therefore, cases like Lee may not experience alienation[19] and cases like Bernadette may experience diachronically inconsistent alienation but both still exhibit a serious deficiency in self-governance because they lack a well-developed external perspective.

So the narrative account provides two ways of understanding how agents in these circumstances *lack* some self-governance. First, agents like Sartwell and Isabel may have a network of intentions that is relatively internally consistent but a part of that network is inconsistent with their wider self-narratives. Those wider self-narratives underpin their negative evaluations of those inconsistent intentions (and the narrative threads in which the intentions are embedded). Second, agents like Lee and Bernadette lack a sufficiently robust self-narrative to make stable evaluations of their intentions (whether those intentions happen to exhibit internal consistency or not).

If this is how we understand a lack of self-governance in these cases, how can such agents improve self-governance? For agents like Sartwell and Isabel who have a sufficiently wide self-narrative to stabilise their evaluations, improving self-governance requires that their intentions (and associated narrative threads) are made sufficiently consistent with that wider self-narrative. We saw how Isabel managed to go about this; she reduced her drug-use and managed to develop a number of replacement narrative threads that she evaluated more positively. However, self-governance is not improved by consistently voicing positive evaluations of unattainable narrative projections; one has to be able to enact those narratives. Therefore improved self-governance requires that the developing narrative threads are not just sufficiently consistent with the agent's evaluative stance but also meet the various reality constraints set out in Chapter 4. Isabel did not just come up with a consistent fantasy of a self-governed life, she connected the new, more consistent narrative threads to her past and ensured that their projections were realistic. In cases such as Lee and Bernadette, where the agent lacks a sufficiently robust self-narrative to stabilise evaluation, the

[19] Agents that do not experience alienation will not lack self-governance by their own standards. To claim that this group of addicts lack self-governance entails that we compare them with an inter-subjective standard of evaluative reflection; they cannot seem to recognize that their life is far worse than it could be. Appealing to an inter-subjective standard for self-governance may commit me to a substantive rather than a purely procedural account of self-governance but I cannot go into that here. In any case, I find it intuitively plausible that we can judge the quality of agents' evaluative skill and that skill may be so impaired that we can question their self-governance.

road to self-governance is more difficult. They need to develop more robust self-narratives to improve their evaluative abilities. As they do so, they also need to ensure that their new narrative threads are consistent with that developing evaluative position.[20]

Summary

Overall self-narrative has a range of effects on self-governance that do not appear on the planning account. First, established self-narratives of addiction limit the plans and policies that appear plausible to the agent. Reinterpretation of one's self-narrative can help the narrator adopt plans and policies targeting more highly valued narrative aspirations. Second, self-narrative context affects the motivational effects of the narrative's constituents including intentions. Capitalising on these effects can counteract motivational effects entrenching addiction and support motivational effects promoting recovery. Third, addicts especially struggle to recover when their self-narrative has a narrow focus on drug-use, treatment and abstinence. Such narratives limit imaginative yet realistic self-narrative projection and leave the agent prone to crises of self-confidence. Fourth, ends-ends inconsistent behaviour can become chronically entrenched through the selective narration of the positive experiences of drug-use into one narrative focus and the negative experiences of drug-use into another narrative focus. This underpins unbalanced decision-making and means that giving up either end appears to require abandoning a significant part of who one is. Narrating links between those narrative foci can help stabilise decision-making, overcoming chronic ends-ends inconsistency. Fifth, self-narratives are co-authored and so the agent should aim to associate with those who will co-author the self-narrative projections the agent evaluates most highly.

Presumably many of these effects are present to some extent in all struggles with addiction, indeed any struggle for self-governed action. However, in some cases, such as those of Isabel and Sartwell, these effects arguably make the difference between continuing addiction and recovery. On the planning account, the addicted behaviour of people like Isabel and Sartwell is a mystery. They apparently have planning skills and options available to them that they plausibly claim to value more highly but they do not recover. When Isabel finally does recover, it does not seem to coincide with any change in circumstance or planning that hadn't been present earlier. I suggest that Isabel

[20] An additional difficulty that can arise in these cases is that because of their evaluative impairment the agent judges that they do not have an evaluative deficiency. Their drug-use has come to dominate not just their practical lives but their cognitive lives as well. At this point, improvement in self-governance requires a paternalistic element with all the ethical issues that entails.

recovered when she did and not before because of the changes she managed to make to her self-narrative (with co-authoring support). In contrast, Sartwell's addiction continued despite his planning skills and the availability of more highly valued options because he could not change his self-narrative of doomed addiction. Finally, the narrative account requires that we adjust the Bratmanian view of self-governance. Self-governance requires more than acting on normatively organised intentions because one can become alienated from one's intentions. On the narrative account, the agent's evaluative judgment can have agential authority even when it clashes with an established intention. Evaluations have agential authority when they display sufficient diachronic stability and are underpinned by the agent's wider self-narrative. Self-governance at its best involves enacting intentions (narrative threads) that are consistent with one's diachronically stable evaluative stance which is underpinned by a robust wider self-narrative.

Narrative concern in treatment

What are the implications of the narrative account for the treatment of addiction? In this, the final section, I sketch out some answers to this question. Recovery always involves changes in self-narrative but the difficulty in making those changes varies. Sometimes agents can develop a recovery narrative relatively easily once the other aspects of addiction are treated, e.g. strong cravings, withdrawals. These people do not suffer from any strong detrimental narrative momentum. Other agents struggle against detrimental narrative momentum and so continue to struggle with addiction even once many of the other factors entrenching addiction are treated. This latter group, in particular, will benefit from treatment informed by narrative agency.

Treatment informed by narrative agency can be divided into two categories and I treat each in turn in this section. First, any therapeutic interaction can benefit from the clinician being *aware* of narrative effects. Clinicians who are alert to the dynamics of self-narrative, co-authoring and master narratives in the recovery process can facilitate the changes in self-narrative required for recovery. I point out the value of narrative awareness in general and then illustrate these points with reference to specific examples of addiction treatment involving opiate maintenance, CBT, and Twelve-Step programs. Second, clinicians can go further and explicitly target the client's self-narrative in treatment. I consider one form of such treatment – Narrative Therapy (NT). NT might be necessary

for recovery in cases where detrimental narrative momentum is particularly strong; however, it should benefit any agent's recovery given that all recovery involves self-narrative change.

General benefits of narrative awareness in treatment

People seeking treatment for addiction are usually assigned a counsellor or social worker who functions as their primary contact with the healthcare system (perhaps in collaboration with a psychiatrist or doctor specialised in addiction). This main contact provides counselling, broadly construed, that ranges from providing food and housing, through to developing coping strategies, administering CBT or motivational enhancement therapy, and monitoring opiate maintenance treatments.[21]

There are a variety of ways that narrative awareness can benefit recovery. Here, I first outline the general benefits of narrative awareness. These benefits apply to any clinician-patient relationship. I then consider a range of more specific treatments and illustrate the therapeutic benefits of narrative awareness in each context.

The primary benefit of narrative awareness is an improved ability to develop a therapeutic relationship. Failing to thoroughly engage with the client's narrative risks alienating them from the treatment process, thereby damaging the therapeutic relationship. Further benefits include being aware of the co-authoring effects at play in the therapeutic relationship and the important idiosyncratic features of each person's case.

These effects are most apparent when we consider the experience of patients whose self-narratives have been ignored or held against them. Interviewee 101 has the following impression of counselling:

> "My experiences have been fairly bad to be honest with you. People don't get it, people just out of uni they just don't get it. It's depressing even talking to them 'cause they just so obviously don't have any idea what they're saying. Having people tell you theory that actually was just really disheartening 'cause theory's just theory. …'cause the people I see down at my level of mental service are 21, 22, 23 year olds just out of uni on the way to a

[21] In Australia, prescription of opiate maintenance (and dose or medicine changes) can only be made by a doctor registered to prescribe methadone and/or buprenorphine.

> much better job, with a better class of clientele so I just get them on the … at the start of their careers. And some of them are very nice, but … yeah they ... you feel like they're quite sheltered and you feel like they … you can't help thinking that they kind of like … you kind of disgust them. They you know they see this sort of old loser and it's you don't get this … any sense of that anybody on your side particularly" (Interviewee 101, from our pilot study).

Here we see a fundamental failure in counselling because the client hasn't been listened to or understood. Trying to understand a person's self-narrative is a crucial part of respecting them as a person because their self-narrative communicates fundamental detail about who they are. If the client feels disrespected they will be unlikely to engage in a therapeutic relationship. Interviewee 101 feels that the counsellors he saw either didn't want to know who he was because they were disgusted by him and preoccupied with their own lives; or, they wanted to help but couldn't due to their inability to understand his narrative. This entrenches detrimental views he has of himself and the stigma he experiences outside the clinic:

> "…they just assume that I'm some sort of horrible loser they can treat however they fucking well like. … often quite contemptuously, and they feel entitled to because who's going to pay any attention to my side of the story anyway" (Interviewee 101).

One would hope that despite the stigma the client faces in society, at least his counsellor would listen to, and try to understand, his self-narrative. Similarly we saw in Isabel's case above that some key workers would ignore the details of her story and just recommend general techniques for recovery – "do this, do that." It wasn't until a key worker took the time to properly consider the details of her narrative that a therapeutic relationship flourished and treatment succeeded.

Narratively aware approaches suggest ways of overcoming such detrimental clinician-patient divides. In order to better appreciate the nuances of another's narrative Charon (2006), a proponent of narrative medicine, recommends skills analogous to "close reading" taught in graduate programs in literature where the reader habitually pays attention not only to the words and the plot but to all the literary devices in the text.

> "When these reading skills are brought to bear on a clinical encounter with an individual patient, they do their work, in part, by bridging some of the relentless divides – arising from the conflicting understandings of mortality, contextualisation, causality, and emotional

> suffering – that separate clinicians from patients. The clinician equipped with the skills of close reading will, we hope, be better able to reach across these divides once equipped with the wherewithal to absorb form, understand plot, hail narrator, follow metaphor, track time, and live in the face of desire. These clinicians will then have the capacity to attend to what patients tell them and to represent that which is heard in a form that honours the narrative acts performed by the patient" (2006, p. 126).

In other words, through practice engaging with narratives and trying to appreciate their nuances, counsellors can get better at properly appreciating who their clients are.[22] Although some divide between clinician and patient (indeed between any two agents) is ubiquitous, these divides are not completely unbridgeable.

That said, the ability to understand another's narrative is dependent on having at least *some* similarity of experience. Wittgenstein alluded to this point when he suggested that if a lion could talk we would not understand him (2001, p. 190).[23] Obviously, counsellors do not face a gulf in understanding of this magnitude and perhaps, with time, a sympathetic counsellor and an articulate client could overcome even a large gap in shared experience. However, the more quickly a therapeutic relationship can develop the better it is for the client. With this in mind it will be better to match counsellors to clients to facilitate the development of the therapeutic relationship. Counsellors would not necessarily need to have experience of drug-use or addiction; an older counsellor may be better placed to understand older clients, or a counsellor with children might be better placed to understand a client with children, and so on. Just having had similar experiences to a client is, of course, no guarantee of an easily developed therapeutic relationship. The counsellor must still attend carefully to the client because there will always remain a myriad of differences in the detail of each person's experience. There is also a risk that a counsellor who has had an experience similar to the client will favour their own narrative interpretation and be less open to the alternate narrative that the client might give. So even counsellors with similar life experiences to the clients need to develop skills of understanding unfamiliar narratives. Narrative-aware

[22] Ideally with the support of experienced narrative interpreters who can point out subtleties to the student.
[23] Of course we do share some similarity of experience with a lion so if the lion could talk we might be able to understand something of his narrative.

counsellors are better placed to reduce the divide of understanding between clinician and patient and thereby foster an improved therapeutic relationship.[24]

Clinicians who have narrative awareness also enjoy the benefit of understanding their co-authoring influence on the client's self-narrative. The client will inevitably pick up on the clinician's implicit or explicit approval and disapproval, their interest or disinterest, and their attempts to make sense of the client's situation. Charon notes that in any clinician-patient interaction the clinician plays an important role in developing the patient's self-narrative.

> "We clinicians donate ourselves as meaning-making vessels to the patient who tells of his or her situation; … the patient cannot always tell, in logical or organised language, that which must be told. Instead these messages come to us through the patient's words, silences, gestures, facial expressions, and bodily postures as well as physical findings, diagnostic images, and laboratory measurements, and it is our task to cohere these different and sometimes contradictory sources of information so as to create at least provisional meaning" (Charon, 2006, p. 132).

When the clinician helps the agent create a self-narrative thread they must attend closely to the patient because the clinician must help develop the *patient's* narrative not the narrative they prefer or assume the patient to have. Charon refers to narrative interpretations of past and present experience but the development of aspirational narrative futures is equally important. This is particularly relevant in addiction cases involving clients who are certain they want to control their drug-use but are not so sure what they want to put in its place. As we saw above the clinician needs to help co-author a future that the client evaluates positively but clinicians have to make sure it is the client's aspiration not one the clinician is foisting on them. This can be done by drawing on various underdeveloped threads that already exist in the client's self-narrative.

Narrative change and opioid maintenance treatments

Consideration of self-narrative change in recovery is relevant to opioid maintenance treatments in two ways. First, a consideration of narrative change in recovery provides support for short-term

[24] Obviously, developing relationships takes time and effort so it will be important to try and maintain successful counsellor-client relationships throughout the process of recovery. Every time a client has to change counsellor the process of developing a relationship will have to begin anew.

maintenance treatment, particularly in the context of poverty. Second, a narrative approach can help us move beyond the debate as to whether recovery requires total abstinence or can involve permanent maintenance treatment.

An opiate addict without a job who has to raise new money for drugs almost every day is stuck in a debilitating cycle; the cycle dominates their thoughts and there may be no obvious way out. The activities they focus on and the co-authoring effects they are exposed to only further thicken the self-narrative threads around addiction. The agent's self-narrative is only likely to improve if they have the time to begin to focus on different activities and engage with people outside that cycle. Maintenance treatment can help the agent change their practical and social context so that re-narration becomes a possibility.[25] For people who haven't fallen into that extremely debilitating cycle, maintenance treatment still provides the benefit of legal opiates and a reduced risk of conviction. So a narrative approach supports the majority view that opiate maintenance treatments are beneficial in the short-term.

Longer-term use of opiate maintenance is more controversial. Some believe that recovery ultimately requires complete abstinence from opiates. Others argue that ongoing opiate maintenance is compatible with recovery and even that attempts at abstinence are too risky. I suggest a narrative approach provides a way of settling this debate.

I begin with the views of those who believe that recovery entails abstinence. Russell Brand, a recovered heroin addict, said the following when presenting evidence in favour of rehabilitation over maintenance treatment to the UK Home Affairs Select Committee:

> "Once I dealt with the emotional, spiritual, mental impetus, I no longer felt the need to take drugs or use drugs. … If you have the disease or the illness of addiction or alcoholism, the best way to tackle it is to not use drugs in any form, whether it is state-sponsored opiates, like methadone or illegal street drugs, or a legal substance like alcohol" (Brand, 2012, p. 2).

Brand makes it clear that, for him, addiction had emotional, mental and spiritual causes that could not ultimately be healed by chemistry. Ben voices a similar opinion:

[25] Whether that re-narration actually occurs or not depends on further factors.

> "Prescribing substitute drugs is not an adequate answer for drug users. It's not what people are looking for really. I believe what's at fault is something far deeper. It's a sense of themselves" (in Addenbrooke, 2011, p. 169).

On the other side of the debate, people argue that a fulfilling life is compatible with maintenance treatment and that abstinence is a goal foisted on addicts by people who do not understand them. A continuing use of maintenance treatment is not *necessarily* indicative of an underlying emotional, mental or spiritual problem. In Sapphire's case it seems that methadone is compatible with a life she evaluates highly. She says,

> "I'm not sure if I'll remain on methadone forever. I have again reduced my dose without relapse, but I feel that taking it is a preventive measure against craving for, or using, any drugs or alcohol. I'm worried that without methadone my life would be filled with drugs again rather than doing the things I currently enjoy doing, like working, travelling and volunteering" (Sapphire, 2013).

Furthermore, coming off maintenance treatment involves a high risk of relapse and there is no point in taking that risk if continued maintenance is compatible with a satisfying life.

> "[Abstinence] doesn't help. It gives people one defeat after the other. They come home clean from treatment and everything is fine for 3–4 months. And then they have some money … and splash, they are stuck again … In my opinion, drug-free treatment shouldn't be used at all, only on people who are younger than 18. When they become 25–27, it's all over and done with" (in Jarvinen & Andersen, 2009, p. 878).

A narrative approach suggests we should make both master narratives available. Being forced to draw from a master narrative that is fundamentally incompatible with your narrative aspirations is likely to be detrimental. The agent will, rightly, feel that they are replacing their addiction narrative with another that also undermines self-governance.

This view is supported by Jarvinen and Andersen (2009) who interviewed people in Copenhagen, where permanent maintenance treatment had become the default for opiate addiction. Only some addicts found that the master narrative of permanent maintenance easily cohered with their narrative aspirations. Others had to be convinced but ultimately enjoyed the relief of giving up their goal of abstinence. They could focus on more easily achievable goals and didn't have to "beat themselves up" over continued opiate use. However, there remained a group of addicts who

continued to resist the master narrative and felt coerced into methadone when they would prefer to work towards abstinence.

> "When I look back now, I can see that I should never have taken the first tablet [of methadone] . . . Methadone is bad because it makes you stay in the [drug] milieu … Methadone does not change things. I have had a huge side abuse, you see. Methadone was just the thing I took in the morning in order to be well and then I hurried out into the streets to make money" (Jarvinen & Andersen, 2009, p. 878).

This interviewee, like Ben and Brand, is likely to feel that the clinicians are deliberately reinforcing their drug problem rather than helping them. We see similar problems when we consider the opposite situation where clinicians pressure all patients towards abstinence. Sapphire has experience of this:

> "It was very frustrating to be stable and not using illicit drugs, only for the CDT [Community Drug Team] to coerce me into reducing my methadone dose as soon as I gave drug-free screens again. Due to the CDT's coercion, I started supplementing my prescribed medication with street-bought methadone. I was sick to death of the treatment system at this point" (Sapphire, 2013).

It appears that both master narratives have potential to harm and heal depending on the individual's aspirations. These master narratives will harm when they clash strongly with the patient's aspirations but will support recovery when they align with those aspirations. If we ignore the variation in individual narrative aspirations and assume one story fits all, then we inevitably force an incompatible master narrative on a subset of agents. If forced to choose between a coerced recovery narrative and an addiction narrative that they created themselves, the continuing addiction narrative might appeal insofar as it is at least their own. A narratively-aware clinician would promote the master narrative that would best help the client form narrative threads between their existing narrative and their aspirational narrative projections.

Cognitive Behavioural Therapy

CBT covers a wide variety of techniques aimed at identifying and improving detrimental thought patterns and beliefs. Detrimental forms of cognition are significant in developing and exacerbating addiction as well as various comorbidities (e.g. depression, anxiety, low self-esteem, and learned

helplessness). The narrative-aware clinician recognises that the client's self-narrative will be shaped by this detrimental cognition and such cognition may have to be overcome if an agent is to develop a recovery narrative.

All forms of CBT are general purpose, that is, they are designed to work for any human who has a type of detrimental cognition. Therefore forms of CBT can be useful without having to be matched too carefully to the individual's self-narrative context. Like opiate maintenance treatment, CBT can help lay a platform for self-narrative change. In the case of CBT, however, that platform takes the form of cognitive capacities and habits. The client could use CBT to reduce impulsivity, improve social interaction, increase sensitivity to one's emotions, and so on. An example can illustrate this process.

Agents with a pessimistic explanatory style tend to interpret adverse life events as their own fault and the effects of those events as permanent and pervasive (Peterson et al., 1995). Such an explanatory style is associated with learned helplessness, depression and addiction. CBT in such cases can train the individual's explanatory style to be more optimistic. If successful, the person should slowly begin to see themselves as more capable and imagine more ambitious goals, and so this form of CBT should encourage development of a self-narrative with beneficial momentum. So the narrative-aware clinician could hope to improve an agent's self-narrative by working on the cognitive habits that are partially responsible for its form.

However, we can see that this form of CBT, like all CBT, leaves the details of narrative reinterpretation up to the agent; a more positive explanatory attitude underdetermines specific interpretations. CBT cannot recommend any specific narrative self-interpretation or projection over another. As we have seen, new self-interpretations and projections need to make sense in light of some aspects of existing self-narrative if they are to be incorporated. In some cases, such as Sartwell's, this might be a problem because the existing narrative context might be particularly resistant to new interpretations no matter how positive your attitude. Therefore some agents may find that a change in explanatory attitude (or any general technique) is insufficient for beneficial narrative change. These agents need treatment that is explicitly narrative-focussed where the specific content and connections of their narratives can be addressed.

Twelve-Step programs

Twelve-Step programs appeal to the narrative-aware clinician because of the character of their inter-subjective environments. These inter-subjective environments provide enhanced co-authoring opportunities for the client.[26] The client's peers provide multiple narrative examples of recovery in progress that the client can borrow from. These narratives are likely to be more compelling when they come directly from the people living them than if a clinician recounts them from an archive. Furthermore the peer group provides real-time reification or challenge to the client's nascent self-narrative changes, which helps strengthen that development.

> "The sponsee [in AA] gradually learns to tell his or her own story appropriately, and in the process acquires the identity of 'sober alcoholic'" (Swora, 2001, p. 365).

The co-authoring effects of a group are likely to be more powerful than the co-authoring of the clinician. The clinician is just one person whose opinion may be easily discounted as being 'out of touch' with the lived reality of addiction. It is harder to ignore a group of your peers who may understand your life even better than you do yourself.

A variety of treatments draw on the therapeutic benefits of peer groups, e.g. therapeutic communities or group therapy, but a benefit specific to Twelve-Step meetings is that people continue to attend meetings even when they have been abstinent a long time.[27] Therefore new members are exposed to successful recovery narratives, not just narratives of people at the same stage as they are. The benefit of being in social contact with people in recovery has been demonstrated by Best et al. (2012) in a study of 205 people who had been dependent on alcohol or heroin sometime in their life but in recovery for at least a year. They found statistically significant associations between having more people in recovery in your social network and better psychological quality of life, social quality of life, environmental quality of life, total quality of life, lower depression, and higher self-esteem.

The narrative approach also reveals a significant downside to Twelve-Step programs. Twelve-Step programs impose a master narrative with a relatively strict format that members must adopt to be part of the group. That master narrative involves hitting rock bottom, suffering from a disease, aiming at abstinence, admitting powerlessness and need of a higher power, never fully recovering

[26] This is also a benefit available in group therapy and therapeutic communities.
[27] Furthermore relationships with sponsors who are in recovery are made available so that there is a supportive contact almost constantly available.

but always being an addict in recovery, and experiencing a spiritual awakening. It should also involve making up for the wrongs they have done others and remaining part of the group to help other addicts. Twelve-Step programs have made their master narrative more accessible to non-Christians by loosening the definition of 'a higher power' to include anything that has a greater power than the addict and that can guide him towards recovery. However, other aspects of the master narrative appear unnecessarily restrictive or, at least, are restrictive in ways that make it an inappropriate treatment for addicts with certain kinds of self-narrative.

By demanding abstinence, the Twelve-Step Narcotics Anonymous program creates a problem for those addicts using opiate maintenance treatments (even in the short-term).

> "In NA … people on methadone maintenance have traditionally not been considered to be "in recovery," and their "clean time" typically wasn't even allowed to start to be counted until they stop maintenance. Some meetings won't even allow people taking maintenance medications to speak, because they are seen as active users who simply have substituted one drug for another" (Szalavitz, 2012).

From a narrative-aware perspective this subordination and new stigmatisation would dramatically decrease the value of the program for people on maintenance. The abstinence requirement has relaxed to some extent so that *some* groups will allow people on maintenance treatment to participate and count as being in recovery (Szalavitz, 2012). Therefore, clinicians need to know the position the local NA group takes before recommending them to someone on maintenance treatment.

The requirement of hitting rock-bottom is meant to be a way of weeding out people who really want to abstain (or encouraging that goal) from those who remain ambivalent; ambivalent addicts are just going to waste their own time and the group's. This attitude assumes that whatever the agent's problems in controlling drug-use, overcoming ambivalence is under his control; something that he needs to address pre-treatment. He needs to evaluate drug-use negatively to succeed in the Twelve-Step program but changing his evaluation is beyond the scope of Twelve-Step treatment. This pressure may work for some people but the narrative-aware clinician can see that conflicting evaluations are not always resolved by fiat.[28] In fact most patients are not ready to change when

[28] This issue is not confined to Twelve-Step programs. "Clinicians, embedded in the same society as their clients, often are anxious for rapid results. Driven by these and other social forces (e.g. managed care requirements for brief treatments), clinicians unconsciously collude and falsely assume clients want to change when they come to the office requesting treatment" (Shaffer & Simoneau, 2001, p. 101).

they first enter treatment (Prochaska et al., 1994). But as we saw above, chronic ends-ends inconsistency might be a symptom of the agent's inability to govern their self-narrative development; they may be narrating their experience into two independent narrative foci. This problematizes the pre-treatment requirement to hit rock bottom because an addict might need help to overcome their ambivalence. If we wait for them to resolve that momentum on their own we might consign them to unnecessary suffering, even death. To demand all clients be wholehearted prior to treatment sidesteps one of the crucial targets of treatment for some addicts.

Other forms of treatment are more accepting of ambivalence. As we saw above, proponents of motivational enhancement strategies, for example, recognise that recovery comes in stages and that in the early stages there may be little motivation for recovery.[29] Clients tend to be ambivalent about recovery *at best*. Part of treatment is to foster ambivalent feelings as a step towards making a settled decision to pursue recovery (Shaffer, 1997; Shaffer & Simoneau, 2001). Ambivalence is seen as a natural part of recovery and it can be managed as part of treatment; the agent doesn't have to overcome it themselves by hitting rock bottom. The narrative aware clinician can tailor treatment to their client depending on how ambivalent their narrative is. If the client is relatively wholehearted in their positive evaluation of recovery then Twelve-Step programs might be beneficial for them. If they remain more ambivalent, however, that ambivalence might not be overcome in Twelve-Step programs and motivational enhancement strategies, for example, might be better.

Finally, the Twelve-Step master narrative's requirement to see oneself as fundamentally an addict no matter how long one has abstained is problematic because that will not be the best narrative for everybody. Some people will be better served in early recovery by the narrative aspiration of being 'cured,' of no longer being an addict, of being an ex-addict, or even of being a person who used to use drugs too much. Equally, later in recovery, these forms of self-narrative might be more protective because they exclude drug-use from the range of options even up for consideration. An addict in recovery might continue to crave, continue to consider what it might be like to use again and thereby risk relapse. But for someone who no longer self-narrates as an addict of any kind, these drug-using possibilities will be less likely to even enter their thinking; they may be able to exclude drug-use more completely from their lives.

[29] Although proponents of this treatment are clear that they aim to enhance existing motivations and remove resistances (Shaffer & Simoneau, 2001). They do not attempt to coerce motivation for recovery where *none* exists.

This is not to say that the there are no benefits of self-narrating as an addict in recovery – it helps avoid complacency that might lead to relapse. It fosters a healing bond between members, 'you're one of us and we're here for you.' It may play an important role in member retention despite longer-term abstinence. If the member continues to consider themselves at risk they will have a selfish as well as an altruistic reason to stay in contact with the Twelve-Step group. The problem, again, is that the Twelve-Step master narrative unnecessarily restricts the recovery narratives of the program's members.

Narrative Therapy

Another treatment approach suggested by the narrative account is to not just administer treatment based on an awareness of narrative effects but to explicitly target the content of the client's self-narrative. This approach should benefit all agents aiming to recover since all recovery involves change to self-narrative. However, in some cases it might even be necessary for recovery because the agent might be unable to make narrative changes despite treatments such as CBT and methadone maintenance, providing helpful conditions for change. There is a range of psychotherapies that focus on changing the content of the client's self-concept but here I only have space to focus on the one that most explicitly deals with self-narrative – Narrative Therapy (NT). NT has been developed by Michael White and David Epston (1990; 2007) and, as its name suggests, it is underpinned by something very similar to the account of narrative agency I have outlined above.

NT is a collaborative process in which the therapist and the client work together to heal a problematic self-narrative. Narrative therapy stresses the importance of a non-hierarchical relationship where the therapist does not have access to an objective knowledge. By abandoning any special claim to objective knowledge the therapist can only build on the material the client provides. Lang (2013), for example, talks about providing invitations to the client, "how about thinking of it like this?" if the narrative suggestion doesn't resonate with the client and they turn it down, then the clinician moves on, looking for another interpretation. This approach reduces the chance for the therapist to (accidentally) coerce vulnerable agents into alienating self-narratives.[30]

[30] This is essentially the mode of relationship the narrative aware clinician would try to develop.

The core NT process begins with the client carefully articulating his self-narrative while the clinician asks questions to get extra detail. People seeking treatment typically have a dominant narrative that is 'problem saturated.' Early in NT, the therapist works with the client to isolate what is problematic in that dominant narrative. The client is then helped to *externalise* these problems. For example, the clinician refers to them as 'a person struggling with addiction' rather than 'an addicted person.' The problem might even be personified, e.g. 'What has addiction tried to steal from you?' (Lang, 2013). Externalisation is intended to help the client stop habitually self-identifying with interpretations promoted by the detrimental dominant narrative (M. White & Epston, 1990, p. 39). This opens up conceptual space for alternative self-narration – if the dominant narrative is incorrect, what is correct?

> "Those aspects of lived experience that fall outside of the dominant story provide a rich and fertile source for the generation, or re-generation, of alternative stories." (M. White & Epston, 1990, p. 15).

Aspects of experience that fall outside the dominant narrative are called "unique outcomes."[31] The clinician and client search the client's past to find unique outcomes, e.g. "Can you tell me how, despite the powerful influence of alcohol, you made the decision to come here today and ask for some help?" (Lang, 2013). Once a unique outcome is found the therapist guides the thickening of narrative threads around that event by asking questions like, 'What does your success at resisting the problem say about you as a person?' (M. White & Epston, 1990, p. 17), thus increasing the understanding of why and how that exception was possible. Then narrative threads are gradually extended from that unique outcome into the past and future to form an alternative and preferred self-narrative. White describes the overall process as an 'insurrection of subjugated knowledges' (M. White & Epston, 1990, p. 27).

During NT outsider witnesses may be invited. Outsider witnesses are significant members of the client's social network or others who have had similar problems to the client. The witness initially just listens to the therapist interview the client. The therapist then asks the witness say what phrase, image or resonances the client's story has for them (and is asked not to evaluate or critique the story). Subsequently, the client describes their own response to the outsider witness's interview.

[31] A term first used by Goffman (1961) based on the idea that people tend to exclude unique features of their experience in favour of features that are commonly shared by others. Within NT this is reconceptualised as a tendency to self-narrative in ways that fit with socially dominant narrative forms.

Thus NT structures narrative co-authoring by providing the client (and therapist) with an alternate perspective on the client's life, perhaps along with an insight into how the client affects others or a promising direction to solve problems.

NT supports self-narrative development using literary means. For example, the client might write letters describing where they started and where they have now progressed to. Those letters are to help reinforce the changing narrative but (with consent) may also be used to help others with similar problems or even the author should they find the problem redevelops. Another example is where the therapist writes a letter at the end of therapy with their predictions for the client's next six months which the client takes away with them. The idea is that the encouraging predictions will be self-fulfilling and support the agent's progress beyond the end of face-to-face meetings (M. White & Epston, 1990 Chapter 3).

We can see how NT would help the recovery process with reference to Isabel and Bernadette. Recall that Isabel's dominant narrative was one of consistent failure to recover despite treatment; even her own family didn't want her to return to treatment since it had been associated with exacerbation of her addiction. Isabel's recovery involved a key worker suggesting an alternate narrative thread for her life. However she did not receive NT so she had to do a lot more of the re-storying work herself.

Isabel's self-narrative had numerous thin (subjugated) threads that didn't easily fit within that dominant narrative. Some exceptions to the dominant narrative include: her therapist telling her she was intelligent, the similarity she notices she has with her mother being able to solve other people's problems, knowing what role to play in social interactions, and getting in touch with her real feelings. She is telling her story retrospectively so we cannot tell how much work she had to do to reveal those aspects. In any case, they were likely to be much less obvious when she entered treatment for that last time. She had to work to knit them together and thicken the narrative around them to develop a serious alternative to the dominant narrative. This process would likely have been more efficient with the aid of a narrative therapist who could guide the process.

In Bernadette's case NT could investigate whether she had developed two independent narrative foci, one around positive experiences of alcohol and the other around negative experiences of alcohol. If that was the case, then NT could help her develop narrative threads between the two narrative foci so that she could more thoroughly consider the pros and cons of alcohol use in

decision-making. She might then be able to embark on a more diachronically stable narrative projection whether that still included drinking or not.

Isabel's recovery contrasts with the NT approach in one significant way. NT focuses on unique outcomes defined as something the agent achieved despite their problems. But Isabel's recovery began with a reinterpretation of her mother's death and subsequent problems with loss, not any particular achievement on her part. Clearly the focus on action is not necessary to begin to re-story the dominant problematic narrative thread. The agent might develop some subjugated narratives by first significantly reinterpreting events where she is passive; so NT may unnecessarily limit itself here. That said, acts leading to unique outcomes are likely to be the best place to start because the aim is to develop a self-governed life not merely interpret it in a better way. In any case there will usually be multiple ways to begin to develop the narrative threads of recovery. Had Isabel been in NT perhaps the therapist would have helped her develop a recovery narrative by building on her achievement of turning up for treatment, or something she managed to do for her daughter.

Summary

All clinicians would benefit from being aware of narrative effects. Carefully listening to the client's self-narrative and narrative aspirations is crucial to understanding the client's problem with drug-use and what would constitute recovery for them. It also helps develop a therapeutic relationship because the client will tend to feel that they are being treated like a person and, therefore, continue to engage. When selecting and designing treatment that will facilitate the development of recovery the narrative aware clinician can match the client's narrative with the most suitable master narratives inherent in different treatment modalities.

For those people who suffer from entrenched detrimental self-narrative threads NT is likely to be a particularly helpful form of treatment. The therapist can help the client's narrative development in the following ways: isolating and directly targeting detrimental narrative threads that need to be re-storied; weaving narrative links between disparate threads causing chronic ambivalence; developing a realistic narrative projection of recovery that fits with the client's evaluative judgment, and; creating narrative threads that link the agent's current narrative with that projection.

Conclusion

In Chapter 4, I argued that the narrative account reveals a range of effects on self-governance that do not appear on the planning account. In this Chapter I have built on that argument by illustrating how these effects are relevant to understanding addiction and recovery. The planning account emphasises networks of intentions and ignores the impact of how the agent interprets the contingencies of his life and how he links his intentions with those contingencies. The narrative account includes networks of intentions but argues that they are developed within self-narratives where they are narratively linked to the agent's interpretations of their contingencies.

These wider considerations reveal the following effects in the efforts of addicts to regain self-governance. First, established self-narratives of addiction limit the recovery-directed intentions that appear plausible to the agent. This can lead the agent to ignore or refuse what would be effective recovery intentions despite the fact those intentions would meet norms of practical reason. Reinterpretation of one's existing self-narrative can make new recovery-directed projections of that narrative more plausible and thus facilitate the adoption of recovery intentions. Second, the context created by a self-narrative influences the motivational effects of the narrative's constituents. Re-narrating one's self-narrative to change the context for those constituents can increase the motivational power of intentions, desires and affective responses that are compatible with recovery and decrease the motivational power of those that are incompatible with recovery. Third, recovery typically requires the development of several narrative foci each requiring imaginative yet realistic self-narrative projections. Creating such projections are more difficult the more the addict's self-narrative has narrowed to focus exclusively on drug-use, treatment and abstinence. Similar effects would be predicted by the planning account yet it underplays them by ignoring the importance of being able to link developing projections with past contingencies. Fourth, some addicts display ends-ends inconsistent behaviour that goes on chronically despite them recognising the problem. The planning account describes the chronicity of the condition in terms of the agent being unable to make the difficult decision to drop an irreplaceable end. This explanation is strained when the ongoing damage done by ends-ends inconsistency appears to clearly outweigh the irreplaceable end. The narrative account suggests that such addicts selectively self-narrate the positive experiences of drug-use into one narrative focus and the negative experiences of drug-use into another narrative focus. This entrenches unbalanced decision-making where the agent either tends to ignore the negative effects of drug-use or focuses on them to the relative exclusion of the positive

effects. Furthermore, when the agent considers overcoming their ends-ends inconsistency it appears as if he has to abandon a significant part of who he is. Narrating links between the divided narrative foci can help stabilise decision-making and overcome chronic ends-ends inconsistency. Fifth, self-narratives are co-authored and so the addict aiming to recover should aim to associate with those who will best co-author the relevant self-narrative projections. The planning theorist might expand his account to consider how others help us plan but that will only ever capture part of the inter-subjective effect. People also influence how we interpret our contingencies, how we connect them with our plans. This puts limits on our planning that go well beyond the norms of practical reason. Many of these effects will be present to some extent in all struggles with addiction, indeed any struggle for self-governed action. However, in some cases, such as those of Isabel and Sartwell, these effects arguably make the difference between continuing addiction and recovery. Therefore the narrative account can help our understanding of addiction and recovery across the board but in some cases it will be essential. Because narrative threads may drive action that the agent feels alienated from we need to correct the Bratmanian picture of self-governance. Agents exhibit greater self-governance when they enact narrative threads that are consistent with the evaluative stance underpinned by their wider self-narratives.

Given these narrative effects it should come as no surprise that clinicians would benefit from being aware of them. Understanding the self-narrative effects operating in the client's case clearly should involve a careful consideration of the client's narrative. This also allows the clinician to discover which narrative projection the agent would consider to count as recovery or, at least, begin to develop that projection. Armed with this information, the clinician can being to select and design treatment that will complement the client's narrative aspirations, avoiding those that would draw on clashing master narratives. This narrative awareness in general should support the therapeutic relationship because the client should feel that their interests are being considered and they are no having any master narrative forced upon them.

Finally, Narrative Therapy directly targets the content and structure of the client's self-narrative. This therapy could complement any other treatment modality and should help in most cases. However, it is likely to be particularly useful in cases like those of Sartwell, Lee, or Bernadette where the agent suffers from an entrenched, narrow detrimental self-narrative or an entrenched ends-ends inconsistent narrative. The narrative therapist can help the client re-story detrimental narrative threads, develop a realistic narrative projection of recovery that fits with the client's

evaluative judgment and narrative past, help build multiple narrative foci, and develop narrative links between disparate narrative foci underpinning chronic ends-ends inconsistency.

Conclusion

I have argued that a complete understanding of agency should be informed by the motivational effects of agents' self-narratives. Self-governance can be improved or undermined depending on how agents self-narrate. As a contribution to the philosophy of action my arguments support the accounts of other narrative theorists, such as Schechtman (1996, 2007) and Velleman (1989, 2005), however there are two novel aspects to my contribution. First, I have developed a detailed taxonomy of how self-narratives affect agency and made it clear that those affects cannot be captured by non-narrative, intention-based theories, such as that of Michael Bratman (1999, 2007). Second, my argument that a self-narrative can be detrimental to self-governance has provided a novel line of argument in favour of Christman's (2009) view that self-governance at its best requires non-alienation from one's intentions.

Drawing on cases of addiction was not just a useful way to highlight the explanatory power of the narrative account of agency, it also enabled me to build on the recent philosophical work on addiction. This work is one of the first attempts in the philosophical literature to apply a narrative approach to the conceptual problems of addicted action. I have argued that a narrative account of agency provides a way to understand a range of cases that other accounts cannot. These are cases where the agent's self-narrative entrenches addiction and where changes to self-narrative aid recovery. The narrative account also suggests a number of considerations that should be relevant to the treatment of addiction.

In the first three chapters I pointed out problems with how Heyman and Ainslie characterise normal agency. They claim that action is caused by the maximum reward the agent takes to be available and they assume that all rewards are commensurable and reward sizes are fixed extra-agentially. The resulting position is one where action is determined by extra-agential rewards and the possibility of *self*-governance is eliminated. These theorists refer to what seem like techniques of self-control: global choice, prudential rule-following, pre-commitments and test case willpower. However, the deployment of these techniques is purely a function of social pressures, the reward profile of the environment, and sub-agential competition between desires. Furthermore, all those techniques can only diachronically stabilise action in repeating scenarios and so they are inadequate to provide diachronic stability to the multitude of unique diachronic goals in human lives.

The normative planning accounts promoted by Holton and Bratman improved on these weaknesses by introducing more plausible techniques of diachronic control and preserving the possibility of self-governance. I argued in favour of Holton's claim that intentions are the most commonly used

tools of everyday diachronic agency on several grounds. Intentions are versatile, providing diachronic stability to unique plans, global choices, prudential rules, and idiosyncratic policies. They allow decisions to be settled ahead of time which allows for greater diachronic coordination in action. Furthermore, intentions do not rely on creating irrational biases in future decision-making; the agent can rationally act on prior decisions as long as he has good reason to trust his earlier decision-making. I agreed with Holton that muscle model willpower is one of the agent's skills for maintaining diachronically stable action, although I suggested it may not only help non-reconsideration of intentions but also help resist akrasia once intentions are reconsidered. I also argued that pre-commitments usefully complement intentions despite being relatively unwieldy.

Bratman develops conceptual space for self-governance by first observing that agents are regularly faced with choices between incommensurate and mutually exclusive rewards. Many rewards are incommensurate partly because they involve fundamentally different experiences and partly because of the agent's natural epistemic limitations. When faced with these choices the size of the reward underdetermines action; the agent needs to commit to one over another without knowing which will be the most rewarding. Therefore, action is not simply caused by the perception of greatest available reward but by the agent's commitment to a course of action with an intention. As we saw, intentions need to meet the norms of diachronic consistency, means-ends coherence and ends-ends consistency. The better the agent follows these norms in developing her network of intentions the better those intentions interlock and the more diachronically stable the network is. This leads Bratman to claim that those intentions represent who the agent is. The better the agent acts in accordance with her normatively organised intentions the more self-governance she exhibits. The agent's efforts are important in developing and maintaining self-governance in several ways. Synchronically, more or less muscle model willpower can be exerted. Diachronically, the more effort the agent puts into designing and organising his intentions (and pre-commitments) to be sufficiently means-ends coherent and ends-ends consistent the better he will avoid scenarios where he risks breaking practical norms.

In Chapter 5 I showed how the theoretical advantages of the planning accounts benefit our understanding of addiction. Planning accounts make sense of paradigmatic cases of addiction where people experience distress, ambivalence, and a struggle synchronic to drug-use. The planning theorist would expect to see such phenomenology in someone struggling for self-governance. It is no surprise to the planning theorist that recovered agents and clinicians describe recovery as effortful because improvement and maintenance of self-governance requires efforts to

redesign one's intentional network and deploy willpower. Planning accounts also provide a way of understanding why certain treatments work. Resigned fatalism and chronically failing recovery plans may be overcome by treatments that improve means-ends coherence in planning. Chronic ambivalence may be overcome by a concerted effort to find ends-ends consistency among highly valued but incommensurate goals. In contrast, choice theorists insist that action is caused by maximum expected reward and so they are restricted to describing paradigmatic addicts as deceptive, self-deceived, or behaving unintentionally. None of these explanations are adequate. Furthermore, choice accounts reduce the process of recovery to a fortuitous change in circumstance but, according to many addicts and clinicians, that is rarely sufficient.

Despite the theoretical advantages of the Bratmanian account of self-governance, I argued in the fourth chapter that it ignores the effects of the agent's wider self-concept on self-governance. That self-concept includes the agent's interpretations of his contingent features and circumstance. Intentions have to make sense in light of this wider self-concept and so the agent can influence his intentions by how he develops this self-concept. Self-narration, I argued, is the way that agents connect their intentions, desires and contingencies so that they make sense in light of each other. Self-governance, therefore, does not just involve creating a network of intentions but this more inclusive process of self-narration. The narrative view adds explanatory power to the Bratmanian view because the motivational character and intensity of intentions, desires and contingencies is influenced by the prominence we afford them in our narratives. Furthermore, our existing self-narratives pre-reflectively set the context for our attention, experience, and thought so they limit the narrative continuations that make most sense. These qualities give self-narratives momentum which explains why they can be especially difficult to change, difficulty that is not anticipated on planning accounts.

In the final chapter, I brought the additional explanatory power of the narrative account to bear on addiction and made the following points. First, narrative momentum can entrench addiction. Established self-narratives of addiction make a future of ongoing addiction seem more realistic than recovery. This can prevent the agent from adopting recovery-directed intentions even if those intentions meet norms of practical reason. Furthermore, the established narrative context can pre-reflectively increase the prominence of drug-using intentions and drug-related contingencies while effacing aspects of self-concept and cognitive contexts that would support recovery. Efforts of self-narrative reinterpretation can overcome these effects. Second, addicts often have a relatively thin and narrow self-narrative focussed on drug-use. Recovery usually requires the development of a

wider and thicker self-narrative but the narrower one's self-narrative has become the stronger the effects of detrimental narrative momentum become. Planning accounts underplay this effect by ignoring the importance of being able to link developing projections with past self-interpretations. In cases of extremely narrow self-narrative, agents' single narrative focus clouds their judgment; they lose the ability to adopt the external perspective which is necessary for reflection on their life. Third, planning accounts can explain chronic ends-ends inconsistent behaviour as the inability to give up a valued but inconsistent end. However, this explanation is strained by the extreme damage caused by the inconsistency. The narrative account provides a more plausible explanation of how inconsistency can be so entrenched for some agents. Such agents develop independent narrative threads, one in which drug use experiences are cast in a predominantly positive light and one in which the negative experiences of drug use are collected. As each thread becomes well established the agent's ends-ends inconsistent actions and evaluations become entrenched as they alternate between the two cognitive contexts of those threads. Initially, the agent avoids combining these threads because they are averse to the feelings of ambivalence it creates. Once the threads are established the agent may be further discouraged from achieving consistency because it appears to require abandoning an established narrative thread (as well as giving up an irreplaceable end). Fourth, the co-authoring effects of other people and archetypal narratives exert a strong influence on whether an addiction narrative remains entrenched or a recovery narrative is developed. The planning theorist can expand his account to consider how others help us plan but the focus on intentions excludes the effects that others have on the interpretation of our contingencies and how we connect our plans with those contingencies. These narrative effects on self-governance will tend to be present to some extent in all struggles with addiction and in some cases they may make the difference between continuing addiction and recovery.

As a result of these points we must redescribe Bratman's account of self-governance; it is not always enough to enact ends-ends consistent intentions. Agents have a deficit in self-governance if they are alienated from any of their intentions (and the related narrative threads), even if those intentions are ends-ends consistent with the other intentions. Agents' wider self-narratives underpin their negative evaluations of those inconsistent intentions and give those evaluations diachronic stability and agential authority. Such agents can improve self-governance if they adjust their intentions (and associated narrative threads) so that they are sufficiently consistent with their wider self-narrative.

Finally, this work suggests that the treatment of addiction would benefit from clinicians who carefully consider the agent's self-narrative and who are alert to narrative co-authoring effects. It is also likely that many addicts would benefit from treatment modalities which explicitly target detrimental self-narratives and help the agent re-narrate them, such as Narrative Therapy.

References

Addenbrooke, M. (2011). *Survivors of addiction: Narratives of recovery*. Hove: Routledge.

Addolorato, G., Caputo, F., Capristo, E., Domenicali, M., Bernardi, M., Janiri, L., . . . Gasbarrini, G. (2002). Baclofen efficacy in reducing alcohol craving and intake: A preliminary double-blind randomized controlled study. *Alcohol and Alcoholism, 37*(5), 504-508.

Ainslie, G. (1975). Specious reward: A behavioural theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496.

Ainslie, G. (1992). *Piconomics:* .

Ainslie, G. (2005). Precis of *Breakdown of Will*. *Behavioral and Brain Sciences, 28*, 635-673.

Ainslie, G. (2011). Free will as recursive self-prediction: Does a deterministic mechanism reduce responsibility. In J. Poland & G. Graham (Eds.), *Addiction and Responsibility* (pp. 55-88). Cambridge: MIT Press.

Alcoholics Anonymous World Services. (2001). *Alcoholics Anonymous* (4th ed.). New York: Alcoholics Anonymous World Services.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). London, England: American Psychiatric Publishing.

Anthony, J. C., & Helzer, J. E. (1991). Syndromes of drug abuse and dependence. In L. N. Robins & D. A. Regier (Eds.), *Psychiatric disorders in America: The epidemiological catchment area study* (pp. 116-154). New York: Free Press.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.

Baler, R. D., & Volkow, N. D. (2006). Drug addiction: The neurobiology of disrupted self-control. *Trends in Molecular Medicine, 12*(2), 559–566.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Baumeister, R. F. (2002). Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function. *Self and Idenity, 1*(2), 129-136.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*(5), 1252-1265.

Baumeister, R. F., Heatherton, T., & Tice, D. M. (1994). *Losing Control*. San Diego: Academic Press.

Baylis, F. (2011). "I Am Who I Am": On the Perceived Threats to Personal Identity from Deep Brain Stimulation *Neuroethics*.

Baylis, F. (2012). The self *in situ*: A relational account of personal identity. In J. Downie & J. J. Llewellyn (Eds.), *Being Relational: Reflections on Relational Theory and Health Law*. Vancouver: UBC Press.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience, 8*(11), 1458–1463.

Bennett, S. T. (2002). *Cultural identity and academic achievement among Māori undergraduate university students.* Paper presented at the National Māori Graduates of Psychology Symposium 2002: Making a difference, University of Waikato.

Berridge, K. C., & Robinson, T. E. (1998). What is the Role of Dopamine in Reward: Hedonics, Learning, or Incentive Salience? *Brain Research Reviews, 28*, 308–367.

Best, D., Gow, J., Knox, T., Taylor, A., Groshkova, T., & White, W. (2012). Mapping the recovery stories of drinkers and drug users in Glasgow: Quality of life and its associations with measures of recovery capital. . *Drug and Alcohol Review, 31*, 334–341.

Blascovich, J., Spencer, S., Quinn, D. M., & Steele, C. M. (2001). African Americans and High Blood Pressure: The Role of Stereotype Threat. *Psychological science, 12*(3), 225-229.

Blomqvist, J. (2004). Sweden's "war on drugs" in the light of addicts' experiences. In P. Rosenqvist, J. Blomqvist, A. Koski-Jannes & L. Ojesjo (Eds.), *Addiction and life course* (pp. 139-171). Helsinki: NAD Publication.

Bouhuys, A. L., & van der Meulen, W. R. M. H. (1984). Speech timing measures of severity, psychomotor retardation, and agitation in endogenously depressed patients *Journal of Communication Disorders, 17*, 277-288.

Brand, R. (2012). *Oral Evidence Taken before the Home Affairs Committee on Tuesday 24 April 2012*.

Bratman, M. (1981). Intention and means-end reasoning. *Philosophical Review, 90*(2), 252-265.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M. (1996). Planning and temptation. In L. May, M. Friedman & A. Clark (Eds.), *Mind and Morals. Essays on Cognitive Science and Ethics*. Cambridge, MA: MIT Press.

Bratman, M. (1999). *Faces of Intention*. Cambridge: Cambridge University Press.

Bratman, M. (2007). *Structures of agency*. Oxford: Oxford University Press.

Brehm, J. (1956). Postdecisional Changes in the Desirability of Alternatives. *Journal of Abnormal Psychology, 52*, 384–389.

Brodt, S. E., & Zimbardo, P. G. (1981). Modifying shyness-related social behaviour hrough symptom misattribution. *Journal of Personality and Social Psychology, 41*, 437-449.

Burkley, E. (2008). The role of self-control in resistance to persuasion. *Personality and Social Psychology Bulletin, 34*, 419–431.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*, 267–285.

Camerer, C. F. (2006). Wanting, liking, and learning: Neuroscience and paternalism. *University of Chicago Law Review, 73*(1), 87–110.

Charon, R. (2006). *Narrative Medicine: Honouring the stories of illness*. Oxford: Oxford University Press.

Christman, J. (2009). *The politics of persons: Individual autonomy and socio-historical selves*. Cambridge: Cambridge University Press.

Clark, J., Eno, C., & Guadagno, R. (2011). Southern Discomfort: The Effects of Stereotype Threat on the Intellectual Performance of US Southerners. *Self and Idenity, 10*(2), 248-262.

Clarkson, J. J., Hirt, E. R., Chapman, D. A., & Jia, L. (2011). The impact of illusory fatigue on executive control. Do perceptions of depletion impair working memory capacity? *Social Psychological and Personality Science, 2*(3), 231-238.

Cohen, D., & Handfield, T. (2010). Rational capacities, resolve, and weakness of will. *Mind, 119*, 907-932.

Collins, R. L., & Lapp, W. M. (1991). Restraint and attributions:Evidence of the abstinence violation effect in alcohol consumption. *Cognitive Therapy and Research, 15*(1), 69-84.

Conway, K. P., Compton, W., Stinson, F. S., & Grant, B. F. (2006). Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 67*, 247-257.

Coombs, R. H. (1997). *Drug-impaired professionals*. Cambridge, MA: Harvard University Press.
Cullity, G., & Gerrans, P. (2004). Agency and policy. *Proceedings of the Aristotelian Society, 14*, 317-329.
Dahl, R. (1986). *The Best of Roald Dahl*: Penguin Books Limited.
Degenhardt, L., & Hall, W. (2012). Extent of illicit drug use and dependence, and their contribution to the global burden of disease. *The Lancet, 379*(9810), 55–70.
Dennett, D. (1991). *Consciousness explained*. Boston: Little, Brown and Company.
Dow, M. G., & Craighead, W. E. (1987). Social inadequacy and depression: Overt behavior and self-evaluation processes. *Journal of Social and Clinical Psychology*(5), 99-113.
Ellingsen, T., & Johannesson, M. (2008). Anticipated verbal feedback induces altruistic behaviour. *Evolution and Human Behaviour, 29*, 100-105.
Elster, J. (1999). *Strong feelings: Emotion, addiction and human behaviour*. Cambridge, MA.: MIT Press.
Fingarette, H. (1988). *Heavy drinking: The myth of alcoholism as a disease*. Berkeley: University of California Press.
Flanagan, O. (2011). What is it like to be and addict? In J. Poland & G. Graham (Eds.), *Addiction and Responsibility* (pp. 269-292). Cambridge, MA: MIT Press.
Foddy, B., & Savulescu, J. (2010). A liberal account of addiction. *Philosophy, Psychiatry and Psychology, 17*(1), 1-22.
Frankfurt, H. (1971). Freedom of the WIll and the Concept of a Person. *The Journal of Philosophy, 68*(1), 5-20.
Frankfurt, H. (1978). The Problem of Action. *American Philosophical Quarterly, 15*(2), 157-162.
Friedman, M. (1992). Feminism and modern friendship: Dislocating the community. In Cole & Coultrap-McQuin (Eds.), *Explorations in Feminist Ethics*.
Galaif, E. R., & Newcomb, M. D. (1999). Predictors of polydrug use among four ethnic groups: A 12-year longitudinal study. *Addictive behaviours, 24*, 607-631.
Galaif, E. R., Newcomb, M. D., Vega, W. A., & Krell, R. D. (2007). Protective and risk influences of drug use among a multiethnic sample of adolescent boys. *Journal of Drug Education, 37*, 755-758.
Galinsky, A. D., Wang, C. S., & Ku, G. (2008). Perspective-takers behave more stereotypically. *Journal of Personality and Social Psychology, 95*(2), 404-419.
Gibbons, F. X. (1987). Mild depression and self-disclosure intimacy: Self and others' perceptions. *Cognitive Therapy and Research, 11*, 361—380.
Giesler, R. B., Josephs, R. A., & Swann, W. B. (1996). Self-verification in clinical depression: the desire for negative evaluation. *Journal of Abnormal Psychology, 105*(3), 358-368.
Goffman, E. (1961). *Asylums: Essays in the social situation of mental patients and other inmates*. New York: Doubleday.
Goldie, P. (2000). *The Emotions: A philosophical exploration*. Oxford: Oxford University Press.
Goldie, P. (2005). Imagination and the distorting powers of emotion. *Journal of Consciousness Studies, 12*, 127-139.
Goldie, P. (2007). Dramatic irony and the external perspective. In D. Hutto (Ed.), *Narrative and Understanding*. Cambridge: Cambridge University Press.
Goldie, P. (2012). *The mess inside*. Oxford: Oxford University Press.
Gotlib, I. H., & Robinson, L. A. (1982). Responses to depressed individuals: Discrepancies between self report and observer-rated behavior. *Journal of Abnormal Psychology, 91*, 231-240.
Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*, 769-792.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, memory and cognition, 31*(5), 1121-1133.

Grilo, C. M., & Shiffman, S. (1994). Longitudinal investigationof the abstinense violation effect in binge eaters. *Journal of Consulting and Clinical Psychology, 62*(3), 611-619.

Hagger, M. S., Panetta, G., Leung, C. M., Wong, G. G., Wang, J. C., Chan, D. K., . . . Chatzisarantis, N. L. (2013). Chronic inhibition, self-control and eating behavior: test of a 'resource depletion' model. *Plos One, 8*(10), e76888.

Hasin, D. S., Stinson, F. S., Ogburn, E., & Grant, B. F. (2007). Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry, 64*, 830-842.

Herrnstein, R. J. (1997). *The matching law: papers in psychology and economics*. Cambridge, MA: Harvard University Press.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan, W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioural Decision Making, 6*(3), 149-185.

Heyman, G. (2009). *Addiction: A disorder of choice*. Harvard: Harvard University Press.

Heyman, G. (2013). Addiction: An emergent consequence of elementary choice principles. *Inquiry, 56*(5), 428–445.

Heyman, G., & Dunn, B. (2002). Decision biases and persistent illicit drug use: An experimental study of distributed choice and addiction. *Drug and Alcohol Dependence, 67*(2), 193-202.

Heyman, G., & Gibb, S. P. (2006). Delay discounting in college cigarette chippers. *Behavioural Pharmacology, 17*(8), 669-679.

Hinchman, E. (2003). Trust and diachronic agency. *Nous, 37*(1), 25-51.

Holton, R. (2004). Rational resolve. *Philosophical Review, 113*(4), 507-535.

Holton, R. (2009). *Willing, wanting, waiting*. Oxford: Oxford University Press.

Holton, R., & Berridge, K. C. (2013). Addiction between compulsion and choice. In N. Levy (Ed.), *Addiction and self-control: Perspectives from philosophy, psycholgy, and neuroscience* (pp. 239-268). Oxford: Oxford University Press.

Holton, R., & May, J. (2012). What in the world is weakness of will? *Philosophical Studies, 157*, 341-360.

Hudson, S. M., Ward, T., & France, K. G. (1992). The absintence violation effect in regressed and fixated child moesters. *Annals of Sex Research, 5*(4), 199-213.

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

Inzlicht, M., & Gutsell, J. N. (2007). Running on empty: Neural signals for self-control failure. *Psychological science, 18*(11), 933–937.

Jarvinen, M., & Andersen, D. (2009). Creating Problematic Identities. The Making of the Chronic Addict. *Substance Use & Misuse, 44*, 865–885.

Jaworska, A. (1999). Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value. *Philosophy & Public Affairs, 28*(2), 105-138.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological science, 16*(3), 175-179.

Johnson, J., Cohena, P., Kasena, S., & Brook, J. (2006). Dissociative disorders among adults in the community, impaired functioning, and axis I and II comorbidity. *Journal of Psychiatric Research, 40*(2), 131-140.

Johnson, W. G., Schlundt, D. G., Barclay, D. R., Carr-Nangle, R. E., & Engler, L. B. (1995). A naturalistic functional analysis of binge eating. *Behavior Therapy, 26*(1), 101-118.

Jones, K. (2003). Emotion, weakness of will, and normative conception of agency. *Royal Institute of Philosophy Supplement, 52*, 181-200.

Jones, K. (2008). How to change the past. In K. Atkins & C. Mackenzie (Eds.), *Practical identiy and narrative agency* (pp. 269-288). London: Routledge.

Kahan, D., Polivy, J., & Herman, C. P. (2003). Conformity and dietary disinhibition: a test of the ego strength model of self-regulation. *International Journal of Eating Disorders, 33*(2), 165-171.

Karadag, F., Vedat, S., Tamar-Gurol, D., Evren, C., Karagoz, M., & Erkiran, M. (2005). Dissociative disorders among inpatients with drug or alcohol dependency. *Journal of Clinical Psychiatry, 66*, 1247-1253.

Karasaki, M., Fraser, S., Moore, D., & Dietze, P. (2013). The place of volition in addiction: Differing approaches and their implications for policy and service provision. *Drug and Alcohol Review, 32*, 195-204.

Karniol, R., & Miller, D. T. (1983). Why Not Wait? A Cognitive Model of Self-imposed Delay Termination. *Journal of Personality and Social Psychology, 45*(4), 935-942.

Kennett, J. (2001). *Agency and Responsibility. A Common Sense Moral Psychology*. Oxford: Clarendon Press.

Kennett, J. (2013a). Addiction, choice, and disease: How voluntary is voluntary action in addiction? In N. Vincent (Ed.), *Neuroscience and Legal Responsibility* (pp. 257-278). Oxford: Oxford University Press.

Kennett, J. (2013b). Just say no? Addiction and the elements of control. In N. Levy (Ed.), *Addiction and Self-Control* (pp. 144-164). Oxford: Oxford University Press.

Kennett, J., Fry, C., & Matthews, S. (commenced 2010). Addiction, moral identity and moral agency. Sydney and Melbourne, Australia: Australian Research Council.

Kennett, J., Matthews, S., & Snoek, A. (2013). Pleasure and Addiction. *Frontiers in Psychiatry, 4*(117). doi: 10.3389/fpsyt.2013.00117

Kennett, J., & McConnell, D. (2013). Explaining addiction: How far does the reward account of motivation take us? , *56*(5), 470-489.

Kennett, J., & Smith, M. (1996). Frog and Toad Lose Control. *Analysis, 56*(2), 63–73.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*, 593-602.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62*, 617-627.

Kirby, K. N. (1997). Bidding on the future: evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General, 126*, 54-70.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128*(1), 78-87.

Korsgaard, C. (2008). Self-constitution in the ethics of Plato and Kant *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

Koski-Jännes, A., Hirschovits-Gerz, T., & Pennonen, M. (2012). Population, Professional, and Client Support for Different Models of Managing Addictive Behaviors. *Substance Use and Misuse, 47*, 296–308.

Kudadjie-Gyamfie, E., & Rachlin, H. (1996). Temporal patterning in choice among delayed outcomes. *Organizational Behaviour and Human Decision Processes, 65*, 61-67.
Lang, C. (2013). Postmodern assumptions in the treatment of addiction [Press release]. Retrieved from https://www.youtube.com/watch?v=Bx0dOTjaEGk
Laudet, A. (2007). What does recovery mean to you? Lessons from the recovery experience for research and practice. *Journal of Substance Abuse Treatment, 33*, 243-256.
Leshner, A. (1997). Addiction is a brain disease, and it matters. *Science, 278*(5335), 45–47.
Levy, N. (2006). Autonomy and addiction. *Canadian Journal of Philosophy, 36*(3), 427-446.
Levy, N. (2007). *Neuroethics*. Cambridge: Cambridge University Press.
Levy, N. (2011a). Addiction, responsibility, and ego depletion. In J. Poland & G. Graham (Eds.), *Addiction and Responsibility* (pp. 89-111). Cambridge: MIT.
Levy, N. (2011b). Resisting 'weakness of the will'. *Philosophy and Phenomenological Research, 82*(1), 134-155.
Levy, N. (2014). Addiction as a disorder of belief. *Biology & Philosophy*, 1-19. doi: 10.1007/s10539-014-9434-2
Lewis, M. (2012). Why addiction is NOT a brain disease. *Mind the Brain.* Retrieved 9 September, 2013, from http://blogs.plos.org/mindthebrain/2012/11/12/why-addiction-is-not-a-brain-disease/
Lloyd, G. (1993). *Being in Time: Selves and Narrators in Philosophy and Literature*. London: Routledge.
Lyons, M. J., Bar, J. L., Panizzon, M. S., Toomey, R., Eisen, S., Xian, H., & Tsuang, M. T. (2004). Neuropsychological consequences of regular marijuana use: A twin study. *Psychological Medicine, 34*, 1239-1250.
MacIntyre, A. (1984). *After Virtue: A study in moral theory* (2nd ed.). New York: Oxford University Press.
Mackenzie, C. (2008). Imagination, identity and self-transformation. In K. Atkins & C. Mackenzie (Eds.), *Practical Identity and Narrative Agency* (pp. 121-145). New York: Routledge.
Major, B., Spencer, S., Schmader, T., Wolf, C., & Crocker, J. (1997). Coping with negative stereotypes about intellectual performances: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*(1), 34–50.
Marlatt, G., & Gordon, J. (1980). Determinants of relapse: Implications for the maintenance of behaviour change. In P. O. Davidson & S. M. Davidson (Eds.), *Behavioural medicine: Changing health lifestyles* (pp. 410-452). Elmsford: Pergamon Press.
Mazur, J. E. (2001). Hyperbolic value addition and general models of animal choice. *Psychological review, 108*, 96-112.
McLellan, A., McKay, J., Forman, R., Cacciola, J., & Kemp, J. (2005). Reconsidering the evaluation of addiction treatment: from retrospective follow-up to concurrent recovery monitoring. *Addiction, 100*(4), 447-458.
Mele, A. R. (1997). Underestimating self-control: Kennett and Smith on Frog and Toad. *Analysis, 57*(2), 119–123.
Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological review, 106*, 3-19.
Millgram, E. (1997). *Practical Induction*. Harvard: Harvard University Press.
Mischel, H. N., & Mischel, W. (1983). The development of childrens knowledge of self-control strategies. *Child Development, 54*, 603-619.
Mischel, W., Shoda, Y., & Rodriguez, M. (1992). Delay of gratification in children. In G. Loewenstein & J. Elster (Eds.), *Choice over time*. New York: Russell Sage Foundation.

Moé, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual differences, 19*(1), 21-27.
Monterosso, J., & Ainslie, G. (2009). The picoeconomic approach to addictions: Analyzing the conflict of successive motivational states. *Addiction Research and Theory, 17*(2), 115-134.
Muraven, M., Baumeister, R. F., & Tice, D. M. (1998). Self-Control as a Limited Resource: Regulatory Depletion Patterns. *Journal, of Personality and Social Psychology, 74*, 774–789.
Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal Improvement of Self-Regulation Through Practice: Building Self-Control Strength Through Repeated Exercise. *The Journal of Social Psychology, 139*, 446-457.
Neale, J. (2002). *Drug users in society*. New York: Palgrave.
Nelson, H. (2001). *Damaged Identities, Narrative Repair*. Ithaca, NY: Cornell University Press.
Newcomb, M. D., Vargas-Carmona, J., & Galaif, E. R. (1999). Drug problems and psychological distress among a community sample of adults: Predictors, consequences of confound? *Journal of Community Psychology, 27*, 405-429.
Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Reischer (Ed.), *Essays in honour of Carl G. Hempel*. Dordrecht: Reidel.
Osbourne, J. W. (1995). Academics, self-esteem, and race:Alook at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin,, 21*, 449–455.
Prochaska, J. O., Norcross, J. C., & DiClemente, C. C. (1994). Changing for good: a revolutionary six-stage program for overcoming bad habits and moving your life positively forward. New York: Avon.
Robins, L. N., & Regier, D. A. (1991). *Psychiatric disorders in America: The epidemiological catchment area study*. New York: Free Press.
Robinson, T. E., & Berridge, K. C. (2003). Addiction. *Annual Review of Psychology, 54*, 25–53.
Ross, C., Kronson, J., Koensgen, S., Barkman, K., Clark, P., & Rockman, G. (1992). Dissociative comorbidity in 100 chemically dependant patients. *Hospital and Community Psychiatry, 43*(8), 840-842.
Samuelson, P. (1938). A Note on the Pure Theory of Consumers' Behaviour. *Economica, 5*, 61-71.
Sapphire. (2013). It should all be about the person. 2013, from http://www.recoverystories.info/sapphires-recovery-story-it-should-all-be-about-the-person/
Sartwell, C. (2008a). Addiction and authorship. 2013, from http://www.crispinsartwell.com/addict2.htm
Sartwell, C. (2008b). Detritus. Retrieved 20th August, 2012
Schechtman, M. (1996). *The Constitution of Selves*. Ithaca: Cornell University Press.
Schechtman, M. (2001). Empathic access: The Missing Ingredient in Personal Identity. *Philosophical Explorations, 4*(2), 95-111.
Schechtman, M. (2007). Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View. *Royal Institute of Philosophy Supplement, 82*(60), 155-178.
Schechtman, M. (2008). Diversity in unity: practical unity and personal boundaries. *Synthese, 162*, 405–423.
Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard: Harvard University Press.
Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. (1976). Facial expression and imagery in depression: An electromyographic study *Psychosomatic Medicine, 38*, 337-347.

Shaffer, H. J. (1997). The psychology of stage change. In J. H. Lowinson, P. Ruiz, R. B. Millman & L. G. Langrod (Eds.), *Substance abuse: a comprehensive textbook (3rd ed.)* (pp. 100–106). Baltimore: Williams and Wilkins.

Shaffer, H. J., & Simoneau, G. (2001). Reducing resistance and denial by exercising ambivalence during the treatment of addiction. *Journal of Substance Abuse Treatment, 20*(1), 99–105.

Sharps, M. J., Price, J. L., & Williams, J. K. (1994). Spatial cognition and gender: Instructional

and stimulus influences on mental image rotation performance. *Psychology of Women Quarterly, 18*, 413-425.

Shi, J., Li, S., Zhang, X., Wang, X., Foll, B., Zhang, X., . . . Lu, L. (2009). Time-dependent neuroendocrine alterations and drug craving during the first month of abstinence in heroin addicts. *The American Journal of Drug and Alcohol Abuse, 35*(5), 267-272.

Shiffman, S., Hickcox, M., Paty, J. A., Gnys, M., Kassel, J. D., & Richards, T. J. (1997). The absintence violation effect following smoking lapses and temptations. *Cognitive Therapy and Research, 21*(5), 497-523.

Spanier, C. A., Shiffman, S., Maurer, A., Reynolds, W., & Quick, D. (1996). Rebound following failure to quit smoking: The effects of attributions and self-efficacy. *Experimental and Clinical Psychopharmacology, 4*(2), 191-197.

Spencer, S., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Sripada, C. (2014). How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind. *Nous, 48*(1), 41-74.

Steele, C. M., & Aronson, J. (1995). Contending with a stereotype: African-American intellectual test performance and stereotype threat. *Journal of Personality and Social Psychology,, 69*, 797–811.

Stilen, P., Carise, D., Roget, N., & Wendler, A. (2007). *Treatment planning M.A.T.R.S. Utilizing the Addiction Severity Index (ASI) to make required data collection useful*. Missouri: Mid-America Addiction Technology Transfer Center.

Stinson, F. S., Grant, B. F., Dawson, D. A., Ruan, W. J., Huang, B., & Saha, T. (2005). Comobidity between DSM-IV alcohol and specific drug use disorders in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug and Alcohol Dependence, 80*, 105-116.

Stinson, F. S., Grant, B. F., Dawson, D. A., Ruan, W. J., Huang, B., & Saha, T. (2006). Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: Results from the National Epidemiological Survey on Alcohol and Related Conditions. *Alcohol Research and Health, 29*, 94-106.

Substance Abuse and Mental Health Services Administration. (2013). *Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings*. Rockville, Maryland: Substance Abuse and Mental Health Services Administration.

Swann, W. B. J., Wenzlaff, R. M., Krull, D. S., & Pelham, B. W. (1992). Allure of negative feedback: self-verification strivings among depressed persons. *Journal of Abnormal Psychology, 101*(2), 293-306.

Swora, M. G. (2001). Narrating Community: The Creation of Social Structure in Alcoholics Anonymous through the Performance of Autobiography. *Narrative Inquiry, 11*(2), 363-384.

Szalavitz, M. (2012). The Beginning of the End of the Abstinence Rule? Retrieved 14/10/2013, from http://www.thefix.com/content/hazelden-maintenance-suboxone-opiate-painkiller8546?page=all

The National Institute on Drug Abuse. (2012). *Principles of Drug Addiction Treatment* (3rd ed.).
Tolkien, J. R. R. (1966). *The lord of the rings* (2nd ed.). London: Allen & Unwin.
Toomey, R., Lyons, M. J., Eisen, S., Xian, H., Chantarujikapong, S., Seidman, L. J., . . . Tsuang, M. T. (2003). A twin study of the neuropsychological consequence of stimulant abuse. *Archives of General Psychiatry, 60*, 303-310.
Tsuang, M. T., Bar, J. L., Harley, R. M., & Lyons, M. J. (2001). The Harvard twin study of substance abuse: What we have learned. *Harvard Review of Psychiatry, 9*, 267-279.
U.S. Department of Health and Human Services. (2014). *The Health Consequences of Smoking — 50 Years of Progress. A Report of the Surgeon General*. Rockville, Maryland: U.S. Department of Health and Human Services.
Velleman, J. D. (1989). *Practical Reflection*. Princeton: Princeton University Press.
Velleman, J. D. (2000). Well-being and time *The possibility of practical reason* (pp. 56-84). Oxford: Oxford University Press.
Velleman, J. D. (2002). Motivation by ideal. *Philosophical Explorations, 5*(2), 89-103.
Velleman, J. D. (2003). Narrative explanation. *Philosophical Review, 112*(1), 1-25.
Velleman, J. D. (2005). Self as narrator. In J. Christman & J. Anderson (Eds.), *Autonomy and the Challenges to Liberalism: New Essays*. Cambridge: Cambridge University.
Vohs, K., Baumeister, R. F., Schmeichel, B., Twengy, J., Nelson, N., & Tice, D. M. (2008). Making choices impairs subsequent self-control. *Journal of Personality and Social Psychology, 94*(5), 883-898.
Vohs, K., & Heatherton, T. (2000). Self-regulatory failure: a resource-depletion approach. *Psychological science, 11*(3), 249-254.
Vohs, K., & Schooler, J. (2008). The value of believeing in free will. *Psychological science, 19*(1), 49-54.
Volkow, N. D., & Li, T. K. (2005). The neuroscience of addiction. *Nature Neuroscience, 8*(11), 1429–1430.
Vuchinich, R. E., & Simpson, C. A. (1998). Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology, 6*(3), 292-305.
Wang, J., Novemsky, N., Dhar, R., & Baumeister, R. F. (2010). Tradeoffs and depletion in choice. *Journal of Marketing Research, 47*(5), 910-919.
Ward, T., Hudson, S. M., & Marshall, W. L. (1994). The absintence violation effect in child molesters. *Behaviour Research and Therapy, 32*(4), 431-437.
Warner, L. A., Kessler, R. C., Hughes, M., Anthony, J. C., & Nelson, C. B. (1995). Prevalence and correlates of drug use and dependence in the United States. *Archives of General Psychiatry, 52*, 219-229.
Watson, G. (1975). Free Agency. *The Journal of Philosophy, 72*(8), 205-220.
Watson, G. (1987). Free action and free will. *Mind, 96*, 154-172.
We Do Recover. (2010, 3/3/2010). Ambivalence in addicted and alcoholic patients. 2013, from http://wedorecover.com/articles/article/ambivalence-in-addicted-and-alcoholic-patients.html
Weddington, W., Barry S. Brown, B., Haertzen, C., Cone, E., Dax, E., Herning, R., & Michaelson, B. (1990). Changes in Mood, Craving, and Sleep During Short-term Abstinence Reported by Male Cocaine Addicts. A Controlled, Residential Study. *Archives of General Psychiatry, 47*(9), 861-868.
West, R. (2006). *Theory of addiction*. Oxford, UK: Blackwell.
Wheeler, S. C., Brinol, P., & Hermann, A. D. (2007). Resistance to persuasion as self-regulation: ego depletion and its effects on attitude change processes. *Journal of Experimental Social Psychology, 43*(1), 150-156.

White, M., & Epston, D. (1990). *Narrative Means to Therapeutic Ends*. New York: W. W. Norton and Company.
White, W. (2007). Addiction recovery: Its definition and conceptual boundaries. *Journal of Substance Abuse Treatment, 33*(3), 229-241.
Wilson, R. (Writer). (2012). Russell Brand: From Addiction to Recovery. In R. Wilson (Producer). England: BBC3.
Wittgenstein, L. (2001). *Philosophical Investigations* (3rd ed.). Oxford: Blackwell Publishing.
Wollheim, R. (1984). *The thread of life*. Cambridge: Cambridge University Press.
World Health Organization. (2011). *Global status report on alcohol and health*. Switzerland: World Health Organization.