Reports from the Department of Philosophy
Vol. 48

An Access Theory of Self-Control

Essays on Self-Control, ADHD, And Stimulant Medication

Polaris Koi

University of Turku
Finland

Copyright © 2021 Polaris Koi

SERIES EDITORS:
Olli Koistinen
Juha Räikkä

Department of Philosophy
University of Turku
FI-20014 Turku
Finland

ISSN 1457-9332
ISBN 978-951-29-8674-3

Written by
Polaris Koi
University of Turku, Faculty of Social Sciences, Department of Philosophy, Contemporary History and Political Science

Supervised by
Professor Juha Räikkä
University of Turku
Doctor Jukka Varelius
University of Turku

Reviewed by
Professor Neil Levy
Macquarie University; University of Oxford
Assistant Professor Teemu Toppinen
Tampere University

Opponent
Professor Neil Levy
Macquarie University; University of Oxford

Chairperson (custos)
Professor Juha Räikkä
University of Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu *Turnitin Originality Check* -järjestelmällä.
The originality of this dissertation has been checked in accordance with the University of Turku quality assurance system using the *Turnitin Originality Check* service.

Acknowledgments

I view philosophy as a community pursuit rather than as something occurring solely inside the minds of brilliant individuals. Accordingly, there are many people, philosophers and otherwise, whose presence, company and counsel have enabled and enriched the work leading up to this dissertation.

My foremost thanks go to my advisors, Professor Juha Räikkä and Dr. Jukka Varelius. Professor Räikkä's encouragement and practical insight, and Dr. Varelius' careful comments and criticisms have been a touchstone throughout the process of drafting this dissertation. Their patience in bearing with me as I gradually honed (or, under a less charitable description, flippantly kept changing) the scope of this dissertation deserves an honourable mention.

I am deeply grateful to the external examiners of this dissertation, Professor Neil Levy and Teemu Toppinen, for their detailed engagement with my work. Their insightful comments and suggestions have considerably improved this dissertation.

A fledgling graduate student could ask for nothing better than sharing an office with someone who at that time was about to defend, generating an ad hoc mentorship. Such a relationship emerged in 2015 between me and Susanne Uusitalo, giving rise to an early collaboration in jointly running the "Improving ourselves as moral agents: The ethics and neuroscience of self-control" reading group and workshop series in spring 2016, enabled by funding from the John Templeton foundation via the Philosophy and Science of Self-Control research project. The seeds of the ideas developed in this dissertation were planted during that spring, and my utmost gratitude goes to the keynotes and presenters of that workshop series as well as to participants of our group including Ninni Suni, Jarno Tuominen, and Nils Sandman. The first article of this dissertation emerged out of these early collaborations with Tuominen and Uusitalo, whose diligence, intellect and enthusiasm makes them the ideal colleagues and co-authors.

The (bio)ethical side of this dissertation has been enriched by interactions with the lively bioethical community in Turku. My heartfelt thanks to everyone at the International Chair in Bioethics Finland unit and the Center for the Study of Bioethics Finland unit.

In 2017, I visited the UEHIRO centre for applied ethics at Oxford, and I wish to thank everyone at the centre for hosting me for a month during 2017. The access to the vibrant philosophical community at Oxford that this period enabled made this a period of intense learning and growth.

From 2016 to 2019, I had the tremendous luck of being able to participate in the Genetics and Human Agency project as a Junior Investigator. Initially, I had no idea what I was getting into, but during these three transformative years, the sustained dialogues with the behavioural geneticists, psychiatrists, psychologists, and philosophers of science involved in this project expanded my thinking beyond ethics and into the philosophy of science. I am beyond grateful to Professor Eric Turkheimer for taking a chance with me. Thanks also to the friends and colleagues I found at this project, in particular M. Carmela Epright, Kathryn Tabb, Lucas Matthews, Matt Lebowitz, Colin DeYoung, and James Morandini.

I have benefited from the mentorship and encouragement of many senior colleagues. I wish to thank Elisa Aaltola, Marko Ahteensuu, Valtteri Arstila, Greg Bognar, Lilian O'Brien, Markus Eronen, Matti Häyry, Antti Kauppinen, Jaakko Kuorikoski, Eerik Lagerspetz, Caterina Marchionni, Hille Paakkunainen, and Helena Siipi for many invaluable discussions.

The mentorship of senior philosophers notwithstanding, perhaps the most instrumental feature in finishing a dissertation is a rich network of peers on which one can depend on for support and encouragement. Even more enriching has been that these peers have spanned analytic, continental, and historical approaches to philosophy. Accordingly, my heartfelt thanks go to all my friends and colleagues (and my heartfelt apologies, to those I fail to mention). Of those not mentioned above, I would especially like to thank Johanna Ahola-Launonen, Tuukka Brunila, Heidi Haanila, Maria Hämeen-Anttila, Minna-Kerttu Kekki, Kaisa Kärki, Jari Kärkkäinen, Guido Löhr, Joonas Martikainen, Sanna Mattila, Laura Puumala, Mikko Puumala, Pieta Päällysaho, Anne Salminen, Elmarie Venter, and Jaana Virta.

On this transformative journey, I have benefited enormously from discussions with a group of brilliant academics in philosophy and beyond. The friendship and counsel of Alex Burns, Christopher Merwin, Robin Norris, and Mark Roblee has helped me cultivate philosophy as a way of life rather than just a topic of inquiry.

I thank the Finnish Cultural Foundation, the John Templeton Foundation (via the GHA project), and the TOP foundation for financial support during the dissertation process.

Finally, heartfelt thanks to my family, who have been incredibly supportive of this process; in particular, to my aunt, Sirpa Kokko, for a fount of wisdom, and my son, Aale, for bringing me joy and purpose.

I dedicate this dissertation to my mother, Paula Huhtanen, who has been unwavering in encouraging me to pursue my passions. If the complex causal pathways of gene-environment interaction have given me half of her commitment and creative drive, I can consider myself lucky.

Abstract

This dissertation consists of four articles and an introductory essay, the overarching aim of which is to investigate the nature of self-control and its implications.

I defend a value-deflationist definition of self-control, according to which self-control is that which is enacted to align one's behaviour with intention in the face of competing motivation. This radical revision of the concept of self-control, which under its received denotation is associated with motivational hierarchies under various guises, allows for a conception of self-control that is consistent even in the absence of clear value hierarchies.

For my account, self-control is defined by its behavioural function, not by its reliance on any specific cognitive or neural process. As a result, self-control can be realized by a large range of functionally equivalent processes. Some such processes are intrapsychic, such as inhibition, rehearsal of reasons, and construal. Others are situational, such as environmental selection and modification as well as social processes of self-control.

Not all self-control processes are equally available for all agents, however. Some processes that are heavily reliant on executive functioning, such as inhibition and rehearsal of reasons, are unavailable for agents with executive dysfunction. For other self-control processes, social and economic constraints delimit their availability. It then becomes of utmost importance that agents are aware of such means of self-control that are a good fit for the agent and her situation.

In this dissertation I analyse individual differences in self-control in terms of differences in access to self-control behaviours, articulating criteria for access to self-control and demonstrating how epistemic and practical constraints contingently modulate access to self-control in a manner that results in aggravated individual differences in self-control.

I also discuss various implications of the access theory of self-control. implications span assessments of responsibility for failures of self-control; accounting for individual differences in self-control, including when these differences are connected to diagnoses such as Attention Deficit / Hyperactivity Disorder; and the prospects of self-control neuroenhancement.

Keywords: self-control, willpower, agency, access, ADHD, stimulant medication, neuroethics, philosophy of psychology, responsibility

Tiivistelmä

Väitöskirja koostuu neljästä artikkelista ja johdantoesseestä, joiden päätavoitteena on tarkastella itsekontrollin luonnetta ja sen implikaatioita.

Puolustan arvodeflationistista itsekontrollin määritelmää, jonka mukaan itsekontrolli on sitä mitä toimija tekee mukauttaakseen käyttäytymisensä intentioon motivaatiokonfliktin kohdatessaan. Kyseessä on merkittävä revisio tavanomaiseen itsekontrollikäsitykseen, joka liittää itsekontrolliin arvohierarkioita eri muodoissaan. Määritelmäni mahdollistaa käsitteen yhdenmukaisen käytön silloinkin, kun selvää arvohierarkiaa ei ole.

Koska itsekontrollia ei määritä mikään tietty kognitiivinen tai neuraalinen prosessi vaan sen käyttäytymistä säätelevä rooli, suuri joukko funktionaalisesti ekvivalentteja prosesseja voi toimia itsekontrollina. Jotkin näistä prosesseista ovat mielensisäisiä, kuten inhibitio, perusteiden kertaus ja konstruointi. Toiset ovat tilannekohtaisia toimia, kuten ympäristön valikointi ja muokkaus sekä sosiaaliset itsekontrolliprosessit.

Kaikki itsekontrolliprosessit eivät kuitenkaan ole yhtäläisesti jokaisen toimijan ulottuvilla. Vahvasti toiminnanohjaukseen nojautuvat prosessit kuten inhibitio ja perusteiden kertaus eivät ole niiden toimijoiden saatavilla, joilla on toiminnanohjauksen häiriöitä. Toisten itsekontrolliprosessien saatavuutta rajoittavat puolestaan sosiaaliset ja ekonomiset tekijät. Siksi itsekontrollikyvyn kannalta on tärkeää, että toimija on tietoinen itsekontrollin keinoista, jotka soveltuvat hänen tilanteeseensa.

Väitöksessä esitän yksilöllisten erojen itsekontrollissa riippuvan itsekontrollitoimien saavutettavuudesta. Selvennän saavutettavuuden toteutumisen ehtoja ja osoitan, että episteemiset ja käytännön rajoitteet kontingentisti rajoittavat itsekontrollin saavutettavuutta tavalla, joka kärjistää yksilöllisiä eroja itsekontrollissa.

Käsittelen myös joitakin esittämäni itsekontrollin saavutettavuusteorian implikaatioita. Näihin implikaatioihin lukeutuvat vastuun arviointi itsekontrollin petettyä; yksilöllisten itsekontrollierojen selittäminen mukaan lukien silloin, kun nämä erot liittyvät psykiatrisiin diagnooseihin kuten aktiivisuuden ja tarkkaavuuden häiriöön (ADHD); sekä itsekontrollin neurobiologisen parantelun mahdollisuus.

Avainsanat: itsekontrolli, tahdonvoima, toimijuus, saavutettavuus, ADHD, stimulanttilääkitys, neuroetiikka, psykologian filosofia, vastuu

Table of contents

1. Introduction 1
   1.1. The broader context of this dissertation 1
   1.2. Aims and claims 3
   1.3. ADHD and self-control 5
   1.4. Remarks concerning approach and methodology 6
   1.5. The structure of the dissertation 7
2. A long farewell to internalism about self-control 10
   2.1. Norms and prudence: divided mind accounts 10
      2.1.1. Problems for divided minds 15
   2.2. Willpower accounts 22
      2.2.1. Problems for willpower accounts 26
   2.3. Complex models of self-control 28
3. Against willpower: the externalist challenge 36
   3.1. Radical externalism about self-control 37
      3.1.1. Defining self-control as delayed gratification processes? 43
      3.1.2. Problems for externalism about self-control 46
   3.2. The temporal dimension of self-control 50
4. Revising our concept of self-control 55
   4.1. Value-deflationist self-control 55
   4.2. What kind of a kind is self-control? 64
5. The access theory of self-control 68
   5.1. Self-control and access 68
   5.2. When access is limited: the case of ADHD 71
   5.3. Notes on self-control neuroenhancement and stimulant medication 74
6. Conclusion 78

Bibliography 80

Summary of the original articles 87

Original articles

1. Self-control in responsibility enhancement and criminal rehabilitation
2. Born which way? ADHD, Situational self-control, and responsibility
3. Accessing self-control
4. Towards informed user decisions about pharmacological cognitive enhancement

1. Introduction

The contexts in which we fail and succeed in self-control are countless. The importance of self-control for most life plans is evident in the psychological literature: a cohort study by Moffitt et al. demonstrates that "childhood self-control predicts physical health, substance dependence, personal finances, and criminal offending outcomes" (Moffitt et al. 2011: 2693). These outcomes match a widespread intuition about the purpose of self-control: in Baumeister, Vohs & Tice's view, self-control is used "for altering one's own responses, especially to bring them into line with standards such as ideals, values, morals, and social expectations, and to support the pursuit of long-term goals" (Baumeister, Vohs & Tice 2007: 351). In short, we often think of self-control in terms of instances where there is a clear moral or prudential hierarchy between the competing courses of action, and where self-control is used to conform behaviour to the service of the loftier or more valuable aim. To be capable of self-control thus is paramount for autonomy (Mele 1990) and for a good life. Yet while all agree that resisting impulses and aligning behaviour with intended goals have something to do with self-control, neither philosophers nor empirical mind scientists presently have a consensus about what self-control is, exactly.

In this dissertation, I ask three interrelated questions about self-control. These are the following:

1) What is self-control?
2) What explains individual differences in self-control?
3) What can we learn about self-control by looking at the case of Attention Deficit Hyperactivity Disorder (ADHD)?

I furthermore discuss the extent to which pharmacological means, i.e. the use of stimulant medication as is used for the management of ADHD, are a feasible and permissible means of improving self-control.

1.1. The broader context of this dissertation

Discussions concerning the nature of self-control have been subject to two persistent presuppositions. First, debates about the nature of self-control have their roots in the problem of akrasia (weakness of will). The problem of akrasia, briefly put, is: assuming that agents choose to do what they judge best to do, how can it be possible that agents sometimes do not do what they

judge best? As a result of this legacy, self-control has been taken to be deeply interconnected with akrasia, and therefore with evaluating and ranking the options between which we choose what to do.

Second, debates about the nature of self-control, while asking whether akrasia is possible, have long taken for granted that it, along with the other faculties of the mind, is internal to the self. This view can be called internalism about self-control. For a long time, self-control was seen as something happening within the soul; in contemporary, materialistic times, it is seen as something happening within the skull. These two notions – that of failures of self-control as entailing a breakdown of rationality, and that of self-control as agent-internal – have only recently become matters of contention.

One source for pressure towards then intentional breakdown approach is a growing apprehension that humans are nonideal agents[1]. Our judgments about what to do, far from resulting only from an abstract all-things-considered calculus, are instead impacted by a plethora of factors that are not commonly conceived of as rational grounds for judgments, ranging from mood to cognitive biases, somatic states, and neural mechanisms. This has made space for alternative explanations of self-control.

The received notion that motivational hierarchy is intimately connected to self-control, where good self-control is associated with forgoing smaller rewards or desires in order to pursue ones that are larger or all things considered better (e.g., Davidson 1969, Ainslie 2001, Mele 1994) has been fleshed out in various ways. Such accounts include descriptions of self-control as a non-active mechanism (Kennett & Smith 1996), as a mental action done by one part of the mind on the other (e.g., Sripada 2014), or as a mental power or resource that can be strengthened but also depleted (Holton 2003). Note that all these approaches continue to describe self-control as happening inside the mind and brain of the individual. All these are examples of looking into the behaviours we typically label as self-control (or as failures thereof), and asking, what underlying mechanism, trait, capacity, or motivational structure unites these behaviours? That these underlying features are agent-internal – that is, existing within the agent, not in the environment – has often been taken as granted.

Recently, internalism has been increasingly questioned as research into situated cognitive science has gained momentum. Early exemplars of this movement were defenders of research into embodied cognition, who argue that cognitive processes are not only caused but also deeply

[1] Examples of ideal accounts of intending and acting can be found in, e.g., Audi (1993), Davidson (1969), Ainslie (2001). For example, Ainslie, while criticizing one form of ideal agency, substitutes it with another – an abstract calculus involving steep temporal discounting that, for Ainslie, agents in fact follow.

constituted by the physical features of the organism outside the nervous system (e.g., Varela, Thompson & Rosch 1991). They gained an ally in Clark & Chalmers' (1998) extended mind thesis, which argues that agent-external objects and features can sometimes act as de facto constituents of memory and belief. They argued for their extended mind thesis using the example of a person whose notebook was not just an *aid* to his memory but served *as* his memory. While the Clark and Chalmers paper was not the sole root of an externalist philosophy of mind, their paper opened the floodgates for a new way of thinking about cognition as not simply happening inside the skull. Instead, externalist philosophers of mind argued, at least some features of the mind, rather than happening within the nervous system alone, were extended into the physical organism as a whole, as well as to the individual's immediate environment. Philosophers of mind grew increasingly interested in the myriad ways the mind is not just impacted by but realized by factors beyond the brain. Even as the extended mind thesis gained critics, externalist philosophy of mind has continued to evolve. (See, e.g., Gallagher 2018, Harris 2020.)

Concurrently with this trend in philosophy of mind, empirical research into self-control has been increasingly engaged in the ways environmental and situational factors impact self-control. Inspired by this turn in the science of self-control, some philosophers have embraced externalism about self-control (Levy 2017, Vierkant 2014). While positing externalism about self-control is not novel to this dissertation, it remains an emerging approach that merits further philosophical study.

In brief, the way philosophers (and scientists) think about self-control is currently in the process of shifting. It is into this context that I enter with this dissertation.

1.2. Aims and claims

The ambition of this dissertation is to add to the multidisciplinary debates concerning the nature of self-control by approaching the problem of self-control starting not from intention but from behaviour. In order to probe whether there is an underlying breakdown on the level of intending that unites failures in self-control, I look at the various behaviours we ordinarily classify as self-control successes or as self-control failures, paying particular attention to individual differences in self-control, such as in the clinical context.

I defend the following four claims:

1) That self-control is best understood as a group of behaviours unified by their behavioural function: self-control is that which is enacted to align one's behaviour with intention in the face of competing motivation.

2) That self-control is best understood in externalist terms, that is, as being in part constituted by the agent's environment.

3) That individual differences in self-control result from differences in access to self-control behaviours.

4) That appreciating the role of the environment in self-control has important implications, e.g., for assessments of moral responsibility, and for the efficacy and permissibility of self-control neuroenhancement, including within the criminal justice system.

Out of these claims, claims one and three – and most of the scope of claim four – are novel to my thesis. Even outside the intentionality-centric approach, self-control is typically thought of in decidedly internalist terms, such as, as a part of the executive functions, which are behaviour-guiding and cognitive functions, associated with the prefrontal cortex of the brain. As a result, these claims require a thorough and empirically informed defence.

In defending these claims, my attention is especially directed to two problems. These are defining self-control and explaining individual differences in self-control, i.e., that some persons are more self-controlled than others. I present two novel contributions to understanding self-control. The first of these is what I term the *value-deflationist definition* of self-control, which defines self-control behaviourally without reference to a value hierarchy among the conflicting motives.

The second novel contribution is an account of individual differences in self-control. I argue for what I term an *access theory of self-control*, which posits that individual differences in self-control result from differences in access to self-control behaviours. Disadvantage on multiple axes of social, economic, and political disadvantage is a major determinant of barriers to self-control behaviours.

While claim two is not novel to my dissertation, the work I do with this claim involves a novel synthesis of views, yielding both theoretical and practical insight into the philosophy and science of self-control.

I furthermore offer novel contributions to ethics. These include discussion of responsibility assessments for failures of self-control, such as in the case of ADHD; and novel discussion concerning self-control neuroenhancement and the role of stimulant medication in improving cognitive function, including the functions underlying some of self-control.

1.3. ADHD and self-control

The primary example I use to probe individual differences in self-control is that of Attention Deficit/Hyperactivity Disorder (ADHD), a neurodevelopmental disorder whose heritability is estimated at around 80%. Many of the characteristic symptoms of ADHD are associated with problems in self-control. ADHD is also associated with deficits in executive functioning (Barkley 1997, Cortese & Castellanos 2015).

There are multiple reasons for this choice of case. First, compared to traditional cases in philosophy of action, such as Frankfurt's unwilling addict (Frankfurt 1971), the case of ADHD is both more complex and more rewarding because the sorts of failures in self-control that burden persons with ADHD extend to many domains of life, from academic and occupational functioning to interpersonal relationships. While addiction has adverse effects across various domains of life, the failures of self-control in addiction (as they are typically conceived) are, by and large, failures in avoiding the specific addictive behaviour, whereas in ADHD there is a larger scope of behaviours with respect to which self-control may fail, from fidgeting to blurting out answers to difficulty staying on task.

Secondly, these failures are similar in kind to self-control failures in persons outside the clinical population. Associated with impairments in executive functioning, ADHD is characterized by behaviours such as acting impulsively, distractibility, motor fidgeting, and blurting out answers (DSM-V). Even as ADHD is sometimes viewed as a gift or advantage, persons with ADHD often struggle with behaviours that are explicable by failures of self-control, including temper outbursts, impulsive purchases, and substance use problems (DSM-V; Barkley 1997, 2015). None of these behaviours are foreign to persons outside the clinical population. Rather, what makes self-control failures a clinical problem for persons with ADHD is the *frequency* at which these failures in behaviour thwart their plans. This makes the case elucidative of non-clinical self-control problems, as well. Studying non-clinical cases with the help of the case of addiction is also helpful, but the case of ADHD, by virtue of the broad range of behaviours with respect to which self-control may fail, helps give a different, broader perspective.

Third, and perhaps most compellingly, at first sight ADHD appears a fitting counterexample for externalism about self-control (and therefore for most of my claims, above). The thought goes, what makes for better evidence that self-control failures are agent-internal than a heritable brain disorder resulting in self-control problems? There is a persistent tendency in laypeople,

professionals, and academics alike to associate the genetic or neurodevelopmental causation of a trait with its being agent-internal and (relatively) immutable (see Dar-Nimrod & Heine 2011). However, as I explain in article 2, this is at odds with the complex reality of genetic causation. Understanding the reality of biological causation underlying self-control, in my analysis, increases rather than decreases credence in defining self-control by its function.

1.4. Remarks concerning approach and methodology

A challenge for this dissertation is that its topic seems to sit at a junction of various subdisciplines of philosophy. It has, as a result, been very hard for me to classify this work squarely within a subdiscipline of philosophy. It combines elements of applied ethics (especially, neuroethics), metaphysics, philosophy of science (especially, of philosophy of psychiatry and philosophy of psychology), philosophy of action, and philosophy of mind. While this makes the dissertation something of a chimera, for the study of self-control, a synthetic approach is best suited for providing results that are both compatible with the current empirical evidence and conceptually sound.

I do have an overarching approach to self-control. That overarching approach is an empiricist and pragmatist one. I do not believe that the concept of self-control can be explicated, or individual differences in self-control explained, a priori, in the armchair. In my view, in order to make headway with this topic, thoughtful engagement with the empirical sciences is necessary. Unlike some other subjects of philosophical investigation, self-control is not what it is because of some universal necessity. Rather, it is a feature of the human condition and as such, contingent on both the biophysiology of Homo Sapiens and on their social and cultural practices. As a result, all four articles, as well as this introduction, are in discussion with research in the behavioural and cognitive sciences, ranging from ‘soft’ ones such as psychology, behavioural economics, psychiatry, and criminology to the ‘hard’ biosciences of behavioural genetics and neuroscience. This dissertation reflects the two-way relationship that philosophy and the empirical sciences can have. Ideally, not only can empirical sciences inform philosophers by helping us question some intuitions that may turn out to result from cognitive bias or from misunderstanding, philosophers can critically engage with the way empirical scientists operationalize their concepts, draw inferences from their data, and

interpret their findings. It is my hope that the work herein reflects an earnest attempt towards this ideal.

One more methodological remark is in order, namely, to explicate my project in asking what self-control is. It is clear that the word 'self-control' is used in a plethora of ways, not all of which are consistent with each other. For this project, my project is not the descriptive linguistic project of outlining all senses of self-control in current lay and professional usage. Rather, this is something where I fall within the philosophical mainstream: I assume there can be an adequate unitary concept of self-control, and that by having such an adequate concept of self-control, the work we do with the concept of self-control (which will be discussed across sections 2 and 3, and in particular, in section 4 where my revision of the concept is introduced) can be done better. I connect my conception to existing usage of the term, not because I would be seeking merely to explicate existing usage, but because having such a connection between a conceptual revision and existing usage is a desideratum. The account of self-control I offer is intended to serve the *aims* of the common usage: to be helpful for the *work we do with the concept* of self-control.

Overlap with standard usage of the term serves as evidence of the applicability of my account of self-control for the term being defined. But I have no qualms about departing from common usage when that is required for principled reasons, such as for reasons of conceptual coherence. The lay concept is subject to a range of ambiguities, resulting in that there is no one standard denotation of self-control. Here, my approach is akin to conceptual engineering, wherein departure from common usage, when it does occur, is not a problem so long as it is reasoned and well-motivated.

I do not assume the unitary approach to be best for all concepts – some concepts may be best understood in variantist terms. I do discuss some disjunctive accounts of self-control in section 2; this discussion, arguably, illustrates that a unitary concept of self-control is more apt.

1.5. The structure of the dissertation

This dissertation is cumulative and features four articles that approach various problems related to the task at hand. The introduction complements these articles, justifying their relevance for the thesis and describing the theory of self-control being developed here.

The first article, "Self-control in responsibility enhancement and criminal rehabilitation", with co-authors Susanne Uusitalo and Jarno Tuominen, explores the complex role self-control has in the context of crime. It is well known that self-control problems and crime are correlated; should, then, future neuroscience seek to improve self-control? In the article, we examine whether such a project

is ethically permissible. We demonstrate that the presence of environmental constraints can thwart attempts to medically improve self-control, especially in underprivileged populations, suggesting that duly accounting for the environmental factors relevant for self-control are key for the ethical permissibility of attempts to enhance self-control. This article is further discussed in section 5.3.

In article 2, "Born which way? ADHD, situational self-control, and responsibility", I discuss the role of self-control in moral responsibility. Pleas for the mitigation of responsibility, for people with ADHD, invoke the control condition for responsibility. They ordinarily refer to the agent's being genetically or neurologically abnormal in a way that persistently thwarts their self-control. However, when the role of the environment in self-control is properly appreciated, it turns out that pleas should instead be situational: the agent was, in those circumstances, prevented from acting in the desirable way. This article will be further discussed in section 5.2.

Article 3, "Accessing self-control", develops my positive account of individual differences in self-control. I defend a functionalist conception of self-control and argue that differences in self-control are analogous to differences in mobility in that they are modulated by inherent traits and environmental supports and constraints in interaction. In brief, I argue that individual differences in self-control amount to differences in access to self-control behaviours; I explicate access as including the three criteria of means, awareness and non-excessive effort, and demonstrate how these criteria are variably satisfied within a population. A summary of this article can be found in sections 5.1 and 5.2.

Finally, article 4, "Towards informed user decisions about pharmacological cognitive enhancement", returns to the theme of stimulant medication initially explored in article 1. Users of stimulant medication in subclinical populations engage in the use of these pharmaceuticals to improve their cognitive capacities and their self-control. Defenders of a liberal approach to enhancement claim that users are well placed to assess the costs and benefits of these pharmaceuticals, and that therefore we should let those who benefit from stimulants to use them at will. In the article, I demonstrate that the reality of stimulant medication is much murkier, and that informed user decisions are difficult to procure in practice; I also explore possible means to provide assistance for this difficult choice. This article is discussed in section 5.3.

As is typical in cumulative dissertations, many gaps remain to be filled for a cohesive thesis to be possible. This introduction endeavours to fill those gaps and to place the articles in the broader context of an empirically engaged philosophy of self-control.

Sections 2 and 3 constitute a critical review of the literature. In section 2, I explore various forms of internalism about self-control, presenting grounds for rejecting them. Section 3 describes and endorses the externalist position about self-control. In it, I also point out some room for

improvement within the externalist position, namely, regarding how it retains a commitment to self-control as value hierarchical.

In section 4, I offer a radical revision of our concept of self-control that I term the *value-deflationist definition of self-control*. I also suggest that self-control is best understood as a practical kind.

My account of individual differences in self-control, which I call an *access theory of self-control*, is described in section 5 which summarizes the articles of this discussion. I claim that self-control is analogous in meaningful ways to capacities such as mobility and argue that individual differences in self-control hinge on differences in access to self-control behaviours. I analyse the ways in which physical, social, and epistemic barriers act as constraints for self-control generally, and significantly constrain the self-control of some people, such as people with ADHD. I also describe some of the implications of my approach to self-control for responsibility assessments, as well as for criminal rehabilitation. I conclude the section with a discussion of the prospects of stimulant medication for enhancing self-control within both voluntary users and as a means of criminal rehabilitation, urging caution on both fronts.

2. A long farewell to internalism about self-control

My aim in this section is to provide a brief critical review of internalist thinking about self-control within philosophy, psychology, and related disciplines. That means that a broad swath of theory is to be covered, yet the task is made easier by the fact that there are striking similarities among internalist theories of self-control. I divide internalist theories about self-control in two broad categories that I call *divided mind accounts* and *willpower accounts*.

Divided mind accounts of self-control describe and define self-control in terms of motivational conflicts: for them, self-control is when a motivational conflict is correctly resolved, such as, when an impulse is inhibited in order to act intentionally, or when a smaller sooner reward is forfeited in order to pursue a larger, later goal. I use the term 'divided mind' in a broad sense, here: in the sense that there are competing or contrasting elements, processes, or contents of the mind, such as an impulse and a goal, and that the relationship of these elements is what sets self-control apart from other action selection processes. For these accounts, that relationship is one of conflict that, in self-control, is resolved in favour of one part of the mind (say, the goal-oriented part) rather than the other (say, the impulse-oriented part). Some of these accounts are divided mind accounts in a narrower sense, conceiving of the person as divided into more than one process, agent, or 'mind' capable of assessing a situation, forming intentions, and directing behaviour. But for most of the accounts I group under the term, divisions into, e.g., impulses and intentions, or proximate rewards and goals, suffice. These accounts can describe self-control in terms of these divisions without relying on postulating a discrete faculty of the mind for self-control.

Willpower accounts posit that self-control involves a discrete faculty of the mind, a dedicated capacity to direct one's behaviour that may also be termed 'willpower'. Willpower accounts also involve a divided mind in the broad sense: there is a motivational conflict that willpower is employed to resolve. They are however set apart by that they explain self-control by postulating a discrete capacity or faculty for self-control, whereas for divided mind accounts, resolving motivational conflicts does not necessitate a discrete faculty but can instead be accomplished by the same processes that govern action selection more broadly.

2.1. Norms and prudence: divided mind accounts

In Plato's Republic, Socrates bemoans the absurdity of self-control:

> But 'master of oneself' is an absurd phrase. For if you're master of yourself you're presumably also subject to yourself, and so both master and subject. For there is only one person in question throughout. (Plato 1987: 430e)

A minimal understanding of self-control would be to understand it in the negative: as the absence of rule by others. That, however, is typically not what we have in mind when we talk about self-control and is discussed under terms such as personal autonomy instead. Self-control and autonomy do not always go hand in hand. One can be self-controlled but not autonomous, such as the soldier who exerts self-control in order to keep going on a long march, but acts not out of intrinsic motivation but rather under threat of severe punishment if he fails to comply. And one can be autonomous and not self-controlled, such as the athlete who, disheartened by a plateau in her progress, caves to her desire to skip practice and go out for a burger instead. Socrates ends up endorsing a divided-mind view of self-control, in which "there is a better and a worse element in the personality of each individual" (ibid., 431a); when the better controls the worse, we refer to that as self-control.

In amended forms, this idea persists. Self-control is characterized as the rule of rational agency over acting on appetites (Aristotle 1146a15, Kalis 2014); as acting on goals versus desires (Kotabe & Hofmann 2015); in terms of system 2 processes over system 1 processes (Kahneman 2011, Levy 2011); in terms of pursuing larger, later goals rather than overemphasizing smaller, sooner rewards (Ainslie 2001, Levy 2017); as acting on long-term goals in the face of temptation or impulse (Magen & Gross 2010), and as acting according to all things considered better judgment as opposed to just any motivation (Davidson 1970, Mele 1987, 1995). For these accounts, there are certain judgments, intentions, or parts of the mind that are preferable to certain others. In self-controlled behaviour, our behaviour aligns with these preferable motivational factors rather than with competing impulses or motives. Divided mind accounts associate self-control with resolving a motivational conflict in favour of the *better, loftier, or more valuable* aim or desire rather than with simply resolving a motivational conflict any which way and acting accordingly.

The hierarchy of aims or desires involved in divided mind accounts is sometimes described in overtly normative terms; the alternative is to endeavour to minimize normativity and evaluate which of the competing motives ought to be suppressed and which enacted in terms of practical rationality. The roots of this tendency are found in the accounts of Plato and Aristotle. In Plato, the

better part of the soul (i.e., reason) controls the worse part. For Aristotle, the agent with *enkrateia*[2] acts correctly in the face of bad or excessive appetites (1146a15).

Perhaps surprisingly, overt normativity does not happen within the domain of moral philosophy alone: social psychologists, among others empirically working on the topic of self-control, have described it in decidedly normative terms. For example, social psychologists Baumeister & Exline (1999) describe self-control as "the master virtue":

> More precisely, virtues seem based on the positive exercise of self-control, whereas sin and vice often revolve around failures of self-control. Insofar as it is fair to regard vice and sin as the opposite of virtue, the centrality of self-control to both is an impressively consistent theme. We submit that it is fair to consider self-control the master virtue. (Baumeister & Exline 1999: 1175).

Approaches like Baumeister & Exline's strike some as surprising because empirical science, according to a received conception of scientific inquiry, is supposed to describe rather than evaluate. On one hand, they introduce self-control as a normative capacity (a 'master virtue' on which virtues are based), whereas on the other, they suggest it is a fit topic for quantitative psychological study – approaches that, to risk comical understatement, are at some friction. A boon of this controversial approach, however, has been the renewed philosophical interest in self-control that Baumeister's work, specifically, has generated. For example, his theories of self-control as a moral muscle (1999) and its further development, the strength model (aka ego depletion theory) (2007) have compelled a philosophical response, with Holton (2003) and Henden (2008) debating how best to account for the role of willpower, as described by Baumeister, in philosophizing about self-control. These will be described in more detail in section 2.2.

Other, contemporary philosophical contributions to theorizing about self-control have responded to a broader range of empirical literature on self-control. Here, I will describe internalist responses, returning to externalists in section 3. Horstkötter (2015) outlines in detail how normative claims about self-control, while typically not as overt as Baumeister & Exline, are ubiquitous in social psychology: in her analysis, social psychologists constantly refer to self-control's desirability and its wide-ranging positive consequences such as academic performance and law-abidingness;

[2] Enkrateia is perhaps best translated as continence, or, if a more literal translation is desired, as self-mastery; its opposite, akrasia, as incontinence. However, most of the literature on akrasia refers to it as weakness of will. The denotations of enkrateia and akrasia do not match entirely to the modern concepts of self-control or lack thereof; however, for my present purposes of discussing these concepts as the origins of the contemporary self-control debate, rather than engaging in exegesis, describing them in terms of self-control will do.

and define it in normative terms, such as, as the inhibition of *inappropriate* impulses. For example, examples of uses of self-control within the psychological literature include to "resist impulses to go back to sleep, to eat fattening or forbidden foods, to say or do hurtful things to their relationship partners, to play instead of work, to engage in inappropriate sexual or violent acts, and to do countless other sorts of problematic behaviors" (Baumeister, Vohs & Tice 2007: 351). The use of various addictive substances is likewise routinely viewed as a failure of self-control (for discussion, see Levy 2013). Within criminology, poor self-control and criminality are often equated, following Gottfredson & Hirschi's (1990) influential General Theory of Crime, which postulates that poor self-control and criminality are two sides of the same coin.

This pool of examples suggests a normative conception of self-control as enabling not any life plan but certain sorts of life plans – namely, life plans where the better, loftier or otherwise more valuable aims are pursued[3]. For such a normative conception, self-control is valued in part because it enables the sorts of behaviours that positive valence is often, in our culture, attached to: waking up early, eating moderately, acting in a friendly and compassionate manner, working instead of playing, refraining from inappropriate acts, and engaging in countless sorts of other behaviours that are not considered problematic.

Rather than arguing that the concept of self-control ought to be purged of its normativity, Horstkötter concludes that self-control is a normative capacity (ibid: 40). Kalis (2017) is of a similar mind. Defending an Aristotelian conception of self-control, Kalis argues that while conceptions of self-control as a psychological mechanism and those of it as a normative capacity are at odds, mechanistic explanations ought to be relinquished in favour of describing self-control in normative terms.

Above, I have focused on accounts of self-control that are overtly value-laden in that they attach self-control to moral value in a manner that that goes beyond the norms of prudentiality. Many more contemporary accounts of self-control, however, attach it to prudential value: to forfeiting smaller, sooner rewards in pursuit of larger, later ones. For the latter, strictly prudential accounts, someone – say, a leader in organized crime – is self-controlled insofar as her behaviour is in the correct relationship to expected rewards, even if these rewards are morally reprehensible,

[3] Does this mean that self-control is not a Rawlsian primary good – something that is useful for any life plan and should therefore be preferred by any rational agent? This depends on how one reads Rawls. I read Rawls (1999) as having the starting point of considering consider what free, equal, reasonable and rational humans find desirable, which reveals that the concept of primary good is about goods for free, equal, reasonable and rational humans. Self-control as involving the pursuit of the better, loftier, or more valuable aim is useful for any life plan that a free, equal, reasonable and rational human could endorse, so self-control, for this normative account, does seem a primary good. The value-deflationist account of self-contro thatl I will endorse is also a primary good due to its utility for the life plans of free, equal, reasonable and rational humans, but it can also be used for the unreasobnable and arational.

make this person a poor citizen, et cetera. Ainslie's (2001) influential hyperbolic discounting theory is a case in point. For Ainslie, humans have a natural inclination to discount distal rewards in comparison to proximate ones. While some such discounting is rational (since it allows for unknown variables, perceptions of waiting time, and probabilistic reasoning about one's environment; see McGuire & Kable 2013), temporal discounting is often conducted too steeply, and as a result, the distal reward may appear less appealing in comparison to the proximate reward even if, according to a correct value calculus, it would be the more valuable one. For Ainslie, failures of self-control occur because we fail to appreciate the full value of the distal aim and end up choosing against what practical reason dictates. The very proximate reward – say, a slice of chocolate cake right in front of a person attempting to diet – yields the curious result that the agent judges they should refrain from eating cake, except for at present. Yet even with steep discounting, human motivation follows a neat calculus – it is just that the calculus it follows is not the one that would maximize benefit.

Accounts of self-control that either follow Ainslie's or have an affinity with it in that they define self-control in terms of prudential value conflicts between larger later and smaller sooner rewards, have been defended by Rachlin (2000), Fujita (2011), and Metcalfe & Mischel (1999), among others. Some of these further explain why delay of gratification fails by invoking a division of the mind into hot and cool mechanisms (e.g., Metcalfe & Mischel 1999; Hofmann, Friese & Strack 2009). These latter, dual-process models of self-control are strikingly similar to Kahneman's (2011) account of human information processing. Kahneman's dual-process model is founded on the idea that there are two distinct systems that process information. Out of these, the "hot" system 1 is fairly automatic and fast, whereas the "cool" system 2[4] is consciously controlled, processes things using language and logic, and is much slower but also more comprehensive in scope than system 2. When working on a task, both systems are activated, with system 2 regulating system 1 in case of a motivational conflict. However, the inhibitory, overriding processes system 2 has to offer are not perfect, and sometimes system 1 gets its way. System 1 is associated with urges and emotions, whereas system 2 is associated with rationality. As a result, part of the appeal of applying the dual-systems theory to self-control may be because it can be interpreted as a scientific reformulation that gives us the results of Aristotelian and later accounts of self-control as, essentially, the rule of reason over passion without beginning with their overt normative commitments.

[4] "Hot" and "cool" here are Kahneman's descriptions, not the present author's colloquialisms.

While the dual-process model has attracted much attention, on closer inspection, it does not seem to fully account for self-control – at least, not for the self-control that we refer to in common usage. One critic of dual-process models is Sripada (2010), who criticizes these accounts by noting that what some want to express with self-control is typically the control of the self by the self – not the control of one inner process by another. Another concern Sripada raises is that many accounts of self-control associate it with acting according to the agent's strongest desire, but it is unclear which of the two motivations – that of system 1 or that of system 2 – is the stronger one.

Note that strictly interpreted, a dual-systems account of self-control would be value-neutral: for it, self-control is when the motivational outcomes of the two systems diverge, resulting in a motivational conflict, and system 2 successfully regulates system 1, causing the agent to forfeit the system 1 motive and to behave according to the system 2 motivation. However, system 2 motivations sometimes are morally and / or prudentially inferior to those of system 1. A proponent of a dual-systems account needs to either bite that bullet and allow for self-control against morality or prudence, or add a further constraint to that definition of self-control, adding, e.g., that self-control is the regulation of system 1 processes by system 2 in pursuit of the all things considered better judgment. If the latter, then Ockham's razor seems to favour the simpler divided mind approach, as in that case the explanatory force of the dual systems for self-control is contingent on whether system 2 gives the correct recommendation.

Given these restrictions, the best divided mind approach seems to be the simplest one: that of self-control as choosing larger later rewards over smaller sooner ones. For such an account, a motivational conflict lies at the heart of self-control, and only when that conflict is resolved in favour of the larger later reward has the individual enacted self-control. However, this best approach also runs into serious limitations.

2.1.1. Problems for divided minds

The pursuit of long-term goals, health, good interpersonal relationships, and moral values are domains of self-control that rightly merit our attention. Restricting self-control to the pursuit of the more valuable, loftier or better aim is also well in line with the perennial philosophical discussion around the possibility of akrasia, that is, acting against one's better judgment (Mele 1990, 2001; Kalis et al. 2008, Kalis 2017). Indeed, much of philosophical debate on self-control simply defines it, like Mele (1990) does, as the opposite of akrasia – that is, as enkrateia. However, this introduces problems and paradoxes from the akrasia literature into the study of self-control. The chief problem,

well discussed in the work of Mele (1987, 1995), is that there seem to be unorthodox cases of self-control as sometimes, it seems, we use self-control to accomplish an aim we do not endorse at all, an aim we see as less morally and prudentially valuable than the competing impulse. Mele's (1995: 60–61) example is that of a youth who, due to peer pressure, commits petty crime despite having strong judgments against doing so. To commit those offenses, the youth needs to muster self-control in order to stay on task and not "chicken out". But as Mele (1997) points out, there may be more than one motivational conflict at stake, and these may be resolved in mismatching ways. The youth's better judgment may favour the impulse of fleeing to carrying out the crime, but the youth may also prefer the exercise of self-control to succumbing to the impulse to flee.

In practice, Mele's solution for resolving the case of "errant self-control" could be described as slicing the action into smaller units. The use of self-control in a task that is against one's better judgment is thus analyzed as a failure of self-control in deciding to do that task, and successful self-control in executing it, defending it from lower passions such as fear (Mele 2001: 62). This option retains a value hierarchy even on part of the execution of the act, but does not require it to be optimal: breaking into a house, in this example, is worse than not doing so, but it is still superior to succumbing to fear.

An alternative solution for unorthodox cases of self-control is to strip self-control of the requirement of pursuing the more valuable aim. A strict interpretation of the dual-process model would be one such solution: in case of a conflict between the dual systems, *whatever* it is that system 2 endorses, if it successfully inhibits what is endorsed by system 1, the agent is enacting self-control. We could then postulate that for Mele's youth, his system 2 endorses the crime while his system 1 endorses 'chickening out'. While this solution would bring with it the concerns raised by Sripada, it eliminates any value-ladenness, prudential or moral, associated with neo-Aristotelian notions of self-control.

However, it turns out that there are occasions that, like Mele's youth, look a great deal like self-control but that seem not to be captured by a conflict of the dual systems, each of which with its own preferred outcome, nor by an enkratic notion of self-control. In what follows, I will discuss these issues at length.

In addition to resisting addictive behaviours, caring for our health, and conforming to our moral principles, we sometimes use self-control for purposes that do not involve dramatic differences in outcome. I will here make use of a distinction between enkratic instances of self-control, that is, self-control in the service of better judgment, and what I term *non-hierarchical* instances of self-control. Non-hierarchical self-control is when there is a clear motivational conflict without a clear motivational hierarchy: more than one option seems equally good to the agent.

These instances of self-control most often occur in unexciting or almost trivial situations: e.g., an agent may have no preference regarding whether to do the laundry or to do the dishes. But as he sets out to do one chore, his attention may wander and he may end up switching tasks. Even as such an event is innocuous in terms of (some) housework getting done, it is a threat to the person's sense of agency.

I propose that these unexciting instances of self-control are common, and for many persons, resolved without much deliberation. in the philosophical quest to understand akratic failures of self-control, however, failures where the competing options are, roughly speaking, equal in value, have received scant attention.

Non-hierarchical cases may also include higher-stakes instances where the agent sees the options as so close to equal in value that she has difficulty deciding which is the preferable one; ambivalence in such decision-making is discussed by Levi (1986). The received notion of self-control seemingly requires that there is no such ambivalence but rather a clear ranking determining which of the competing courses of action is the 'goal', which the 'temptation'.

Let me clarify the notion of non-hierarchical instances of self-control by contrasting it to other, similar cases. There are mundane, unexciting instances of enkratic self-control, too, such as mustering oneself to take out the trash, which is better than living in a smelly home: it is not the quotidian character that distinguishes non-hierarchical instances from enkratic instances of self-control. Rather, non-hierarchical self-control is set apart by the absence of a clear subjective evaluative ranking of the options. This can be either because the competing courses of action are *in fact* equal in value (or close to equal, so that determining the difference in value would come at a greater cost than gain), or because *the agent* does not *construe of* these options as having a clear value hierarchy.

Out of the latter sort, we can make a rough distinction between two reasons why competing courses of action may seem to the agent to not have a value hierarchy. First, the agent may be missing some information about one or more of the options, or have false information about one or more of them: the information the agent possesses makes the options seem equal even though they are not. For example, Andy is deliberating whether to do the laundry or the dishes; he only has time for one of these chores before going to an event for the rest of the night. These options seem equal to Andy. But yesterday at dinner Andy did not notice that he spilled wine on his favourite shirt. He can only successfully remove the stain if he does it today. Therefore, it would in fact be more valuable for Andy to do the laundry, because if Andy does the dishes, his shirt is ruined. Not knowing this, Andy musters self-control to stop deliberating and to do one of these tasks under the

impression that the options are equal. I maintain here that this instance should count as an instance of self-control, regardless of which chore Andy does.

Second[5], the options may appear to the agent (or in fact be) incommensurable, or they may otherwise result in ambivalent responses. For example, Joanna is upset because her husband has done something bad – say, invited his openly misogynist cousin to stay despite their previous agreement that the cousin is not welcome in their home. Joanna has an impulse to yell hurtful words at her husband, although on the other hand, she knows that doing so would hurt his feelings. She is feeling ambivalent: while Joanna believes it is bad to hurt her husband's feelings and is therefore motivated to refrain from yelling, she also believes he deserves to be yelled at, and is motivated to yell. She experiences these two motives as incommensurable, and is not able to rank their value, subjective or otherwise. When she controls herself and refrains from yelling, she remains unsure if she should have. I think it reasonable that this instance, too, should count as an instance of self-control.

Now, if I am correct that there are cases where agents exert self-control when the competing courses of behaviour have equal value, and cases where agents exert self-control to enact one course of behaviour despite being in a state of decisional ambivalence, then these cases present a challenge for value hierarchy-centric divided mind accounts. They present such a challenge regardless of whether they have a neo-Aristotelian moral component to them, such as Kalis' approach, or whether they are founded on prudential value alone, such as Ainslie's account. My main thrust against divided mind accounts therefore is that for such an account, only if the agent resolves a motivational conflict by forfeiting the less valuable aim does this count as an instance of self-control. This seems incompatible with calling the cases of Andy and Joanna cases of self-control.

A philosopher who is committed to the idea that differences in the value of options can always be found may argue that there are in fact value hierarchies in all such cases that I call non-hierarchical: the agents (and I) are simply mistaken in believing otherwise. For example, Joanna's case may cause such a philosopher to wonder if Joanna does, after all, prefer not to yell. For this objection to the case, her experience of ambivalence may be misguided, and her reasons not to yell may be stronger than she is able to articulate. Thus, Joanna has false beliefs about the relative value of these competing courses of behaviour. Here, my response is twofold. If the critic's point is that we do not know whether Joanna actually prefers not yelling – that we can envision both Joanna_1, who has such a preference, and Joanna_2, who does not have one, then the criticism is harmless as my account only

[5] This list is intended as illustrative rather than be exhaustive – there are likely to be further relevant sorts of cases.

requires that there are indeed instances where self-control is utilized to enact behaviours one does not prefer; the existence of some Joanna$_2$ instances suffices to make my case, and I willingly grant the existence of some Joanna$_1$ cases. It seems uncontroversial to me that agents sometimes act on reasons they are unaware of, and likewise that they may act under self-deception.

Now, if the critic's claim is the stronger claim that from Joanna's decision, we can infer that Joanna actually preferred not to yell, then that is a claim that I must deny. Such a philosopher would likely have a similar objection to Andy, arguing that even if the agent sees the options as equal, they ultimately choose according to what they subconsciously find most valuable, being simply unaware of doing so; in economics, this claim is called the revealed preference theory. And revealed preference theory is one that I must deny in order to claim that non-hierarchical cases exist, and are important. But it is not one that I would be alone in denying. Indeed, while revealed preference theory continues to have its proponents, it has also faced a sustained onslaught of criticism for the last 50 years or so[6]. I won't here make an attempt to argue against revealed preference theory, as that undertaking falls outside the scope of my thesis. In section 4, I will also add that my positive account of how self-control ought to be conceived is compatible with revealed preferences.

Note also that proposing an unconscious value hierarchy in such cases would be unhelpful for the work we do with the concept of self-control: roughly, we use the concept of self-control to assess whether we need to deploy our capacity for self-control in a certain situation or not – how to do it, and what can be achieved by it; to assess whether an attempt at self-control has been successful; and to seek and provide help if one keeps failing at it. Most of the time, we do this implicitly; nevertheless our conceptions of agency and control impact these assessments, whether implicitly or via explicit reasoning. For the concept of self-control to be useful for tasks like these, it ought to be feasible, at least in most cases, for agents to self-assess whether they have successfully controlled themselves. Again, it seems more helpful to look at whether agents have successfully guided their behaviour to conform to their intentions. Whatever it is that causes Joanna to intend to refrain from shouting, if she successfully does so despite a conflicting impulse, she has controlled herself.

Here, a philosopher might object that Joanna need not be interested in whether she has controlled herself: if her self-assessment as an agent is at stake, then Joanna should be more interested in whether she is autonomous, is enacting certain virtues, and so forth – not in whether

[6] Some pertinent criticisms of revealed preference theory can be found in Levi (1986) and Sen (1973). For example, Ainslie's (2001) account of self-control relies methodologically on revealed preferences. A recent philosophical defense of revealed preference theory can be found in Thoma (2020).

she exerted self-control. But this criticism strikes me as peculiar, as these are not at odds with each other: agents may be interested in both.

Non-hierarchical cases seem compatible with a dual-process account of self-control. For a dual-process account, whichever of Andy's chores system 2 endorses, resolving the motivational conflict in favour of that chore counts as self-control. But here we run into another kind of problem. It seems inaccurate to describe Andy's system 2 as endorsing one chore while system 1 is endorsing the other. Rather, it appears that Andy has system 2 reasons to do either of the two options. Indeed, people who have never found themselves having a tough time deciding between laundry and dishes may say that Andy is *overthinking* it.

Joanna's case seems more compatible with the dual systems view: it seems that system 1 is motivated to yell, whereas system 2 is inclined to inhibit that. But it also seems that Joanna has both system 1 and system 2 processes at play for both courses of action. She has an urge or an affective impetus both to yell and to protect her husband from hurt feelings. Her system 2 deliberation does not recommend either course of action clearly.

A similar defence is available for the dual-systems theorist of self-control than is for the value hierarchy theorist: first, we may simply choose to describe these cases as something else than self-control; and secondly, the dual systems division may covertly apply to these cases even if it overtly seems, to the agents and to me, that it does not. But here, as well, the helpfulness of the concept of self-control is at stake. Self-control is a concept used for describing success and failure in bringing oneself to act in one way over another. Were self-control described in a way that does not serve that purpose – e.g., in such a way that it would be opaque to agents whether or not they have successfully controlled themselves – the concept would fail to serve this purpose. In making this point, I am not claiming that agents would always have a complete understanding of their behaviours and motivational states. I am simply claiming that the concept of self-control, in order to be useful for agentive self-understanding, needs to refer to phenomena that agents can be expected to be privy to with reasonable frequency.

Further objections towards divided mind accounts have been raised by Fujita, Carnevale & Trope (2016). They point out that while the dual-systems theory seems to suggest that in order to succeed in self-control, we ought to pause and think more, empirical evidence suggests that deliberative processes are often employed to rationalize acting on impulse. They cite research by Weber and colleagues (2007), which suggests that "when asked to think carefully and deliberately, people are better able to generate reasons to indulge than not to indulge" (Fujita, Carnevale & Trope 2016: 284). Likewise, the prediction that self-control would fail in the absence of deliberation seems false: by employing implementation intentions, agents can create automatized self-control

mechanisms for specific situations (ibid.). In those situations, deliberation has occurred in a prior, planning situation, but need not occur in temporal proximity to the action in question. For example, Joanna may decide, "if I ever get angry at my husband, then I will not yell", over time creating a mechanistic link between the temptation and the selected behaviour[7]. Fujita, Carnevale & Trope suggest that instead of postulating that self-control is when a certain part of the mind controls another part, we should envision self-control as involving the wholehearted pursuit of a cohesive life plan.

Unsatisfied with the divided mind approach, Fujita, Carnevale & Trope suggest an alternative. For them, self-control is when motivational elements that diverge from such a cohesive plan are managed by means of top-down mechanisms. Their chosen metaphor is that of a "senate of the mind", where competing motives are akin to senators, each with their own agenda; in this metaphor, a combination of negotiation and majority rule determines the senate's joint agenda. They illustrate this with the following example: "[an] individual senator attempting to exert influence against the objections of the broader collective forms the heart of the self-control dilemma. Although eating the cake would satisfy the hedonic eating senator, it undermines the collective will of the senate" (ibid., 286). But whereas Fujita, Carnevale & Trope are correct in acknowledging that situations where self-control is enacted are often more complex in terms of their motivational structure than simply a tug-of-war between an automatic impulse and a reasoned response, their argument risks resting on a runaway metaphor. Rather than dividing the mind in two, they are now dividing it into a whole senate of parts of mind, treating each part as an agentive force of sorts and treating individual persons as group agents. It is unclear what sorts of mental entities or processes count as "senators"[8].

Perhaps the most sweeping objection against divided mind accounts is in the observation that agents rely on a range of practices, including environmental practices, to regulate their behaviour, and that these practices are not just supporting systems for 'self-control proper' but rather, they just are self-control. This objection will be further discussed in section 3, where I describe externalism about self-control. First, however, we should discuss another, overlapping class of internalist accounts about self-control: accounts that describe self-control as crucially involving a discrete mental faculty. I term these willpower accounts of self-control.

[7] Of course, such mechanisms are sometimes less than reliable, and this example may be a case in point. But the creation of habits, more generally, is a self-control strategy that works much like the implementation intentions that Fujita, Carnevale and Trope (2016) discuss in terms of relying on system 1 rather than system 2 once established.
[8] Conceivably, construing of motivational conflicts in terms of the senate metaphor may itself be a form of self-control: Bermudez (2018) has argued to the effect that self-control often works via framing effects, and that thinking of a behaviour as self-controlled or as exhibiting willpower is just such a framing effect. See section 3.1.

2.2. Willpower accounts

A class of accounts of self-control within both philosophy and psychology describe self-control as crucially involving a discrete, agent-internal, mental faculty or resource (Henden 2008, Holton 2003, Muraven & Baumeister 2000, Baumeister, Vohs & Tice 2007, Baumeister & Tierney 2011). That mental faculty may be inhibitory control or, more poetically, willpower. A subclass of such accounts states that self-control *is reducible to* such a discrete mental faculty. Some accounts of delayed gratification, such as Ainslie's, speak of willpower without postulating a discrete mental faculty; for those accounts, their usage of 'willpower' is rather more metaphorical. Ainslie's theory (2001) describes neither will nor power, but rather, it is a theory about commitment to evaluative judgment.

In philosophy, two willpower accounts have become particularly well known. Richard Holton (2003) argues that there is such a mental faculty as willpower. For Holton, however, this faculty is distinct from what he terms self-control: in Holton's usage, willpower is a faculty used for sticking to resolutions, and self-control is equivalent to enkrateia[9]. While Holton's view of willpower is not equivalent to self-control as the term is used in the present dissertation, there is enough overlap for it to be of interest here.

Edmund Henden (2008), by contrast, argues that willpower is a *necessary part* of a form of self-control. That is, for Henden, there is such a discrete mental faculty, and it is of utmost importance for much of self-control; but Henden allows that it would not be active in all instances or forms of self-control.

Both Holton and Henden draw their inspiration from what could be termed willpower theories of self-control within psychology: theories that describe self-control in terms of ego depletion (e.g., Muraven & Baumeister 2000; Baumeister, Vohs & Tice 2007) or in terms of inhibitory control (e.g., Cohen, Berkman & Lieberman 2013). Baumeister, especially, has led the way in popularizing a conception of self-control as relying on a dedicated mental faculty he repeatedly refers to as willpower, including in a popular science volume describing his decades-long work in self-control and describing willpower as "the greatest human strength" (Baumeister & Tierney 2011). Thus, in contrast to Holton and Henden, Baumeister's view seems to be that self-control is equivalent to the same discrete mental faculty he terms willpower. In section 2.1, we also saw that Baumeister and colleagues have described willpower as "the master virtue" (Baumeister & Exline 1999). But while

[9] Holton discusses his usage as follows: "I would rather use self-control to describe the related but distinct phenomenon which is the contrary of akrasia: on this usage one lacks self-control if one does other than that which one judges best, even if one does not thereby violate one's resolve (and hence is not weak-willed" (Holton 2003, p. 56).

some see willpower as the greatest human strength and the master virtue, others argue willpower does not even exist, save as metaphor. Before moving to these objections, I will describe how willpower accounts describe the human capacity for self-control.

Perhaps the simplest sort of willpower accounts are inhibitory control accounts of self-control. These accounts define self-control simply as inhibitory control. Inhibitory control refers to the executive function that allows an individual to suppress habitual, instinctive, and other prepotent responses to stimuli, whether those responses are thoughts, behaviours, feelings, or the like. Inhibitory control, for such an account, is not just a faculty or mechanism that is part of self-control: it is the *very same* faculty that we refer to as self-control. Inhibitory control accounts share quite some overlap with ego depletion accounts, but it adds clarity to treat ego depletion accounts as a further elaboration of inhibitory control accounts rather than as equivalent to them. Examples of accounts explicitly defining self-control as inhibitory control include Muraven et al. (2006), who define self-control as "the overriding or inhibiting of automatic, habitual, or innate behaviors, urges, emotions, or desires that would otherwise interfere with goal directed behavior" (p. 524), and Cohen, Berkman & Lieberman (2013) whose account of self-control postulates that various forms of self-control – that is, motor inhibition, inhibition of temptation, and emotional inhibition and inhibition of risk-taking – lists the various domains of inhibitory control as the various domains of self-control (p. 417).

More often than not, however, the operationalization of self-control as inhibitory control is not explicit. Fujita (2011) suggests there is an implicit acceptance of self-control as inhibitory control in the psychological literature: "This tradition of research proposes that self-control requires recognizing these impulses as undesirable and inhibiting them. […] This proposed intrapsychic conflict between impulse activation and effortful impulse inhibition is evident in many theoretical accounts of self-control in psychology" (p. 354). Fujita lists the ego depletion and dual systems theories, among others, as subscribing to this line of thinking.

For inhibitory control accounts, self-control is when an agent gets an impulse to behave in a certain manner and inhibits that impulse by means of a faculty referred to as inhibitory control. The classic example of this is addiction: for this account, a smoker wishing to quit must inhibit the urge to smoke in order to refrain[10]. To illustrate this model with another example, consider Dani. She picks up a novel, intending to read it. But as she opens the novel, she is struck by a wealth of competing impulses, such as to do a load of laundry, to bake bread, to aimlessly browse social

[10] The stance that recovery from an addiction would rely on inhibitory control is not uncontroversial; for the purposes of the present discussion, I believe it more useful to steer away from this complex topic and towards other examples. For nuanced discussions of self-control and addiction, see Uusitalo (2015) and an edited volume by Levy (2013).

media, or to put the novel down and read poetry instead. According to the inhibitory control account, what happens in Dani's failure to keep reading the novel is essentially this. As Dani intends to read the novel, she gets an impulse to do something else. To stay on task requires that she inhibit any impulse to switch tasks. When her ability for inhibitory control – in some theories, such as that of Kotabe & Hofmann (2016), coupled with other factors such as motivational strength or effort – directed into suppressing any oncoming impulse trumps the strength of that impulse, she is successfully using her self-control and ends up reading the novel.

Inhibitory control is a well-documented executive function whose role in some instances of self-control is not under dispute. We should, however, be wary of simplistically reducing self-control to inhibitory control: such a reduction could not explain, for example, why it is that some people sometimes succeed yet sometimes fail in self-control (other than by temporary malfunction of inhibitory control processes, which does not begin to explain why successes and failures in self-control so often are domain-specific, e.g., a person who cannot reliably hold her temper nevertheless can reliably regulate her appetite). This challenge is pertinent because executive functioning ability in individual executive functions, such as inhibitory control, stabilizes by late adolescence (see, e.g., Friedman et al. 2008, Friedman et al. 2016). While environmental effects continue to modulate executive functioning – e.g., it is harder to redirect attention from the television to a textbook if the television is loud – theorists wanted to explain variance by agent-internal factors in addition to merely environmental ones.

A range of modified inhibitory control accounts sought to solve this problem, most prominently the ego depletion model of self-control (e.g., Job et al. 2010, Inzlicht & Schmeichel 2012, Levy 2007, Hagger et al. 2016, Inzlicht & Friese 2019), also sometimes called the strength model of self-control (Baumeister, Vohs & Tice 2007) and the willpower theory (Clarkson et al. 2016). This model postulates that in addition to inhibitory control, successful self-control relies on a dedicated resource of the mind or brain: in what is termed the ego depletion effect, the effortful use of one's "willpower" reserves depletes them, and thus after a long day of inhibiting prepotent impulses, an agent may be less capable of inhibition. It further posits that there is individual variance in the amount or strength of willpower individuals have, but that willpower can be trained to strengthen it, much like a muscle (Baumeister, Vohs & Tice 2007). The model therefore endeavours to explain both intertemporal and interpersonal variance in self-control.

In studies examining how willpower can be improved and the depletion effect slowed, moderate effects have been found for both glucose (Gailliot et al. 2007) and methylphenidate (Sripada, Kessler & Joniker 2014) in postponing ego depletion in healthy subjects, in effect enabling them to exert more willpower. However, the ego depletion effect seems also susceptible to

a range of psychological effects: political beliefs (Clarkson et al. 2016), motivation (Vohs, Baumeister & Schmeichel 2012), affirmation of core values (Schmeichel & Vohs 2009) and strikingly, beliefs about the nature of self-control (ibid., Clarkson et al. 2016) all have been documented as modulating the ego depletion effect, although success in replicating these studies seems variable (Clarkson et al. 2016).

The inhibitory control model yields a readily quantifiable conception of self-control: in laboratory conditions, inhibitory control can be studied with means such as Stop-Signal, Go/NoGo and Stroop tasks. This research-friendliness has helped psychologists and neuroscientists make considerable advances in the study of inhibitory control, including deciphering the neural correlates of this ability and activity (e.g., Tabibnia et al. 2011; Cohen, Berkman & Lieberman 2013). Various parts of the frontal lobe have been demonstrated to have a vital role in inhibitory control, including the inferior frontal gyrus, which is associated with all domains of inhibitory control, ranging from motor to emotional control. Inhibitory control can also be manipulated, including with non-invasive transcranial direct current stimulation (Hsu et al. 2011) and with prescription stimulants such as methylphenidate (Ilieva, Boland & Farah 2015). There is also some research into the prospect of training inhibitory control (see, e.g., Beauchamp, Kahn & Berkman 2016).

In and of itself, defining self-control simply as the inhibition of a prepotent impulse presupposes no value hierarchy among the competing goals. In practice, however, social psychologists defending willpower accounts of self-control pair an inhibitory control approach with dual-systems and temporal discounting approaches: recall Muraven et al. (2006) defining self-control as the use of inhibitory control to *promote goal directed behaviour*. In these accounts, self-control is either implicitly or explicitly defined as the inhibition of a prepotent impulse by a reflective process in the pursuit of a larger later goal over a smaller sooner one. Internalist conceptions of self-control as involving a motivational hierarchy on one hand, and as involving the effortful inhibition of impulses on the other, come together, either implicitly or explicitly. In section 2.3, I will discuss complex models of self-control that outline how these various features of self-control could come together (Buckholtz 2015; Kotabe & Hofmann 2016; Gillebaart 2018).

Neither construct, that of inhibitory control nor that of temporal discounting, was originally intended to serve as a comprehensive definition encompassing all of self-control. Rather, both originally demarcated a process involved in many or most instances of self-control. It is, however, not uncommon to treat either account as a definition of self-control in empirical sciences, and this has had an impact also on both philosophical and lay thinking about self-control and willpower. Holton, for example, sketches an account of willpower that matches, and is directly inspired by, Baumeister's 'muscle model' of self-control qua inhibitory control: for Holton, willpower is best

understood “as a block on reconsideration” (Holton 2003, p. 64) that works “very much like a muscle, something that it takes effort to employ, that tires in the short run, but that can be built up in the long run” (p. 49). Other philosophers, such as Henden, have made a case for generalizing the role of willpower (or, perhaps, inhibitory control) to map across much of self-control. Henden allows for there to be diachronic self-control without willpower, but holds it to be necessary for synchronic self-control[11]:

> To sum up, unlike Holton who appears to believe that will-power and self-control are separate, I believe will-power is an essential component of synchronic self-control. The self-controlled agent uses will-power to direct her attention away from a rebellious desire in order to form, retain or execute an intention to perform the act she takes herself to have most reason to perform.
>
> (Henden 2008: p. 84.)

I will next look at the various objections that willpower accounts of self-control face.

2.2.1. Problems for willpower accounts

The most prominent objection for willpower accounts of self-control is to deny that all self-control involves struggling against active impulses. This line of argumentation has been previously employed by Fujita (2011) and Levy (2007, 2011, 2017) among others. It maintains that self-control is also enacted using a variety of cognitive and environmental skills and practices aimed at *preventing* oncoming impulses or diffusing them before they become so salient as to require active resistance (see, e.g., Duckworth, Gendler & Gross 2016). While some (e.g., Sripada 2020) deem this an unusual usage for the term self-control, the prevention and early diffusion of competing motivations is fairly well in line with how self-control is commonly conceived. For example, we hold that to quit smoking requires self-control; but smokers trying to quit may do so by throwing their cigarettes in the trash rather than by demonstrating willpower by engaging in a staring contest with a pack. We hold that focusing on studying for an exam requires self-control, but students may do so by going to the library and muting their phones rather than engaging in willpower athletics by trying to focus in the midst of a social gathering. We likewise think it can take self-control to go vegan, but a person who has newly pledged to go vegan, who enters a café and notices a small

[11] The synchronic-diachronic divide in self-control is a matter of some discussion that I will address in section 3.2.

hankering for the dairy ice cream they sell, can grab coffee to go in order to stick to that resolution and thus avert the hankering turning into an intense craving[12]. From examples like these, we can infer first that environmental practices are at least sometimes functionally equivalent to willpower: they serve the same behavioural function, namely, the avoidance of one course of action in order to complete another; and second, that as we commend those who regularly engage in such practices for their self-control, this is at least not entirely removed from common usage.

The above examples suggest that self-control is multiply realizable, and that while some instances of self-control rely heavily on inhibitory control, for others, the effortful inhibition of impulses serves merely an auxiliary role, if any.

Unlike willpower accounts, divided mind accounts easily accommodate multiple realizability because they do not pin self-control down to any single process or function of the mind or brain. Multiple realizability, in my discussion, refers not just to the idea that the same cognitive process can be realized by a variety of physical bases, but rather, to the idea that the same behavioural upshot can be realized by many types of processes and mechanisms. This line of thinking will be discussed further in section 3 where externalism about self-control is described and endorsed.

If the above discussion is accurate, then inhibitory control seems not to be necessary for self-control. Furthermore, not all inhibitory control happens in the service of self-control (under ordinary usage of the term). It also occurs in behaviours such as in hesitation: if I reach my hand out to grab a piece of chocolate, then pull it back – just to grab another, identical piece, I have successfully inhibited my impulse to take the first piece of chocolate. But hesitating about whether to have a piece of chocolate is not self-control. Rather, it illustrates that inhibitory control is a process that subserves a variety of cognitive and behavioural processes, including hesitation.

We could attempt to rescue the inhibitory control account from this second shortcoming by qualifying it, and such attempts have been made: for example, a common one is to define self-control as inhibitory control in the service of goal-orientedness (Muraven, Shmueli & Burkley 2006) or of better judgment (Mele 1990, 2001). But in such approaches, first, the presence of inhibitory mechanisms in self-control ends up having less explanatory force for self-control than the motivational structures do; second, this account would still force us to exclude Ulysses arrangements and similar environmental practices, often used for self-control purposes, from our definition of self-control when they do not require inhibitory control; and third, the qualification

[12] Here, a line of criticism available for the willpower theorist is that these behaviours, while well and good, are something other than self-control. They might, for example, be examples of arranging one's life so that self-control proper is not needed. This line of criticism will be expanded on in section 3.1.2.

restricts this account to specific motivational structures, which is a problem if I am correct in asserting that there are non-hierarchical instances of self-control.

While inhibitory control as an executive function is well documented, the ego depletion or "willpower" theory of inhibitory control has faced serious replicability issues. Doubts surrounding the ego depletion theory began to surface early on, and the 2016 multilaboratory preregistered replication of the ego depletion effect (Hagger et al. 2016) was considered by many to be the final nail in its coffin. Hagger and colleagues across 23 laboratories internationally replicated the ego depletion effect using a sequential task experimental paradigm and found a null effect. A popular hypothesis is that what studies measuring ego depletion were in fact registering was simply fatigue (Inzlicht & Friese 2019). Yet there are some who continue to work with ego depletion theory; their work encompasses revisions of the theory. One suggestion has been that rather than individuals' *capability* for self-control being depleted by previous self-control tasks, it is their *willingness* to control themselves that diminishes – fatigued subjects being less likely to choose to control themselves (ibid., Baumeister, Tice & Vohs 2018). Baumeister, Tice & Vohs (ibid.) scoff at the credence placed on the replication studies, arguing that a better designed replication – at the time of that article, being conducted by Vohs et al. – would yet vindicate the effect. Yet the results of that replication study vindicated not ego depletion but the null hypothesis (Vohs et al. 2021). In the Vohs et al. study, the only significant correlate of measures of ego depletion was self-reported fatigue.

These difficulties in replicating ego depletion theory urge us to set it aside until a modified ego depletion effect is substantiated – the prospects of which are slim. These difficulties however do not reflect poorly on the well-attested idea that inhibitory control is an important *part* of *some* self-control. I next review a handful of models joining motivational hierarchies and inhibitory control in unified theories of self-control. These models are persuasive and demonstrate the appeal of a multidimensional, mechanistic approach to self-control. But as I will demonstrate in section 3, these theories, too, are vulnerable to an externalist challenge, i.e. the view that self-control is sometimes enacted in ways that involve the environment.

2.3. Complex models of self-control

Both willpower accounts and divided mind accounts capture important aspects of self-control. As definitions of self-control, however, each has important limitations: willpower accounts struggle to accommodate the range of strategies individuals use to enact self-control, whereas divided mind

accounts have difficulty explaining self-control in situations where the competing motives do not fall neatly into the proposed framework. Joining aspects of both sorts of accounts into a unified, complex model of self-control therefore has intuitive appeal. Multiple such complex accounts exist. Here, I will briefly overview three such accounts in psychological theory: Kotabe & Hoffmann's (2015) integrative model of self-control, Buckholtz's (2015) multidimensional account, and Gillebaart's (2018) 'operational' definition of self-control (scare quotes Gillebaart's). This discussion is not meant to exhaust all complex models of self-control within psychology, but rather, my aim here is to analyse diverse examples to illustrate both the value of complex models of self-control, and the sorts of challenges they face[13].

The accounts overviewed here each pursue a complex model of self-control that does not reduce the various aspects of self-control into the sort of a divided-mind account described above. Complex models of self-control can be further classified as conjunctive accounts of self-control, which merge dimensions of self-control into one coherent construct, and disjunctive accounts of self-control, which describe self-control, e.g., as either inhibitory control *or* temporal discounting, with the upshot that there are more than one phenomenon that we call self-control, and that the term thus does not yield to a single description or definition.

Each of the models being reviewed was developed for research purposes and thus a reader may question whether it is 'fair game' to treat these models as attempts to fully describe or define self-control: rather, they ought to be treated as operationalizations for the purposes of specific fields of research. I fully endorse making this distinction. In any scientific operation, however, it can be hard to distinguish between studying an operationalization of x and studying x itself; likewise, whether and how new information on an operationalization of x should be taken to imply that we have new knowledge about x is a topic of ongoing debate in philosophy of science. Empirical scientists often make the claim that new information about an operationalization of x equals new knowledge about x; but taking this for granted smacks of operationism, the idea that the extensions of concepts are identical to the operations we do with those concepts (Chang 2019). Operationism can be seen as a form of pragmatism, but a specific form of that: it privileges scientific practice, rather than the various purposes of common usage, as determining the extension of the term in question. In contrast, my approach is that in defining a term in widespread usage, such as self-control, the definition ought to help accomplish the purposes that the common usage denotation has.

[13] It is worth mentioning that the divided-mind approach is itself a type of combined or complex account of self-control since it nests inhibitory control (or 'willpower') within the various means of delaying gratification. Its difference from the compound approaches discussed below is that these compound approaches involve the idea that the divided-mind approach does not exhaust self-control: for these approaches, delayed gratification is not the 'umbrella' under which all self-control can be pooled, but rather is itself nested within a larger framework.

In the case of self-control, such purposes include to help determine whether, in a given situation, one needs to deploy self-control; to determine whether a self-control attempt has been successful; and to inform us in how to provide assistance for individuals who struggle with self-control. An operationalization may or may not succeed in that mission. With these caveats, I now proceed to discuss the models at hand.

Buckholtz's (2015) multidimensional account of self-control is an example of a disjunctive account of self-control. His interest is in using cognitive neuroscience to explain why some people fail to control their impulses, especially in the case of antisocial individuals. To better help these people, Buckholtz holds, it is helpful to understand the neural functioning that enables antisocial behaviour.

In order to understand errant behaviour, we need to understand behavioural control. Like most, Buckholtz starts from divided minds (or, in Buckholtz's more decision theory -based terminology, value-based decision-making) focused on prudentiality on one hand and inhibitory control on the other. However, unlike most accounts, Buckholtz's account denies that these two domains of self-control would always work in synergy. This is where Buckholtz's interpretation of current work in cognitive neuroscience gets philosophically interesting. Buckholtz interprets the seeming mismatch between ways to test, measure, and evaluate motivational conflicts within the domain of decision-making and inhibitory mechanisms and processes in the domain of action control as showing that rather than being 'two sides of the same coin', these are two dissociated dimensions of cognition. Further evidence in neuroimaging, Buckholtz holds, demonstrates that these distinct dimensions of cognition furthermore have dissociable neural substrates (ibid.: 123). Therapy for antisocial individuals could in the future involve ways to help these two domains work in synergy. The below figure highlights Buckholtz's dissociation.

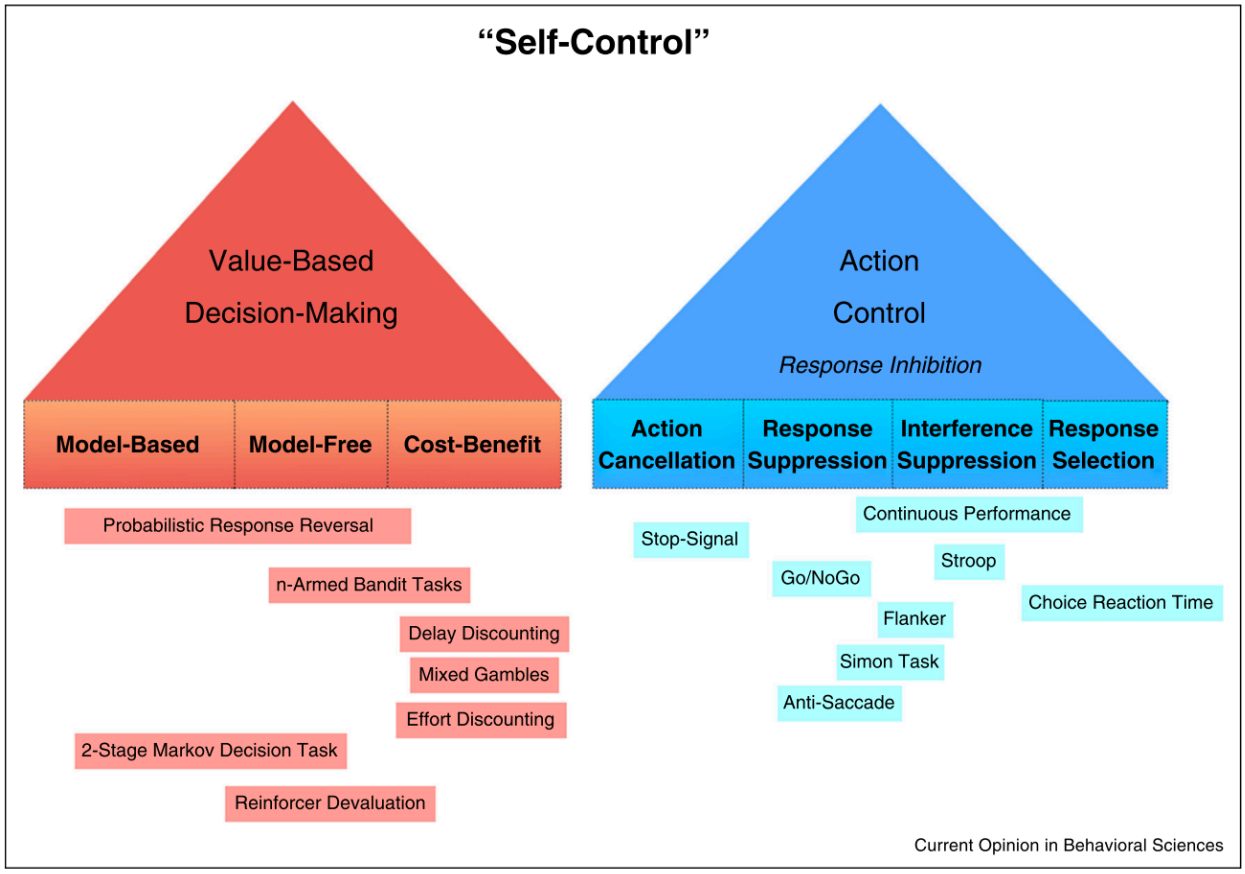

Figure 1. This figure, depicting Buckholtz' disjunctive account of self-control, courtesy of Buckholtz. Reproduced with permission from Buckholtz (2015), p. 124.

Buckholtz makes another surprising statement. In many accounts of self-control involving the two domains, what is implied is that the chain of command goes from resolving the motivational conflict to behavioural control. But Buckholtz argues for the contrary position, at least in the case of antisocial individuals. Ultimately, Buckholtz says, it is value-based decision-making that guides the behaviour that is awry in antisocial individuals; response inhibition and further considerations such as empathy factor into the value calculation (ibid.: 127).

Evaluating Buckholtz's multidimensional model based on how these divided minds and willpower approaches have been discussed above, Buckholtz seems to say that by self-control we mean not one thing but two things. But perhaps the more accurate interpretation is that there is not one but two domains of cognition serving the sort of behaviours that we call self-control. Like any description of the workings of the mind, the multidimensional model is an incomplete, speculative picture – it suggests that some of the ways in which we think about self-control, such as that it would be a single cohesive system, are incorrect, but explicitly makes no claim to be a complete picture of self-control. Buckholtz allows that there may be more domains to self-control than the

two he outlines. His account stresses the heterogeneity and divergent nature, in both expression and (neural) causation, of the behaviours we call self-control.

What does an account like Buckholtz's imply for our conceptualizations of self-control? One outcome could be to conclude, as Herdova (2017) does, that the various phenomena discussed under the term 'self-control' are too diverse to be brought under any unificatory umbrella. In such a case, the best we can do is to develop a taxonomy of the heterogeneous referents of self-control and related concepts. Buckholtz, however, seems to gesture in a contrary direction. Rather, he implies that his two disjunct dimensions of self-control do come together into a unified model that quite resembles divided-mind models, described above. For Buckholtz, not only are there divisions between conflicting motives, there divisions track distinct processes that may guide behaviour in discrepant ways.[14].

Buckholtz's account privileges value-based decision-making, and is therefore susceptible to the same sorts of criticisms as levelled against divided-mind accounts in general in section 2.1.1.: non-hierarchical cases are beyond it given that it explicitly requires a value hierarchy – or, in Buckholtz's terms, "estimating the subjective value of different choice options and selecting actions based on optimal utility" (p. 123). This is not to say Buckholtz's account is false, other than as a definition of self-control *tout court*. Rather, this is to say that it is a description of *some* processes underlying self-control – unquestionably, evaluative processes are involved also when the evaluation ends up in a tie – and accounts such as Buckholtz's should not be mistaken for a description of self-control itself. In and of itself, this account leaves many questions about the nature of self-control and the definition and scope of the term open. One such open question is, on what conceptual grounds do certain domains of behavioural guidance count as domains of self-control?

Rather more complex is the integrative model of self-control developed by Kotabe & Hoffmann (2015), an example of a conjunctive account bringing together various aspects of self-control into a cohesive model. In the integrative account, a self-control process is activated by the conflict between a goal and desire, corresponding to the temporal discounting account of self-control; the control motivation arising from the conflict is then processed, roughly, as an inhibitory control process, and if self-control effort trumps the strength of the desire, self-control is successful — that is, unless environment constraints, which the model places as the final arbiter of self-control success and failure, rule otherwise. The model draws from a broad range of empirical literature and

[14] In the dual-systems model, Buckholtz's action control would likely fall squarely within system 1 processing; in Buckholtz's model, motivations arising from both systems alike fall within the domain of value-based decision-making, with action control being a further process.

describes self-control as having seven distinct components: desire, higher order goal, desire–goal conflict, control motivation, control capacity, control effort, and enactment constraints. That makes for a model both complicated and intuitive, making explicit many of the implied connections between aspects of self-control on various levels of explanation within the psychological literature.

The model is ambitious in that it strives towards exhaustiveness, describing a flow chart of sorts of the self-control process and placing various domains of self-control study at various stages of the process. The purpose of the model, for Kotabe and Hoffmann, is primarily to help scientists working with self-control relate to each other's work; however, they also propose that the model highlights new directions of research and plausible self-control interventions, and argue that the model can be used to calculate predictions of self-control success or failure.

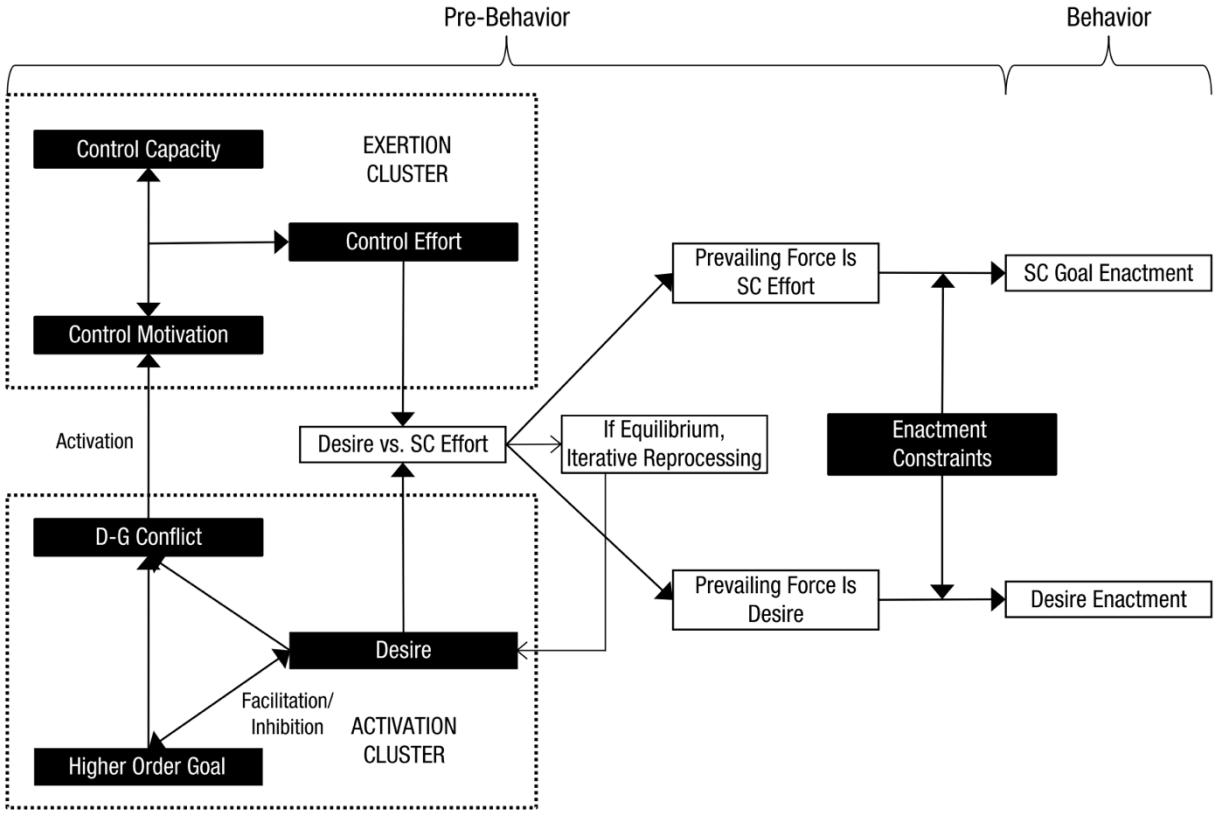

Figure 2. Kotabe & Hoffmann's diagram of their integrative model of self-control. Reproduced with permission from Kotabe & Hoffmann (2015), p. 621.

The Kotabe-Hofmann model efficiently draws together previously unconnected strains in the research of self-control, and rightly shows that environmental enactment constraints can ultimately make or break self-control: even when Ulysses' self-control fails otherwise, his being tied to a mast keeps him from plunging into his death. Describing the multiple nodes that contribute to self-control processes, it draws attention to the richness and nuance involved, enriching the construct of self-control beyond mere delayed gratification. By only counting instances of inhibitory control that

are activated by a desire-goal conflict of the temporal discounting variety, it also rules out trivial cases of inhibition. It is particularly promising in serving as a framework for classifying self-control failures: as several psychological components are required for self-control success, there are multiple junctions along the path to self-controlled behaviour that can cause self-control to fail. However, by doing so, it also loses the content-neutrality of the inhibitory control account and ends up committing self-control to only concern cases where a ranking of motives is present[15]. Additionally, the multiple realizability that gives the temporal discounting model such explanatory force for the self-control successes and failures is lost. While the Kotabe-Hofmann model acknowledges that enactment constraints have the role of a final arbiter in self-control, it does not clearly account for the psychological processes, such as construal, that may act as constituents of self-control, nor does it recognize the role of other environmental scaffolding than last-ditch constraint. As such, despite its many merits, it ends up describing a more limited set of self-control behaviours than either divided-mind or willpower accounts do[16].

Gillebaart's (2018) 'operational' definition of self-control is explicitly a definition of self-control aimed at furthering multidisciplinary interaction as well as the dissemination of research results with the public. This, in short, strives to be a comprehensive definition that brings together both 'traditional' (i.e., delayed gratification, dual systems, inhibitory control) and new directions in self-control research in a manner that would clear up any confusion about what self-control is, and what self-control is not. I can only agree with this aim, given that clearing up the same confusion is what I, too, attempt to do in this dissertation. Gillebaart's starting point, too, is promising. She begins by asking what distinguishes self-control from other forms of self-regulation. Her solution is to refer to the idea of self-regulation as a feedback loop comprising of standards, monitoring, and operating. Standards are the intended effects that the agent desires; monitoring involves the individual noticing whether they are currently behaving in alignment with these standards; and operating is any behavioural interventions that the agent enacts in order to correct course or to enable future behaviour that coheres with the standards. Gillebaart identifies self-control with operating (ibid.: 3).

[15] Kotabe & Hofmann note that in addition to desire-goal conflicts, symmetric desire-desire and goal-goal conflicts are possible; however, they express uncertainty about whether these count as self-control, and refrain from accounting for them in their model.

[16] The Kotabe-Hofmann model has many similarities in structure with Henden's (2008) account, in which "Self-control is a capacity to bring one's actions into line with one's self as it is embodied in what one takes oneself to have most reason to do" (2008, p. 85), a statement well aligned with the desire-goal structure in the Kotabe-Hofmann model. Henden also argues that "willpower" is necessary for particular instances of self-control to succeed, echoing the similarly central role Kotabe and Hofmann give control capacity.

There are many advantages to Gillebaart's approach. First, it allows multiple realizability, and this is clearly one of Gillebaart's priorities: she notes that while temporal discounting and effortful inhibition are part of the bigger picture of self-control, self-control also is realized by various situational strategies (ibid.: 2–3) as well as by creating and fostering habits. People who rate high in self-control respond to environmental cues in ways that support rather than detract from their long-term goals (ibid.: 3). Second, it is easily communicable and helps distinguish self-control from other forms of self-regulation (i.e., monitoring and standard setting).

Gillebaart's account is however too lenient in its definition of self-control: taken at face value, it allows for *all* intentional action to count as an instance of self-control. For example, Kris monitors his internal state and his environment (he is thirsty and in a party), judges what standards are pertinent to the situation (prosociality and comfort), and operates accordingly (walks to the bar, smiles at the bartender and orders a beer). In doing so, Kris brings his current state better into alignment with his standards, but I would not call this an instance of self-control as Kris had no current or anticipated motivational conflict. Instances of self-control are clearly instances of operating, but identifying self-control with operating overlooks a significant difference between instances of self-control and other instances of operating, namely that it is conducted in response to, or in order to avert, a conflicting impulse or motive.

The shortcomings of internalist accounts of self-control, as well as the multiple realizability highlighted by complex models of self-control, highlight a need to foreground the role of interaction between the individual and the environment in self-control beyond what any of the approaches discussed above have done. This is the topic for the following section.

3. Against willpower: the externalist challenge

Perhaps the most famous literary case of self-control is that of Ulysses and the sirens. In Homer's Odyssey, Ulysses wants to hear the Sirens' song, but also knows that any who hear it are doomed to death. He instructs his men to fill their ears with beeswax, to tie him in the ship's mast and to not let him loose however much he pleads. This ingenious plan works, and Ulysses gets both to hear the Sirens and to keep his life. The ties holding Ulysses to the mast prevent him from plunging to his death. Whether or not these constitute a form of self-control remains controversial. In what follows, I discuss accounts that would answer in the affirmative.

Arrangements that resemble that of Ulysses are commonly used to help individuals control themselves. The routine advice for dieters and drinkers striving to learn a healthier lifestyle is to keep one's home clear of their substance of choice, be it chocolate or cognac. These arrangements – the absence of liquor from the home, for example – are the enactment constraints that models like that of Kotabe & Hoffmann (2015) acknowledge. Thus, a bare-bones argument in defence of both multiple realizability and externalism about self-control could be to claim that Ulysses arrangements are functionally equivalent to other forms of self-control, and to argue that if functional equivalence is sufficient for identification with self-control, then there are ways to enact self-control that involve the environment.

However, internalists about self-control need not be too bothered by Ulysses arrangements. They can do precisely what Kotabe & Hoffmann do, which is, to say that there is a certain way self-control works that is decisively agent-internal; environmental constraints are merely something that may, ultimately, prohibit a behaviour that the agent would have otherwise selected. They can then depart from Kotabe & Hoffmann and deem those environmental constraints to fall outside self-control proper: the self-control process, for such an account, would stop somewhere just before environmental constraints. While they can make or break *the behavioural output* of self-control processes in some cases, they do not require a re-evaluation of what self-control is: rather, they simply illustrate that sometimes any given behaviour is unavailable for environmental reasons. For such an account, Ulysses, knowing his self-control will fail, temporarily strips himself of autonomy – for that reading, Ulysses is not self-controlled, but due to his being tied to a mast the behaviour he would succumb to is unavailable, and thus his self-control failure does not lead to his death.

In this dissertation (section 5 and articles 2-3), I defend a much more robust role for environment in self-control than mere Ulysses arrangements can illustrate. Building on previous

scholarship, I will argue that agents learn, select, and enact self-control behaviours in continued interaction with their lived environment. That means that environmental practices are not merely involved in Ulysses-type scenarios where they provide a final constraint for behaviour, but rather, environmental practices and processes may be involved in self-control all the way through. I want to suggest that rather than there being inherently weak-willed individuals and individuals with willpower to spare, our lived environments support self-control behaviours for some individuals while undermining them for others. Some individuals with certain sorts of capabilities and circumstances are enabled to control themselves, whereas others may find epistemic, physical, etc. barriers to the sorts of self-control behaviours that would be suited for their capabilities and circumstances. Self-control thus becomes a matter of access. And if that account is correct, then social, economic, and political axes of disadvantage become central determinants of self-control.

But to make my case, I first wish to discuss other externalist approaches to self-control which have been an inspiration for the novel developments defended here. These include accounts of self-control put forth by Neil Levy (2007, 2011, 2013, 2016, 2017), Tillmann Vierkant (2014, 2015), and Thomas D. Connor (2014) within philosophy and by psychologists such as Angela Duckworth (Duckworth, Gendler & Gross 2016; Duckworth, Milkman & Laibson 2019). It also draws inspiration from feminist and disability-centric accounts of autonomy, such as those of Radoilska (2012), MacKenzie (2008) and Shoemaker (2015), from discussions of the causal and constitutive interaction of the environmental and the endogenous in psychiatry and psychology, as well as from accounts of the socioenvironmental constitution of disability (e.g., Shakespeare 2006, Barnes 2016). In this chapter, my aim is to review the relevant externalist literature that led me to the questions I examine in this dissertation, with a particular eye to the questions raised by these accounts. While my positive theory of individual differences in self-control is described in full in articles 2–3 and section 5, this chapter therefore serves to contextualize it by describing its precursors and the room for amendment I find in those precursors.

3.1. Radical externalism about divided minds

Within philosophy of mind, externalism is the idea that the mind is not fully located within the skull: some of the constituents of the mind may be located outside the individual, or interaction with the environment may be seen as a necessary constituent for the mind. A form of externalism, the extended mind thesis, holds that contents of mental propositions, such as memories, may be outside the agent: for example, information in a notebook is treated as a de facto part of the agent's

memory, given certain constraints (Clark & Chalmers 1998). Further philosophical work has explored the idea that cognitive processes could also be carried out in part outside the skull, comprising of a variety of interactions and feedback loops between the brain, body, and environment (Clark 2007). A burgeoning new research program, that of 4E cognition (e.g., Varela, Thompson & Rosch 1991), has developed to explore the mind as embodied, embedded, enactive, and extended.[17]

Rather than committing to the extended mind thesis per se, in this dissertation I ask whether the cognitive processes associated with self-control – such as inhibitory control and delayed gratification – could also be extended. Clark & Chalmers argued for a form of functional equivalence pertaining to memory; could some form of functional equivalence also pertain to self-control? The case of Ulysses suggests they may be the case under certain sorts of special arrangements, similarly to the Clark & Chalmers thesis. By contrast, the 4E thesis is more demanding: it suggests that it is *characteristic* of cognitive processes that they are not limited to the nervous system, not just that there are some *cases* where ingenuity makes up for internal restraint. In the present thesis, externalism about self-control is defended. The reader need not be worried that accepting externalism about self-control would commit them to 4e stances generally: one might hold that some agentive processes, such as self-control, are situated, whereas others are not.

Neil Levy (2007, 2011, 2016, 2017) presents a compelling case for externalism about self-control. Levy argues against willpower accounts and in favour of divided mind accounts, relying on a nuanced analysis of empirical literature on self-control. I will here discuss Levy's account of self-control in some detail because the account presented in this dissertation builds on it, agreeing with it on multiple points yet suggesting a move away from the naturalist divided mind approach that Levy endorses. After summarizing Levy's views, I will describe mine and explain how they revise and build on Levy's. Other externalist literature on self-control within philosophy (Vierkant 2014, 2015; Connor 2014) is scant, and I will end this section by relating it to my account.

Levy's argument for externalism about self-control starts with Mischel's (1989) experiments with delayed gratification. In these experiments, children were instructed to wait alone in a room with a treat, marshmallows being the notorious choice, with the promise that if they waited until the attendant comes back, they would get two treats instead of one. Mischel found that children who succeeded in a delayed gratification task did so by means of a variety of skills and practices, such as

[17] The extended mind hypothesis has also been subjected to criticism. See, e.g. Harris (2020); Gallagher (2018). Here I will make no attempt to defend it generally, but rather note it as it provides some context for the present work. However, while this contextualizes my contribution to the study of self-control, note that one can accept self-control (or other cognitive processes) as extended without holding that the mind itself is.

self-distraction (directing one's attention elsewhere) and construal (conceiving of the object in a different way in order to make delay easier). The children in Mischel's experiments self-distracted by, e.g., singing to themselves or trying to fall asleep; they construed of the marshmallow as a tiny cloud. Mischel found that some children were adept at these strategies by the age of five. However, the skills and practices seemed to Mischel to be learnable, and his group was able to teach previously unsuccessful children to delay. As a result, Levy (2007: 204, 2017) argues, self-control depends on a set of skills and practices. These skills and practices are vital for thriving: Levy (2007: 205–206) suggests that Obsessive-Compulsive Disorder, trichotillomania, and kleptomania are obvious cases of self-control failures, as are addictive behaviours, such as in substance use disorders. He claims that rather than postulating addiction as a distinct psychological kind, we ought to conceive of addictive behaviours as the same sorts of self-control failures as any failures in delayed gratification tasks. Levy argues that mechanisms of delayed gratification underlie all self-control, and that various skills and practices, many of which involve the interaction of the agent with their environment rather than happening within the agent's mind, can be deployed to support the subject in delaying. These mechanisms, however, likely subserve other behaviours as well (Levy 2017). In the image of self-control painted by Levy, delayed gratification mechanisms kickstart self-control, but they are not sufficient for sustained control; for that purpose, various skills and practices are utilized.

These skills and practices do not stop at construal and self-distraction. Levy (2017) describes how in standard cases where self-control may fail, such as dieting, restricting alcohol consumption, and financial prudence, complex indirect strategies of environmental navigation and manipulation are used:

> I might plan my walk through town so as to avoid the bakery with its tempting smells or choose a supermarket where candy is not placed at the checkout. I may buy ice cream in a small tub, so that if I want to eat more than the tub contains, I have to go to the effort of going out again. I might employ Ulysses strategies, such as placing a time lock on my drinks cabinet so that I can't drink before 6 p.m., or putting my money in a term deposit account which penalizes early withdrawal. All of these are actions I may take that restrict my choices. … Nevertheless, people may choose to do this sort of thing in order to exercise self-control indirectly.
>
> (Levy 2017: 201.)

Defining self-control as the capacity to pursue larger, later rewards that conflict with smaller, sooner rewards, Levy (ibid.) argues that we often enact self-control by structuring our environments. Often these strategies are used to make the smaller sooner reward, which is tempting because it is the easier option, seem *less* easy (or, conversely, to make the larger later option seem less hard: e.g., by laying everything out nicely on one's desk so that beginning to work on a difficult manuscript seems less daunting)[18]. I would add that we furthermore use environmental structuring to create cues and reinforcement for construal and self-distraction. Someone trying to decrease their afternoon tippling may, instead of the Ulysses arrangement of the time lock, structure their environment in ways that promote other activities such as exercise e.g., by putting up sticky notes cuing them to those activities, or sign up for a daily "stop tippling" e-mail newsletter that reminds them of the negative (e.g., decreased sleep quality) rather than the pleasurable (e.g., flavour) aspects of drinking. Conversely, others too can structure our environments to either promote or undermine self-control the use of self-control skills and practices, using strategies similar to nudging.

Relying on empirical literature, Levy (ibid.) suggests that thinking of self-control as a character trait is misleading: self-control is, rather, a set of skills and practices. Levy demonstrates this by analysing measures of trait self-control and how they relate to success in delayed gratification. Based on this analysis, he suggests that rather than measuring an innate trait such as willpower, trait self-control ends up measuring aptitude in the skilful avoidance of temptation. This implies that there are individual differences in the sets of self-control strategies employed by various individuals (i.e., some are more reliant on situational strategies, whereas others are more likely to rely on their inhibitory control); and that it is the former sort, rather than the latter, that is more successful in self-control.

Notwithstanding, for Levy, processes of delayed gratification are necessary for self-control. He suggests that if a temptation persists, eventually, individuals' capacity to delay will fail (2007: 212); strategies of self-control are vital for diffusing temptations before our resistance gives in. This seems to be an accurate picture of most instances of self-control: at most instances, agents appear to use a collection of self-control strategies, either simultaneously or consequentially, rather than relying on one strategy alone.

Levy's conclusion is supported by a range of empirical literature on self-control. The various measures, skills, and strategies of self-control, both internal and environmental, have been a

[18] I owe this point to discussion during Levy's lecture, "Externalist self-control", on self-control at the Turku self-control cluster group's workshop in February 2016. Levy's example was that a time lock installed on a liquor cabinet does not make drinking impossible: it simply makes it harder, as the agent would need to either make a trip to the liquor store or break into the cabinet.

topic of sustained study within psychology and related disciplines. I will here focus on findings by Duckworth and colleagues (Duckworth & Kern 2011, Duckworth, Gendler & Gross 2016, Duckworth, Milkman & Laibson 2019). Duckworth, Gendler and Gross (2016), operating under a self-control concept that closely matches Levy's, classify self-control strategies into five 'families' based on at which part of the self-control process the strategy is enacted*: situation selection, situational modification, attention deployment, cognitive change,* and *response modulation.* These families are grouped into situational (i.e., environmental) strategies and intramental strategies.

Duckworth, Gendler & Gross's (2016) situational strategies comprise *situation selection*, which includes, e.g., avoiding bars and boozy parties during recovery from alcoholism, studying at the library rather than at home (where distractions abound), and surrounding oneself with people whose behaviour one wishes to emulate; and *situational modification*, which includes e.g. Ulysses arrangements, setting environmental reminders, and asking friends to help one stay on track with a resolution. While these strategies match what has above been discussed as environmental selection and environmental manipulation, Duckworth, Gendler & Gross also explicitly acknowledge the role of the interpersonal sphere in self-control. Environmental selection (or situation selection) can involve not just avoiding a bakery, but also, e.g., packing healthy lunches (so as to avoid the need to make choices at lunch restaurants). Environmental manipulation (or situational modification) can include not just setting environmental cues to remind one of one's resolutions but also asking friends and family for support. Kimberly, striving to quit smoking, may tell her friends, "if you see me asking for a cigarette, remind me that I've quit. If you see me with a cigarette, confiscate it and give me a talking to." I would add that Ulysses, too, would have been in dire trouble without the help of his social support network – his sailors. In addition to the modification of physical and social situations, Duckworth, Gendler and Gross also expand the sphere of situational modification to include body modification, such as by gastric bypass surgery to reduce overeating. These situational strategies work by influencing the agent's circumstances and internal state in various ways:

> [O]ur physical, social, and somatic circumstances can influence self-controlled behavior in three principal ways. First, where, with whom, and in what physiological state we find ourselves can influence what we pay attention to. Second, our situations can influence the expected costs and benefits of responses we have yet to enact. Third, our situations can influence our response options themselves.
>
> (Ibid., p. 43.)

Duckworth, Gendler and Gross classify intrapsychic strategies into four groups: *attentional deployment* is when focus is directed to aspects of a situation that support rather than counter our attempts at self-control; strategies of self-distraction are pertinent here. Another intrapsychic strategy of self-control is *response modulation*, that is, suppressing or amplifying a response in the moment, such as by means of inhibitory control. The third intramental family of strategies, *cognitive change,* includes any means to decrease the perceived value of the proximate reward or increase the perceived value of the larger later good: an example is construal, such as imagining the marshmallow as a fluffy cloud, or when dieting, conceiving of the chocolate cake as a blob of fat; another effective method is conceiving of our actions in more abstract terms, increasing perceived emotional distance. (ibid.)

Relatedly, illustrating the various forms that cognitive change can take, Bermudez (2018) has suggested that the *concept of willpower itself* is a framing device that is useful for just such cognitive change because we attach value to willpower. Bermudez argues that when we conceive of these options *in terms of* self-control, since we wish to see ourselves as self-controlled rather than as lacking willpower, this framing effect will decrease the perceived value of resisting the SS reward. For Bermudez, delaying gratification until the larger later reward is attained is less attractive than having successfully resisted a smaller sooner temptation, particularly if the agent believes one or more of the following: that successful resistance is self-reinforcing, that successful resistance is good news about the agent's character (e.g., evidence that she has plenty of willpower), or if the agent attaches symbolic value, such as religious significance, to resisting temptation (ibid.: 201–202).

Duckworth, Gendler and Gross do not stop at stating that situational strategies are among self-control strategies: they also argue that they are more effective than intrapsychic strategies are. This is because first, situational strategies can, to use their term, nip a temptation in the bud: they can help avoid or diffuse a competing motive before it becomes so salient as to require strategies of attentional deployment, cognitive change, and response modulation. Furthermore, they cite empirical evidence, such as the poor performance of healthy adults in simple executive functioning tests, as evidence that intrapsychic strategies, especially the strategy of response modulation, are highly unreliable. While no family of strategies is fool-proof, situational strategies are the better bet (Duckworth, Gendler & Gross 2016) – or, in their more evocative language, "the most effective way to do battle with our inner demons may be, in fact, by taking the battle outside of the mind" (ibid., p. 50). As Duckworth, Milkman & Laibson (2019) later argue, this has important implications for policy purposes. (This dissertation's practical implications are discussed in sections 5.2 and 5.3). This is a conclusion also defended by another self-control externalist, Tillmann

Vierkant (2014). Discussing willpower approaches to self-control, he suggests that the resources of willpower are limited and using them requires effort, whereas using environmental strategies for self-control is effortless (ibid.: 58). In Vierkant's (2014, 2015) approach, both internal ('willpower') strategies and external Ulysses strategies of self-control amount to resolutions, but external strategies are the more effective means of sticking to those resolutions. Vierkant's claim, it should be noted, is slightly overstated – environmental strategies, too, require effort at various temporal junctures across the self-control process. Ulysses, for example, must make *some* effort to persuade his men to tie him to the mast. And for some other situational self-control measures, such as time locks on liquor cabinets, they might facilitate self-control but not entirely ensure it, making it less effortful but not effortless.

Both Levy (2017) and Duckworth, Gendler & Gross (2016) make an empirically well founded, persuasive argument to the effect that self-control is best understood as a heterogeneous set of skills and practices, and that environmental practices are not just something that can act as a last-resort constraint on behaviour but rather are a *central* feature of self-control processes and may figure at any stage of a self-control process.

Both Duckworth and colleagues (2016, 2019) and Levy (2007, 2017) suggest that delayed gratification is a defining feature of self-control. This definition of self-control is a good fit with the heterogeneity and environment-sensitivity of self-control practices described. However, the problems of unorthodox cases of self-control (Mele 1987, 1995) and non-hierarchical instances of self-control, as described above in section 2.1.1., are as relevant for externalist divided mind accounts of self-control as they were for their internalist counterparts. The next subsection is dedicated to this conceptual issue in defining self-control. I will discuss how Levy (2007, 2016, 2017) characterizes self-control. His account elegantly evades the problems associated with most divided mind approaches by defining self-control via the *underlying processes of delayed gratification* rather than as delayed gratification itself. That, paired with the overt externalism of Levy's account, makes his account attractive. However, the evasion manoeuvre connects the concept of self-control to epistemic and metaphysical problems. I next discuss these issues and suggest that a different way of defining self-control would help resolve conceptual issues, including these, in his definition.

3.1.1. Defining self-control as delayed gratification processes?

Levy's account of self-control is squarely within the divided mind approach but develops over time. He starts out from a view of self-control as being used to conform behaviour to values endorsed by

the agent: "I control myself, roughly, when my actions stem from my non-self-deceptively endorsed values" (2007: 200). And because self-control is used to conform our behaviour to what we value, it itself has (instrumental) value. Levy later characterizes self-control as the processes by which smaller sooner rewards are forfeited in order to pursue larger later ones (see Levy 2017: 198, Levy 2016: 814, Snoek, Levy & Kennett 2016: 102). However, unlike most divided mind accounts of self-control, Levy's approach underscores that environmental factors are crucial for self-control. His account is among the first to emphasize rather than merely acknowledge a role for environmental modification in self-control success. Lived environments can both support and undermine self-control, and the self-controlled individual knows how to navigate and manipulate those environments. Importantly both from the perspective of childhood development as well as from the perspective of social policy and justice, others, too, can shape environments to undermine or to support self-control (Levy 2007: 215–216). However, for Levy, self-control also involves mechanistic processes. Specific (divided mind) mechanisms are a central feature of self-control: environmental practices facilitate enacting self-control through those mechanisms.

Levy has, across the years, adjusted his account of what exactly are the mechanisms of self-control in response to empirical findings. Initially characterizing successful self-control as the use of strategies such as self-distraction and environmental constraints in time before the ego depletion effect kicks in and our resources of self-control are depleted (ibid.: 212–214, 2016: 814)[19], Levy finally endorses conceiving of self-control as involving such *mechanistic processes and practices as have evolved for the purpose of pursuing larger, later rewards in the face of competing smaller, sooner ones* (Levy 2017: 198, Levy 2016: 814, Snoek, Levy & Kennett 2016: 102). This allows Levy to include instances like Mele's wayward youth in this definition while holding that delayed gratification is primary:

> I do not claim that all exercises of self-control involve resisting SS rewards in order to pursue LL rewards. There may be "paradoxical" exercises of self-control, in which an agent effortfully resists the temptation to pursue an LL reward and pursues an SS reward instead. If there are such cases, though, I suggest that what makes them exercises of self-control is that they engage mechanisms that evolved for the purposes of resisting SS rewards when they conflict with LL rewards.

[19] Levy's stance concerning the role of blood glucose in self-control is particularly interesting: examining empirical research on the topic, he initially suggests that self-control operations literally expend glucose (2007: 210–211), later revising his stance on grounds of contrary empirical evidence, developing the novel suggestion that glucose is not a direct resource after all (Levy 2016). Rather, glucose restores self-control capacity by means of an evolutionary mechanism that interprets glucose as a signal of environmental richness (ibid.)

(Levy 2016: 814).

I will refer to Levy's group of evolved processes and practices as delayed gratification (DG) processes, for short.

While identifying self-control with DG processes, Levy is cautious about suggesting specific DG processes. He mentions inhibitory control as a subsidiary for self-control and discusses the role of glucose in the bloodstream; overall, he argues that the role of specific underlying mechanisms has been overemphasized, and that self-control relies more on (environment-centric) skills and practices than on specific underlying mechanisms. In the face of this cautious approach, it can appear surprising that in his identifying self-control with DG processes, Levy, nevertheless, seems to hold mechanisms (whichever mechanisms they are[20]) to be a defining feature of self-control. This is a point that should not be overemphasized: in discussing the related concept of trait self-control, Levy (2017b) notes that concept taxonomy is less interesting to him than understanding how it is that people succeed in pursuing larger later rewards, and he does not dwell on justifying his specific account of self-control but rather, engages in exploring how the behaviours we label self-control work. Elsewhere, he has maintained that he is uninterested in definitions.

In what follows, I will however treat his account of self-control as DG processes qua a definition. I justify this move as follows: first, Levy's choice of phrasing ("what makes them exercises of self-control…") is strongly suggestive of the interpretation, whether or not intended, that reliance on DG processes is both necessary and sufficient for something to count as an instance of self-control. Second, even if this is not intended as such by Levy, others might, not unreasonably, maintain such a view. Third, part of the contribution of this dissertation is to offer a definition of self-control (in section 4), and so I find it worthwhile to explain why I could not simply employ Levy's characterization as such a definition but instead need to develop an alternative to it.

With the above caveats, then, I will here sketch out some reasons why Levy's characterization would not work as a definition of self-control.

Characterizing self-control as DG *processes* groups together mechanisms and processes that count as self-control in virtue of their evolutionary purpose. Compared to identifying self-control with delayed gratification, such as, e.g., Duckworth and colleagues do, this has the advantage of elegantly including unorthodox instances of self-control, such as Mele's wayward youth, as well as non-hierarchical instances of self-control as described in the present work. These instances count as

[20] Furthermore, in psychology, 'mechanisms' can include almost any causal factors leading up to outputs, including higher-order cognitive processes – heterogeneity that is in part due to differences among psychological traditions (Koch & Cratsley 2020).

self-control simply because they rely on the same processes as delayed gratification does. Situational instances of self-control, too, can rely on such processes. Indeed, the relevant processes might well be processes by which individuals manipulate and navigate their environments.

If I am right that non-hierarchical instances of self-control are not atypical but, rather, a common sort of self-control, it may be objected that conceiving of self-control in terms of DG processes regardless overemphasizes the role of delayed gratification instances in self-control. On its own, I think this objection lacks force: processes subserving both non-hierarchical and hierarchical instances of self-control, even if their evolutionary history is that of enabling delayed gratification, could simply be labelled something else which would yield us the de-emphasis requested. However, there is further reason to think identifying self-control with DG processes (under any name) to be unhelpful as a definition, for reasons unrelated to the variety in motivational structures in self-control. That is, if employed as a definition, this turns out overly inclusive. Elsewhere (Levy 2017), Levy, too, suggests that DG processes are not dedicated features of self-control but subserve other behaviours, as well. However, he does not discuss the implications of that insight for his account of self-control further.

This is something that I will further discuss in section 4.1., where I will explain why I think relegating DG processes to the same status with any other processes subserving self-control is the better move, and propose a simpler definition of self-control that in my view gives philosophers and scientists working on self-control what we need.

3.1.2. Problems for externalism about self-control

Some criticisms against externalism about self-control have been mentioned and briefly responded to in the above sections. Here, I wish to expand on what I consider the most sweeping objection against externalism: the objection that agents using situational strategies are not engaging in self-control, but rather, in something other than self-control, such as, in the avoidance of the need for self-control. I will also respond to an objection concerning the usefulness of a broad conceptualization of self-control for science.

There are two reasons why a philosopher might claim that situational strategies fall outside the scope of the term self-control: one, she may defend a narrower conception of self-control, e.g., a conception of self-control as inhibitory control. Such narrow conceptions have been argued against in section 2, but it makes sense to respond to this objection here in any case.

The other reason why someone might claim situational strategies to fall outside the scope of the term self-control is that they hold that an adequate conception of self-control is one that is well in line with the common usage of the term (e.g., Sripada 2020 endorses this view). I hold that same view, albeit in a qualified form – in my view, departures from common usage are problematic only when unjustified.

Let us begin with the latter form of the objection, i.e., with the claim that situational strategies would fall outside the common usage of the term. If read as a categorical claim, the claim is false. Of course, natural language is subject to interpersonal variance and thus it would be a fool's errand to claim that everyone's conception of self-control included situational strategies. That being said, we do routinely commend the self-control of people who go on morning runs having set their sneakers by the door, who use apps or nicotine patches to help themselves quit smoking, or who, having decided to go on a diet but finding a tray of cookies in their hotel room, toss the cookies in the bin. We would label these people foolish rather than more self-controlled if they instead opted to make no adjustments in their living circumstances to make running easier, prided themselves in accepting no help, whether social, digital or chemical, in quitting smoking, or engaged in hotel room staring contests with the cookies they wished not to eat. It is true that some situational strategies are such that it would be unusual to call them self-control; Duckworth, Gendler & Gross's gastric bypass example is one such strategy. But I think here, Duckworth and colleagues are biting the bullet and following their claims to their logical conclusion, even when that conclusion surprising; This is exactly what any conceptual revisionist ought to do (as long as there are no strong reasons to reject the surprising conclusion).

While the externalist conceptualization of self-control is not an exact match to common usage, if this is a problem then it is a general problem for all conceptual work and for most of philosophy, rather than a problem for my conception of self-control. To sum up, externalist accounts of self-control, as given by Levy (2017) and Duckworth, Gendler & Gross (2016) capture roughly the right sorts of behaviours in their conceptions of self-control and natural language-based objections do not give sufficient grounds to their rejection.

Now to the first form of the objection, namely, that it is not a problem that situational strategies would not be universally called self-control; rather the problem is that they *are not* self-control, regardless of whether people label them as such. Instead, they are something else, such as examples of arranging one's life so that self-control is not needed. This form of objection presumes that there is a way to distinguish between self-control and non-self-control that is not equivalent to our linguistic practices, and that this distinction is one that would exclude situational strategies, specifically.

For this line of counterargument, my response is to insist on the multiple realizability of self-control. Again, I mean multiple realizability in a broad sense: not only are there multiple physical ways to realize a given process or mechanism, multiple sorts of processes and mechanisms can realize the same behavioural function. For the view endorsed in the present dissertation, equating self-control with a discrete mental faculty makes the concept both too narrow and too broad, as candidate faculties such as inhibitory control are also used for processes that are not self-control. There seems to be no discrete process we could label self-control; rather, self-control is distinguished by its function. Even for accounts equating self-control with inhibitory control, inhibitory control may be realized by various processes, and they qualify as inhibitory control processes because of their shared function (of inhibition). I agree with the inhibitory control theorists about self-control that it is rather intuitive to define self-control functionally. The controversy concerns *which* function – that of delayed gratification, that of inhibitory control, or the one I suggest in section 4 – is the relevant function.

However, *all* the candidate functions are, arguably, subject to multiple realizability. The above discussion makes it clear how delayed gratification is subject to multiple realizability; but how about inhibitory control? Here, I presume that one who believes externalism about self-control to be overly extended beyond self-control proper would not have that same objection to defining self-control as inhibitory control, either because she would equate self-control with inhibitory control or because she would hold that inhibitory control often is part of self-control.

As discussed above, however, inhibitory control and situational strategies are not quite as far apart from each other functionally as they are materially. Indeed, a case could be made for the *functional equivalence* of some or all inhibitory control and (some) situational strategies of self-control. My way of working out their uniting function will be expanded on in section 4, where I define self-control in a way that responds to the problems faced by Levy's account of self-control above, but that retains the multiple realizability of self-control. The functional equivalence of situational strategies and inhibitory control has been discussed, among others, by Heath & Anderson (2010), who argue for it, and by Vierkant (2014), who accepts it but defends an even stronger connection between the use of 'willpower' and external props[21]. Given that it is functional equivalence that unites self-control in narrower senses of the word (such as self-control qua inhibitory control), and that situational strategies can fulfil those very same functions, it seems like the philosopher who wishes to dismiss situational strategies as non-self-control would require some

[21] For Vierkant, the two are not just functionally equivalent, but also, the 'will' itself is extended. Here, I won't discuss that very controversial claim further, as for my purposes, the functional equivalence thesis suffices for dismissing the objection in its present form.

further rationale for accepting some instances of functional equivalence but not others in order to make her case.

A final sweeping objection to externalism about self-control is purely methodological, and touches on one desideratum for conceptual work – namely, that the concept should be useful for the work we do with it. For self-control, the concept is also used for scientific research. If self-control is realized, *as well as researched*, by such diverse means, to what extent does evidence about specific operationalizations of self-control for the purposes of specific experimental designs translate into evidence about self-control, as a whole? Can we meaningfully, e.g., think that evidence concerning inhibitory control tasks at laboratory settings would be informative for self-control, given that self-control is so highly environment-dependent?

A meta-analysis by Duckworth & Kern (2011) provides some reassurance. The meta-analysis, including 236 studies, whether there is significant and substantial correlation among measures as strikingly different as executive functioning tests, delay of gratification tests, and questionnaires, operating not just under different methodologies but also, often under different (but overlapping) concepts of self-control. Duckworth and Kern found moderate convergence overall. The correlations were strongest among informant-report questionnaires, self-report coming next; weakest correlations were found within executive functioning tasks. All measures of convergent validity were found statistically significant. The implication is that while the phenomena being measured by these varied research methods may not be the same, they are correlated. This suggests that measures of self-control by all these methods do have evidential value for self-control; however, this evidential value is qualified. The results of the meta-analysis are consistent with the idea that self-control is realized by a heterogeneous set of skills and strategies.

However, a further study by Saunders et al. (2018) involving a Bayesian meta-analysis of self-control measures indicated that self-report measures of self-control have a null relationship with executive functioning tests associated with inhibitory control, such as the Stroop task. These correlations had been found weak yet statistically significant by Duckworth and colleagues. Meanwhile, the trait self-control scale was highly correlated with measures of grit and of the conscientiousness personality trait. These observations are consistent with the worry that some operationalizations of self-control within personality psychology would simply measure conscientiousness by another name[22].

These observations indicate a need for cross-disciplinary conceptual work on self-control. First, there is broad consensus that self-control and conscientiousness are distinct phenomena –

[22] I owe this point to discussions with personality psychologist Colin DeYoung.

recall that none of the accounts of self-control described above conflated it with conscientiousness. Conceptual work, such as is done in the present thesis, may help build operationalizations that improve the separability of individual differences in self-control from individual differences in conscientiousness.

Second, conceptual work may prove helpful in finding out how exactly it is that inhibitory control and self-control connect. There is a possibility that an operationalization of individual differences in self-control that would better distinguish it from conscientiousness would correlate better with individual differences in inhibitory control as measured by executive functioning tasks. However, we should not be too concerned if this were not so. There is no question that executive functioning plays an important role in many self-control behaviours. However, it is less obvious that particularly high measures of it would be needed for particularly high measures of self-control. When it comes to inhibitory control, if various situational, construal, and other strategies are as effective as they seem, a little inhibition may go a long way.

The speculation offered in the above paragraph suggests that psychological science could benefit from conceptual work, such as is offered here; and that narrow definitions of self-control, such as ones that describe it via inhibitory control, may not be the best way to move psychological science forward. Sometimes a broader construct is best suited for the task at hand. Externalism therefore need not be rejected for progress in psychological science to be made.

3.2. The temporal dimension in self-control

The literature review of this dissertation is drawing to a close. One last theme in the literature remains to be briefly described, as it would be a gross omission to neglect to mention it. This is the temporal dimension of self-control. For many, the experiences of warding away an ongoing temptation and of preventing the later onslaught of one are sufficiently different and require a distinct treatment each. The result, for these authors, is divide self-control into two kinds: synchronic self-control and diachronic self-control.

Those defending the distinction often reckon that situational self-control is always diachronic. It is true that many standard examples of diachronic self-control, such as Ulysses arrangements, are also examples of situational self-control. However, the agent-internal / situational distinction is not the same distinction as the diachronic / synchronic distinction. Indeed, as the reader may have remarked, I have not said much about the diachronic / synchronic divide prior to this section, and will not say much about it after this section. Rather, I've attempted to avoid that divide altogether as

I hold it has some puzzling features that would either require a distinct investigation to be resolved, or else they might confuse rather than clarify.

Much of philosophical debate concerning self-control has revolved around the possibility of synchronic self-control, which some find puzzling. The puzzle arises from two claims that have been interpreted as standing at a paradoxical relation: one, that whenever we do something, we do what we most want to do; and two, for synchronic self-control, one tries not to do what one most wants to do at the time (see Kennett & Smith 1996; Mele 1997). As understanding of human psychology has progressed, the claim that our actions always match what we judge best or most desirable no longer enjoys the axiomatic status it perhaps once did. In this dissertation, sections 2.1 and 4.1 include discussion of agency in the absence of, or contrary to, best judgment or of value hierarchies in other guises. When assumptions of revealed preferences and of the supreme motivational force of best judgment are dropped, synchronic self-control ceases to puzzle.

Following the debate on this puzzle, however, the distinction between diachronic and synchronic self-control has persisted in the philosophical literature. Synchronic self-control is self-control that occurs "at the same time as the rebellious desire" (Henden 2008: 74) or self-control "directed at a desire that is currently active" (Sripada 2020: 2). Diachronic self-control, by contrast, is anticipatory in nature. For Sripada, diachronic self-control, which he defines as preventing "an unwanted, non-occurrent, anticipated future desire from becoming active", is not self-control at all (ibid.); Sripada would characterize most of what I call situational self-control as something other than self-control simply because much (though not all) of it is anticipatory.

Henden (2008: 74) characterizes diachronic self-control as follows: "we anticipate at an earlier time ourselves losing control at a later time, and seek to avoid this". These definitions diverge: for example, a smoker trying to quit desires cigarettes as soon as he walks into a bar (having habituated into associating bars with cigarettes). However, he anticipates himself losing control only after a few beers. When he tells his friends to not let him bum cigarettes from other patrons, he's enacting situational self-control, which is synchronic for Sripada (since the desire is occurrent). For Henden, it would be both synchronic and diachronic (since the desire is currently active, but the loss of control is anticipated).

This divergence in demarcations of the synchronic-diachronic divide helps show that the divide is not as clear as it is sometimes taken to be, or at least, there is no received understanding of how that division is to be drawn. There seem to be multiple components to self-control that, for self-control to occur, need not all begin and end at once. These components include the intended course of behaviour; the conflicting motivation; and the measures, whether agent-internal or situational, taken to steer behaviour. As the two diverging definitions demonstrate, it is not clear by

which part of self-control we ought to set our stopwatches. Sripada's account is helpful in making it clear that for his conception of self-control, it counts as synchronic when the relevant sequence of processes occurs concurrently with the desire: for Sripada, behaviours that happen before a desire is active – such as Ulysses asking his men to tie him to the mast – are not self-control at all. But it is less clear how we ought to treat responding to weak desires that one foresees strengthening over time. Consider Matt, a diligent academic and an avid sunbather who is preparing a laboursome report for a funder on a sunny day. Going to his office, Matt looks at the sunny weather, sighs and thinks longingly of the beach – a desire that is currently weak enough to not require intervention. Matt, however, considers it likely that the desire grows stronger over the day. He also considers it likely that as it grows stronger, he will drop the task halfway and flee his office to the beach. He decides to use a social strategy of self-control to help himself complete the report, and calls up a friend, asking to go to the beach together but only after the report is done. Many instances of self-control seem to be like Matt's: upon having a weak desire that conflicts with another motivation we wish to align our behaviour with, we may project that the desire is likely to strengthen and decide to intervene by enacting measures of self-control[23].

Another reason to consider the divide not particularly helpful is that all behaviour, including cognition, takes time. Even reflexive behaviour takes precious milliseconds. For many behaviours typically classified as synchronic self-control, they take a surprisingly long amount of time: for example, resisting the urge to look at one's phone during a boring lecture (an example of synchronic self-control gleaned from Sripada 2020) may involve extended mental deliberation and the rehearsal of one's reasons. So synchronic self-control is not synchronous in the strictest sense: awareness of the desire and the instance of self-control both are temporal processes and may only overlap in part. The difference between "synchronic" and diachronic self-control then seems to be that "synchronic" self-control extends over a short stretch of time, and diachronic over a longer stretch, resulting in obvious sorites problems. This need not be a problem if synchronic and diachronic self-control are seen to be "two sides of the same coin", but for holders of stronger views, e.g., that only diachronic self-control is actional (Kennett & Smith 1996) or that there is no diachronic self-control (Sripada 2020), a bright line between synchronic and diachronic self-control is needed if one is to be able to tell for any behaviour whether that counts as actional, as self-controlled, etc. using those accounts.

[23] For the conception of self-control I defend, it makes no difference whether the motivation is weak or strong: Matt would be enacting self-control if he called up his friend even if he presumed the desire to stay weak. I mention this since it is relevant for authors working under the DG paradigm, for whom are the relative strength of desires, goals, impulses and the like is of the essence.

One easy solution for this worry would be to specify that the synchronicity of desire and resistance in synchronic self-control does not refer to synchronicity in terms of objective measurements of time, but rather, to both occurring within the specious present – i.e., within the agent's impression of the present moment.

But this latter solution, while promising, becomes less so when we consider that desires are considered active, I presume, when the agent is conscious of having a desire. Such desires, however, flit in and out of consciousness. The smoker trying to quit, while at the bar, may forget about his desire to smoke for periods of time – say, he will forget about it whenever his friends have interesting things to say and remember the desire whenever the company is boring. More so, the agent may cease to desire a cigarette altogether when his friends are particularly entertaining; or, at least, the agent may believe to have ceased to desire it. What, then, should we think of a desire that has become put 'on hold' in that manner? These desires are clearly not ones in the specious present. If an agent enacts behaviours for purposes of self-control during such a break in a desire, does this count as synchronic or diachronic self-control? Is such a desire still occurrent in the sense Sripada maintains? Or are such desires non-occurrent desires that are anticipated on grounds that one just had the very same sort of desire, and thus judges it likely that one has it again soon?

Also counterintuitive is the further implication that successful synchronic self-control ceases to be self-control: when an agent is successful in synchronic self-control and, e.g., successfully suppresses the desire to smoke, then the desire is no longer concurrent. In that case, synchronic self-control can be either attempted or it can fail, but synchronic self-control directed at suppressing desires can never succeed because upon succeeding it ceases to fulfil the definition of self-control Sripada endorses.

Of course, one may define the synchronic / diachronic divide ostensively, by reference to paradigmatic cases of each. But if the picture I have been painting of self-control across this dissertation is correct, then a large swath of self-control is non-paradigmatic, utilizing a combination of strategies rather than just one strategy, occurring in 'messy' ways including under 'messy' motivational architectures. And if that is correct, then an ostensive divide would have limited applicability.

The following problem may be posed for proponents of diachronic self-control. This approach, too, can be seen to give rise to a temporal sorites problem of its own, namely how long a swathe of time may elapse during the self-control process. If we hold that Ulysses, from entering into an agreement with his men to successfully sailing past the Sirens, is enacting self-control, what about similar arrangements that take a very long time? For example, consider Mike, a 20-something entrepreneur. When Mike's grandmother is diagnosed with Alzheimer's yet refuses to take her

medication, Mike becomes concerned that he, upon growing old might do the same. But Mike wants to prevent his future self from succumbing to a desire to reject medication. So Mike, consulting his attorney, draws up an advance directive ensuring that if future Mike refuses Alzheimer's medication, the medical staff will not listen to his pleas any more than Ulysses' men listened to his cries. Is Mike enacting self-control?

This is a tricky question, and I am somewhat agnostic about how one ought to go about it. One possible approach would be simply to bite the bullet and accept that this is a form of self-control; another would be to place some special constraints on what counts as self-control, such as that self-control is not the proper denotation of such forward-looking agency that involves predicted personal transformation. For such a line of argument, Mike is simply too different from future Mike with Alzheimer's for self-control to be the right concept for the purpose; a similar line could be pressed for transformative experiences that take a lesser time, say, measures undertaken to act in a certain way after college or after becoming a parent, if Paul's account of transformative experiences (Paul 2015) is correct. For such a constraint for the concept of self-control, self-control requires, very roughly, that the agent can reasonably expect to have roughly the same sorts of psychological states throughout the temporal extent of the self-control process. I do not wish to make any claims about the validity of this way of delimiting the scope of self-control one way or another; rather, the purpose of this quite underdetermined sketch is just to illustrate that there are indeed possible lines of defence against the extreme temporal scope criticism. For now, however, I am happy to leave this question open. In any case, the existence of problem cases like Mike's seems not to be reason enough to push for a sharp synchronic / diachronic divide, given the problematic nature of the divide.

Likewise, I have made no attempt here to resolve the question of whether desires flitting out of consciousness are occurrent, but have simply mentioned this to note that this is a complication for those wishing to base their notions of self-control on the synchronic / diachronic divide. As such, the main purpose of this section was to motivate my lack of utilizing the synchronic / diachronic divide in the bulk of my work on self-control by showing that this divide is not very helpful. If clear ways to distinguish between synchronic and diachronic self-control are developed, then there is some hope that the divide turns out more helpful than it currently seems. In effect, this section has perhaps raised more questions than it has answered. This, however, is a feature it shares with its subject matter of the synchronic/diachronic divide.

4. Revising our concept of self-control

In this section, I specify the behavioural function that unites different forms of self-control, and urge a radical revision of the concept of self-control. I claim that our concept of self-control should be revised to allow for there to be self-control in the service of any intention. I explore difficulties with restricting self-control to the service of intentions aligned with value hierarchies, such as is done within the DG paradigm. These difficulties arise from the use of self-control against one's better judgment, and the use of self-control under decisional uncertainty. Conceiving of self-control without recourse to value hierarchy makes the concept consistent across each motivational landscape. After I have made my case for the value-deflationist revision, I explore a charge made by Marcela Herdova (2017) that self-control, not being a natural kind, would be a gerrymandered concept, and suggest self-control is best understood as a practical kind.

4.1. Value-deflationist self-control

As outlined above, agents sometimes use self-control to enact something that is contrary to their better judgement or to their larger later goals, instead using it to pursue an aim contrary to their better judgement (recall Mele's wayward youth) or a smaller sooner goal. Agents, furthermore, are not always in possession of a clear hierarchy between the better and worse aim, or between the larger or smaller goal: many decisions are made under ambivalence. Agents make choices *without* having resolved what is best when deciding under an unresolved value conflict as well as when options seem equal to them. It is unclear why the concept of self-control should be restricted only to cases where the agent's intention matches her value hierarchy. Instead, self-control is best understood as something that agents can succeed and fail in regardless of how their intentions relate to their evaluative judgments. To achieve that understanding, I recommend a radical revision of the concept of self-control, as follows:

> *Value-deflationist self-control*. Self-control is that which is enacted to align one's behaviour with intention in the face of competing motivation.

Intention, here, is to be understood in the Bratmanian sense, as a commitment to a plan of action (Bratman 1987). Motivation refers to any competing course of action with some pull on the agent, whether that pull is due to reason, affect, desire, or urge.

What received accounts of self-control and the account I defend here have in common is that for each of them, self-control is a thick concept: it is evaluatively loaded in addition to its descriptive content. However, my account differs from other accounts in locating the normative content of the concept of self-control not within the self-control process itself (such as in forgoing inappropriate behaviours in favour of appropriate ones) but rather, outside it, in our attitudes *towards* self-control.

Due to its value-deflationist character, this revision allows for separating self-control from intention formation. While intention formation and efficacy processes often interact, much confusion about self-control dissipates if intention formation is treated as distinct from self-control and the latter is treated as pertaining to agential efficacy. There is enkratic self-control, but it adds something further to self-control – namely, a specific sort of value-hierarchical intention formation process that is not contained in self-control *per se.* The further question concerning what grounds one's choice to align one's behaviour in one way instead of another is not a question about self-control specifically, but rather a question about intention formation more generally.

The relationship of the resulting behaviour to value hierarchies, if any, is then not a defining characteristic of self-control. Rather, promotion of the larger or more valuable aim is an accidental feature of some, but not all, instances of self-control. For this conception of self-control, there is no "paradoxical" self-control.

In claiming that my value-deflationist definition of self-control is preferable to the DG approach, I, naturally, need to explain with regard to which desiderata it fares better. For that, my foundation will be the work we do with the concept of self-control.

Some concepts, such as 'mitosis', need to be very narrowly defined in order to be useful, whereas other concepts, such as 'shelter', are used to denote a fuzzy cluster of similar phenomena, brought together by their behavioural function. Self-control is more like the concept of shelter than it is like the concept of mitosis. It is too broad a concept for carving nature at the joints, or for fine-grained scientific work about human biology or (neuro)psychology: these disciplines have an array of narrower concepts that are better suited for scientific advances, such as the concepts of delayed gratification and inhibitory control. These concepts characterize subprocesses that are present in some, but not all, self-control. Self-control should be *associated* with delayed gratification and inhibitory control (as with all other processes that subserve it, such as attention), not *identified* with it.

Behaviours we typically label self-control are a diverse set. They include both intramental behaviours, such as 'steeling oneself' against an impulse, rehearsing one's reasons to forgo temptation, construing the competing motive as less appealing (such as, when dieting, thinking of a piece of cake as a blob of fat); and situational strategies, such as Ulysses arrangements, physically avoiding situations where unwelcome desires are present (such as avoiding bars if recovering from alcohol abuse, or by walking away when being heckled or slurred gives rise to a violent impulse). As Levy (2011, 2017) argues, self-control is thus subject to multiple realizability. It denotes a diverse set of ways in which agents regulate their mental states and interact with their environments, rather than an intramental capacity for 'willpower'.

This heterogeneity has prompted Marcela Herdova (2017) to argue that self-control is a "gerrymandered property" and a "hodgepodge" (ibid.: 731) rather than a natural kind. Here, I hope to demonstrate that even as the concept is not a natural kind concept, contrary to Herdova's claim, it can nevertheless be conceptually coherent and informative (for work on concept utility, see, e.g., Egré and O'Madagain 2019).

Even as the concept of self-control is not used to inform fine-grained scientific work, it is used for a variety of other purposes. These purposes can be grouped in two classes: the purpose of *communication* and the purpose of *agentive self-understanding*. The communicative function of the concept of self-control is employed across a range of domains. Within disciplines such as social psychology or philosophy, the concept is used to pool those processes together by their behavioural function – such as the function of suppressing impulses or the function of guiding behaviour towards the larger later goal. The concept of self-control is used to communicate across disciplinary boundaries (such as psychology, neuroscience, and philosophy). It is furthermore used to communicate between professionals and laypeople, such as within the setting of education, and in the clinical setting, where a patient who bemoans she has problems with self-control is hopefully met with understanding by the clinical practitioner. Finally, the concept is used by laypeople to communicate with each other ("I have been trying to quit smoking, but it seems I have no self-control").

The second use of the concept of self-control is *agentive self-understanding.* Individual agents either implicitly or explicitly use their conception of self-control to identify whether a situation requires self-control, what behaviours are available as self-control strategies, and what can be achieved by those behaviours; to assess whether an attempt at self-control has been successful or not; and to seek and provide help in self-control (among peers, from professionals, or within a caregiving relationship).

The DG conception of self-control does a decent job in all these tasks – for those cases of self-control that match it. However, it struggles to accommodate cases of self-control where the intention does not reflect the agent's larger later goal, whether because the agent has formed an intention that is contrary to that goal, or because she has not judged which goal is the 'larger later' or more valuable one: sometimes, decisions about what to do are made from what Isaac Levi (1986) calls 'a position of suspense'. As I will show, the value-deflationist conception of self-control represents an improvement on the utility of the concept because it enables the use of the concept for communication and agentive self-understanding also in cases of decisional ambiguity or acting against one's better judgment.

Levy's definition of self-control as DG *mechanisms* instead of DG itself sidesteps this problem: *if* the same processes of self-control are used from the 'position of suspense' as are used in DG instances of self-control, then we end up with the correct set of cases. For Levy, what makes errant cases of self-control count as self-control is that they make use of the same underlying processes and mechanisms that have "evolved for the purpose of" standard self-control (Levy 2016: 814)[24]. While Levy does not mount a defence of these claims, I will here discuss them at length because they offer a naturalist strategy for granting primacy to value-hierarchical self-control. That, paired with the convincing account Levy makes for externalism and multiple realizability about self-control, makes his account extremely attractive. However, this attractive account has epistemic and metaphysical limitations that I will discuss below, due to which there is space for improvement in Levy's account.

These problems stem from the heterogeneity of self-control behaviours. Given the extreme breadth of the various means to enact self-control on either the standard conception or the one I am proposing, are there any discrete processes that would have evolved for the purpose of standard self-control? Levy suggests that delayed gratification processes would be such a unifying feature. When these processes are engaged to enact behaviour other than *de facto* DG, that behaviour would thus count as self-control.

Here, however, lies a problem. Considering, as Levy does, the multiple realizability of delayed gratification and the striking heterogeneity of practices, skills, and mechanisms that undergird it, we are forced to include a wealth of mechanisms and processes into the scope of processes involved in the enactment of delayed gratification. These would include not just

[24] One way to criticize Levy's approach here would be to object to its teleological interpretation of evolution. However, I am doubtful Levy intends to commit to teleology here – telic shorthand is ubiquitous. Rather, his phrasing is better interpreted as shorthand referring to mechanisms (of an evolutionary origin) whose *adaptive function* rather than *purpose* is the pursuit of larger later rewards in the presence of smaller sooner temptations, regardless of whether the teleological interpretation is true.

inhibitory control mechanisms, but also processes of attentional selection and attention guidance, cognitive processes such as memory, problem-solving, and imagination (such as are relevant for construal, framing effects, and the like), and processes of environment selection and manipulation. Social interaction mechanisms are involved, e.g., in the case of Ulysses, who asks his men to tie him in a mast; or in the case of Descartes, who made numerous promises to deliver a manuscript for the explicit purpose of thus coaxing himself to finish it (see Davenport 2006: 66).

As a result, it looks like almost any cognitive mechanism can participate in DG. But of course, the *participation* of a process or mechanism was never at issue: it is noncontroversial that a variety of processes subserve self-control. Rather, the question was whether, as Levy suggests, self-control can be defined by reference to some processes or mechanisms *whose function is to enable DG,* and if yes, which ones? My best suggestion here would be attention guidance and forward-looking planning processes, as those participate in most self-control strategies and would therefore be good candidates for Levy's DG mechanisms. However, they are subject to the problem of generality, which I suspect would be an issue with any process or family of processes that would be present in all of DG. Levy (2017) notes the generality of DG processes, but does not pan out its implications; I will undertake that here.

Any mechanism or process that are likely to subserve, and have "evolved for the purpose of", DG would necessarily be too basic and general a cognitive process to be helpful for demarcating self-control: a process that general would necessarily be involved in a broad swath of human functioning, not just in self-control. The processes and mechanisms involved in DG are too many to be helpful for demarcating self-control, and too general to be helpful for distinguishing self-control from other forms of agency. For example, inhibitory control could be such a process, but it is also active in hesitation. Likewise, evaluative processes are an integral part of DG, but they are also active in all preference formation, including when there is no motivational conflict, and if I am correct then an occurrent or anticipated motivational conflict is necessary for something to be self-control. Of course, we can be hopeful that only a small set of the many processes that subserve DG have "evolved for the purpose". But even such processes would be basic enough that they would also subserve behaviour that is not self-control, such as other sorts of action selection, endogenous regulation, and behavioural guidance. Another problem is that reliance on processes in demarcating self-control makes us unable to distinguish between self-control successes and failures, as failed attempts, too, include the same sorts of processes as successful ones do.

If the problems of generality and multiplicity are as foundational as they seem, defining self-control in terms of processes or mechanisms – whether on the basis of an evolutionary explanation or not – gives us a definition that is not helpful in distinguishing between self-control and non-self-

control. To know whether Mele's wayward youth, Andy or Joanna have in fact enacted self-control or not, we would need to know whether the processes their behaviour rests on were the right sorts of processes. But the problems of generality and multiplicity prevent us from finding that out, at least in the foreseeable future. And insofar as whether these agents are enacting self-control or not is opaque – to them, to their loved ones, to philosophers, to clinicians, or to experts in human evolutionary biology – we are prevented from doing the work that the concept of self-control was supposed to help us do. On these grounds, I would conclude that the mechanistic solution falls short of the mark.

Not only is the mechanistic solution epistemically problematic, it is also not what we, typically, do when we identify something as self-control. When we ask whether Johnny has enacted self-control, we do not typically ask about cognitive processes or evolutionary functions. Rather, the sort of questions we typically ask include questions about his mental states and his overt behaviour. We ask whether he ended up doing as he intended. We ask whether he had to tackle a competing motivation on the way from intention to action[25]. When Johnny notices that the situation seems to require self-control, it is because he has identified an occurrent or anticipated motivation that runs counter to his intention and judged that he needs to rely on certain practices in order to stay on course.

We often enact self-control because we wish to pursue a larger later goal. But that our value hierarchies often give us *reason to* self-control does not require that we include them in its definition. I want to maintain that the behaviours like the ones Joanna enacts to hold her tongue or the ones Andy enacts to do a chore count as self-control, even if no clear value hierarchies are pertinent to their cases. Ambivalence[26] and indifference are motivational landscapes within which self-control may be more difficult, but there is no compelling reason to define our concept of self-control so as to make self-control in those motivational landscapes impossible.

We sometimes act under ambivalence, and sometimes, to act under ambivalence requires self-control. Indeed, there is reason to believe that ambivalent cases are of *particular* importance for self-control. This is because after a choice has been made that the agent is ambivalent about, that agent is often half-hearted about what they have chosen. In half-hearted action, agentive efficacy,

[25] This is to be taken ostensively, not exhaustively: I do not intend to claim that everyone asks this specific set questions, just that these sorts of things fall within what we typically (philosophers, laypeople) are most interested in.
[26] Ambivalence is not merely a feature of moral dilemmas like Joanna's, but it can also occur in prudential choices: Isaac Levi's example is that of an office manager selecting the most capable secretary out of three candidates, each of which has a different skill profile. Levi maintains that the manager need not choose which skill he values most to make the selection (1986: 11–13). I will not expand on the case here, as my only aim in introducing it is to show that ambivalence also occurs in the prudential domain.

i.e., the capacity of agents to act according to their intention, is at particular risk. As Shepherd (2017) argues, half-heartedness is often expressed in diminished control: the agent, being less motivated to do what they intend to do, has more trouble acting accordingly.

Self-control is difficult enough in paradigmatic cases, where an agent, intending to do what is clearly most valuable, nevertheless has a hankering for that which she values less. But when one has settled on a course of action without having judged that course of action to be best, that struggle may be much more challenging. Not only does the agent, having formed an intention, face a competing motivation, that competing motivation is not less preferable to her than her intention. Yet in order to exhibit agentive efficacy, she must align her behaviour to what she intends and not to the competing motivation. This is harder because first, the agent may be less committed to her choice; and second, some self-control strategies, such as rehearsing one's reasons to prefer the intended course over the competing one, are unavailable (because she has no compelling reasons to rehearse). In situations where the agent is ambivalent about which course of action is best, whichever option she settles on, self-control is that much more important for efficacy – and that much harder to enact[27].

The value-deflationist approach to self-control equally captures non-hierarchical, errant and value- hierarchical cases of self-control. For the value-deflationist approach to self-control, there is no paradox in errant self-control: self-control is simply the coaxing of one's behaviour to align with any intention, regardless of whether that intention reflects the agent's values or better judgment in any way. That, of course, does not mean that the relevant intention, or the resulting behaviour, would not be 'errant'. Mele's wayward youth's intention to join the break-in is errant, though it's plausibly an intention born less out of choice and more out of peer pressure. But that the youth's *intentions* are errant ought to be irrelevant for whether his behaviour counts as self-controlled. Both Mele's wayward youth, and agents pursuing more palatable aims such as dieting, overcoming phobias, giving up indulgences that are harmful to the environment, or finishing manuscripts, are engaging in self-control if they are aligning their behaviour with their intentions rather than with competing motives.

[27] Relatedly, Richard Holton, in describing weakness of will, has argued that weakness of will should be understood as independent of value judgment; see Holton (2006). There are some similarities between Holton's account and the one I offer here. However, the account proposed here differs from Holton's in several ways. First, Holton's account is an account about weakness of will, a concept that is related to but distinct from that of self-control. It is unclear whether Holton would hold the same claims about self-control as he does about weakness of will. Second, in Holton's account, weakness of will is the over-ready abandonment of a resolution, but for his account motivational conflict is not a necessary feature for strength of will, as it is for self-control in the present account. Furthermore, the present account does not concern itself with any conception of will or willpower, and is therefore compatible with both Holton's account and with skepticism about willpower, such as is expressed by, e.g., Ryle (1949: 72–73) and Levy (2011).

There are two main lines of objection worth responding to, here. First, one may insist that even in cases like Andy's and Joanna's, the agent in fact does prefer the course of action that they promote with self-control. It's just that the relevant motivational conflict for assessing self-control is not the conflict between dishes and laundry, or the conflict between yelling and holding one's tongue, but rather, e.g., the conflict between continuing to deliberate and doing a chore; or even the conflict between not making a difficult decision and making one[28]. In other words, we get a value-hierarchical case if we hold the conflict to occur between a different set of motives. This is exactly the solution that Mele proposes for the wayward youth: for Mele's case, the youth does promote what he thinks best – it's just that the relevant conflict is not the conflict between breaking into a house and not doing so, but rather, e.g., the conflict between sticking to one's resolution and 'chickening out'.

Many such descriptions are highly plausible, yet I have two worries about this approach. First, while it is true that any decision outside the laboratory yields itself to multiple descriptions, it is not entirely clear why we should privilege the description that gives us the received sort of evaluative structure. That this option should be given explanatory priority would need a further justification. Second, I worry about the results of this approach for conceptual utility. If the relevant conflict for self-control is often not the one that agents perceive it as being, then agents do not have access to information about whether they have enacted self-control, or whether they have succeeded or failed in it. Empirical research on self-control success and failure, too, would become contingent on researchers nailing the correct description. The value-deflationist definition avoids these worries as for it, we do not need accurate information about the evaluative hierarchy of the competing courses of action in order to, e.g., assess whether self-control has occurred.

Secondly, it may be objected that value-deflationist as it is, my approach to self-control throws the baby out with the bathwater. For many, it is precisely the pursuit of larger later goals or better judgment that makes self-control philosophically and psychologically interesting. I am sympathetic to these views; however, those interested in the value-hierarchical uses of self-control need not dismiss the value-deflationist conception. That self-control is best understood as devoid of value hierarchy does not mean we should not be interested in its use *for* intentions that match what we find valuable. Indeed, during an earlier stage in this manuscript, the manuscript stated that granting the value-deflationist conception of self-control, we may be justified in directing most of our attention to conventional cases: we greatly care about those cases, and experience deep regret when, due to a lapse in self-control, we have forfeited something of great value and gained only

[28] I owe this point to the external examiners of this dissertation.

little. Some such failures are frustrating, such those of Descartes upon procrastinating. Others are tragic, such as when life and well-being are on the line. But now I am not willing to make that claim about the relative importance of conventional cases. Cases of self-control, or the failure thereof, that involve ambiguity are also a source of deep concern for us. We care not just about the choice we make and our subsequent behaviour, but also about the alternatives we had and forfeited. Some of these choices may continue haunt us years later, and as they do, we may ask: what if I had chosen differently? And if, having resolved one way, our self-control has failed and we have ended up acting to the contrary, we may ask: what if I had managed to do as I resolved to do[29]?

The value-deflationist approach is preferable not just because errant and non-hierarchical cases cause it no difficulty, but also because the simplicity and transparency it lends to those cases makes it preferable for the work we do with the concept of self-control, both in the domain of communication and in the domain of agentive self-understanding.

Within the domain of communication, the value-deflationist approach is useful in communication between professionals and clients, such as within the clinical context. When a patient bemoans a lack of self-control to her psychiatrist, her problems on the level of intentionality (e.g., making poor life choices) can, through dialogue and other clinical instruments, be distinguished from self-control problems, on the level of efficacy. This distinction may be clinically useful. Likewise, for interdisciplinary and lay communication: in each of these domains, the value-deflationist conception allows for distinguishing between intention formation and efficacy. Within the domain of agentive self-understanding, this conception of self-control is directly utilizable by agents in assessing their self-control, including in cases where the agent enacts self-control to follow a poor life choice or a choice made under ambiguity.

The proponent of revealed preference theory would reject the above cases of Joanna and Andy but could embrace my account for reasons of avoiding redundancy. Namely, it is unclear why a value hierarchy ought to be included in our *conception* of self-control. This is for two reasons, the first of which is triviality. If the strong axiom of revealed preferences is true and all actions involve acting according to a value hierarchy, whether we perceive them as such or not, while this feature of agency can help explain self-control failures (as Ainslie (2001) has done), this feature does not help distinguish successful self-control from other behaviour. It is then the guidance of behaviour according to intention in the face of competing motivation, rather than the ubiquitous value

[29] I do not mean to imply that we would not also ask these questions when self-control was *not* pertinent to the choices we made. We can, and do, wonder about our past choices. I merely intend to point out that we also wonder about such matters concerning cases where self-control *is* pertinent, and when we do, then our conception of self-control becomes relevant for ascriptions of self-blame, for self-understanding, and so forth.

hierarchy, that characterizes self-control *specifically* rather than human agency *generally*. In other words, for revealed preference theory, valuation is not particular to self-control, and therefore the proponent of revealed preferences may consistently adopt my definition of self-control even as she holds revealed preference theory to be true. For the proponent of revealed preferences, even as she will not be swayed by talk of ambivalence, what I propose has the dual advantages of avoiding redundancy and Ockham's razor going for it.

4.2. What kind of a kind is value-deflationist self-control?

One way to explore the differences between various concepts and how they relate to the world is in terms of kinds. Kinds are categories, distinguishable from other categories by the manner in which these categories group entities into them. A notion originating in Aristotle is that some kinds are natural kinds: they 'carve nature at its joints', grouping entities together in ways that follow our observations of nature rather than in ways that respond to human interests in demarcation. 'Gold' and 'mammal' are received examples of natural kinds, whereas 'president', 'banknote', and 'virtue' are not. Within philosophy of biology and philosophy of medicine, the metaphysics of natural kinds have been increasingly problematized: the complexity of biological processes makes some received natural kinds subject to debate (see, e.g., Beebee & Sabbarton-Leary (eds.), 2010). The debate about natural kinds is nowhere as vibrant as within philosophy of psychiatry, where whether we believe that a sustained inquiry into the causal patterns underlying mental disorders would help us discover disorders that are natural kinds, or whether we instead conclude that mental disorder categories ought to be defined in terms of what best helps accomplish the practical work of psychiatry (see, e.g., Varelius 2009, Zachar 2014, Kendler, Zachar & Craver 2011, Borsboom, Cramer & Kalis 2019a), is a pressing issue. As Haslam (2002) argues, it is unnecessary to presume that all psychiatric disorders would be the same kinds of kinds: some may be natural kinds even if others – or most – are practical kinds. In this section, I apply arguments from within the conceptual and metaphysical debates in philosophy of psychiatry to the concept of self-control.

Haslam (2002) offers a helpful taxonomy to clarify what sorts of kinds psychiatric categories capture. He divides psychiatric categories into non-kinds, practical kinds, fuzzy kinds, discrete kinds and natural kinds. Non-kinds are categories that represent something without clear boundaries. Haslam offers continua as an example; in the context of conceptions of self-control,

Herdova's view that self-control is a gerrymandered construct could also be phrased that self-control is a non-kind.

Practical kinds are categories that are defined on the basis of the practical work to be done with these concepts: for example, disorder concepts are used, e.g., to provide treatment, assign treatment priority, and negotiate accommodations at school or workplace. For some disorders where symptoms are seemingly on a continuum, there may be such pragmatic considerations that warrant drawing the boundaries of the diagnostic category in a certain way. The boundaries of a diagnostic category, such as Major Depressive Disorder, do not then rest at an entirely arbitrary place on the continuum, but rather fall where the behaviours associated with the disorder reach a degree of intensity that warrants medical attention. (Haslam 2002.)

Haslam argues that two further categories of kinds fall in between natural and practical kinds: fuzzy kinds and discrete kinds. Fuzzy kinds have "real, discoverable discontinuities" (ibid.: 208), that is, they are categories that characterize multidimensional phenomena, where the dimensions may be only moderately correlated. In psychiatry, fuzzy kinds and practical kinds are alike in that the boundary of the category is on a gradient and subject to some negotiation. Out of the models of self-control discussed above, Buckholtz's multidimensional account of self-control could be analysed as treating self-control as a fuzzy kind. Discrete kinds, by contrast, have crisp boundaries but may not nevertheless be natural kinds: membership in the category may not be up to a single property, as in an Aristotelian view of natural kinds as essences. Many of the accounts of self-control described in the previous sections are akin to describing self-control as a discrete kind: they posit that there may be heterogeneity in which biological processes help realize self-control, but the boundary between self-control and other phenomena that are not self-control is nevertheless crisp and clearly identifiable rather than a subject of practical negotiation. Moreover, it is identified in terms of the discoverable structures of the world – such as, in terms of a group of cognitive and neural processes that may be identified in the future – rather than in terms of the practical use of the concept.

As we saw above and as Herdova (2017) has demonstrated, the concept of self-control does not refer to a natural kind. But this need not mean it would be a non-kind. We have three further options: fuzzy, discrete, and practical kind.

Could it be a fuzzy or discrete kind? Typically, kinds (other than practical kind) are distinguished based on observable phenomena such as processes and mechanisms, rather than based on the behavioural function of a kind; on that basis, self-control would not be a candidate for a fuzzy or discrete kind (based on the generality and multiplicity of these processes demonstrated above). There does not seem to be any discoverable, objective feature of the world – such as a

process or function – that would, alone or together with other such features, converge into something identifiable as self-control. But if we entertain the idea that a category could be discrete based on a discrete behavioural function, then my value-deflationist conception of self-control would seem to be a candidate for a discrete kind. After all, it seems, at first sight, to be clearly demarcated: something either is that with which behaviour is aligned with intention in the face of competing motivation, or not. Sticking with this thought for a while, however, the discrete kind classification turns out not to be accurate. An agent may do something that is both a means of diffusing an oncoming motivation and a means to some other end. Prayer strikes me as an example: religious persons facing self-control troubles may use prayer both as a self-control strategy and to express and cultivate their spirituality. An instance of self-control is thus also an instance of something else. It strikes me as plausible that instances where a behaviour serves both the function of self-control and another function, where some such instances of self-control may be "self-control heavy" with the other functions as secondary, and for other instances, self-control may be a secondary function. Which of these we should call 'self-control', and under what circumstances we should call them, depends, in my view, on the work done with the concept – including the work of agentive self-understanding. However, this does not make participation in the category of self-control arbitrary: if I am correct, then there are requirements for something to be called self-control, namely, the presence of an intention, i.e., a practical commitment, as well as current or anticipated contrary sources of motivation, and that the behaviour is used to guide the agent to stick to the former.

Herdova's criticism also appears to foreclose the option of defining self-control as a practical kind: such practical kinds that are defined as sets of discoverable traits, mechanisms and processes, but the boundaries of the category are set on the basis of our practical needs, would also face some of the same issues. Contrast, here, self-control with Major Depressive Disorder. In MDD, certain symptoms, such as flat affect and distress are present to varying degrees. These are also present in some of the general population. Identifying the degree to which they must be present for the symptom presentation to qualify as MDD is a matter of some debate. However, these symptoms are not *general*: they do not pertain to all agents at all times, nor are they a general feature of agents of another category (say, rational agents, male agents, or oppressed agents). All agents may experience these at times, but their persistence constitutes a meaningful differentiating factor between depressed and healthy agents, even if the threshold of clinical interest is subject to negotiation. The mechanisms of delayed gratification, proposed to be underlying self-control, however, are general to *all rational agency*. The same mechanisms are utilized also in other processes than self-control. Nor are degrees of intensity or frequency, so useful in the case of spectrum disorders, helpful in the

case of self-control: attention guidance and inhibitory control are highly active, for example, when cramming for an exam – even if no conflicting impulse is present; and highly inactive also when failures of self-control are not imminent, such as, during sleep. Researching how it is that these mechanisms and processes contribute to self-control is of unquestionable value. But as I suggested in section 4.1, they are unhelpful for the purposes of *defining* self-control.

However, approaching the problem bottom-up, and defining self-control via its function for behaviour, seems to do the trick. The utility of the concept of self-control for the practical work of agentive self-understanding, communication, research, and clinical practice rests on the clarity of the behavioural function it refers to. If the concept of self-control would need to refer to processes and mechanisms in order to have conceptual utility, then Herdova's claims would likely be correct and self-control would be a non-kind. However, the concept of self-control defended in the present dissertation is not subject to that criticism.

As I have suggested above, other concepts such as temporal discounting and inhibitory control work well for the purposes of researching those processes. The work of the concept of self-control is the work of communicating about behavioural functions that those specific processes sometimes contribute to. The 'hodgepodge' pool of behaviours is united by their helpfulness for all these purposes, even as the uniting feature – as the 'prayer' example shows – is not entirely crisp across human behaviour. I find this gives us strong grounds for considering self-control to be a practical kind.

5. The access theory of self-control

Above, I have worked out my preferred way to define self-control, according to which self-control is any behaviour that is enacted to align one's subsequent behaviour with intention in the face of competing motivation. In this section, I turn my attention to another question, namely, what explains interpersonal differences in self-control? To do so, I draw from advances in philosophy of disability.

The previous sections have laid out the context and groundwork for my ideas concerning individual differences in self-control. This section, by contrast, is much denser and presents arguments in summary alone. This is because the full argumentation for the access theory of self-control that I defend can be found in article 3. A full discussion of its implications for responsibility and ADHD can be found in article 2. Discussion concerning self-control neuroenhancement, especially using stimulant medication, can be found in articles 1 and 4. This section, then, will only summarize the main thrust of those arguments and explicate how the four articles relate to each other rather than repeat their discussion in full.

5.1. Self-control and access

An insight within disability studies and philosophy of disability is that many of the differences in bodily function associated in disabilities are not, in and of themselves, disabling. Rather, the restrictions and difficulties for conducting daily activities and pursuing life plans stem to a considerable extent from physical barriers and social stigma. The relationship between disabilities and their constitutive differences has been much debated: for example, the social model of disability (Oliver 1996) distinguishes between disabilities, which result from the social and built environment, and impairments (the biological foundation that stigma and barriers attach to). This model is not without critics: as Tremain (2015) argues however, physiological differences do not always themselves impair; Barnes (2016) and Shakespeare (2008) point out that presocial bodily phenomena, such as pain, also disable. Yet the insights that the social and built environment is the source of much of the constraint in disability, and that dismantling these barriers rather than

attempting to ‘fix’ others’ bodies to fit the preferences of the social and built environment may be the better way to go, are the sustained legacy of the social model.

However, even with this focus on the social and built environment in the study of disability, there has been no systematic philosophical analysis of the concept of access. I take tentative steps towards this project in this dissertation, asking what is required for an agent to have access to something.

As the articles 2 and 3 demonstrate, during this tentative work my views have evolved. In article 2, I sketch the criteria as follows: “For a self-control behaviour to be accessible for an individual, requires, at minimum, that the behaviour is feasible and that the individual is aware of the availability and efficacy of this behaviour”. In article 3, I both clarify and formalize these criteria and add a third one: that of non-excessive effort. In brief, I discuss access to x as requiring that the agent has some practical means by which she can x; that she is aware of the availability and efficacy of these means; and that these means do not require excessive effort for the agent to enact. See article 3 for a full discussion of these criteria.

Variance in the extent to which these three criteria are fulfilled, for my account, explains both interpersonal and diachronic variance in self-control. As self-control simply is the various behaviours we label self-control (under either my value-deflationist conception or any conception based on defining self-control on the basis of its function), individual and diachronic differences in self-control amount to access to those behaviours.

The criteria of means, awareness and non-excessive effort are satisfied to different extents in different persons; and the distribution of means and awareness is not random but rather, tracks various axes of disadvantage. Diachronic variance is explained by the fact that some means may be available at certain times but not at others, may be more effortful at certain times (say, after a sleepless night) than at others, and agents may be aware of them at certain times but not at others.

For persons with difficulties in executive functioning including people with ADHD, self-control strategies that rely on inhibitory control and attentional shifting may be either not such a feasible means or may be excessively effortful. It then becomes of paramount importance that these agents are aware of alternative means of self-control that are better suited for them. By contrast, many social strategies are only available in the presence of supportive social relationships – Ulysses would not have stood a chance had his crew at been mutinous when they passed the Sirens. And the availability of many paradigmatic situation selection strategies tracks socioeconomic status, as many such strategies require some expenditure that, even if a bargain for the middle-class academic, may be prohibitive for someone living in poverty.

For those who experience disadvantage on multiple axes, the pool of feasible means to self-control grows small but not non-existent. However, whether these agents are aware of this set of self-control strategies is another story. Particularly, the discourse on self-control that highlights mental effort and "willpower", obscures that other strategies may be available.

The third criterion of excessive effort is admittedly, and perhaps necessarily, fuzzy. There are clear cases of excessive effort, such as when the effort required to do something is so large as to make this pursuit self-defeating. And there are clear cases of non-excessive effort. I hope to do further work on excessive effort in the future – but for my purposes here I think it sufficient to note that access can be foreclosed by the excess of effort required.

As a result, out of the pool of all possible self-control strategies and processes, agents only have access to feasible means that they are aware of and that do not require excessive effort. Barriers to feasibility, barriers to awareness, and barriers to non-excessive effort prevent the successful use of some strategies and processes of self-control. For those who are disadvantaged on multiple axes, this can result in that there is no such self-control strategy or process that the person would have access to. However, this is not, in such cases, a personal flaw, but rather is something that is caused by, and could be ameliorated by, the norms, structures and practices of our society. By altering those norms, structures, and practices, barriers to self-control can be lifted. In this dissertation, space does not allow me to do justice to the effects of specific axes of disadvantage implicated. However, there is a need to expand on, in future work, the entangled effects that intergenerational poverty, disability, crime, and mental illness, as well as other social determinants, have on access to self-control.

One upshot of thinking of self-control in terms of access is that it undermines the tendency to think of self-control successes and failures as personal successes and failures. Self-control success and failure, for the access theory, are not evidence of character but rather, of the use of resources both intramental and environmental. Sometimes, self-control success results from the clever, persistent, or resourceful use of such resources, and thus we are not necessarily wrong to praise it. But a similar degree of cleverness, persistence, and resourcefulness may not result in self-control success under different socioeconomic and (neuro)physiological conditions.

Accepting the access theory, then, does not require us to cease blaming and praising each other for our successes and failures in self-control. Instead, it urges that if we do so, we should blame and praise with an eye on the situational character of self-control success – i.e., on how our

circumstances, including wealth, the built environment, and the people who help and hinder us, have helped or hindered our success[30].

The access theory suggests that those who are fairly consistently successful in self-control have a robust 'toolkit' of self-control behaviours that are well suited for them and that are applicable to a variety of contexts. Given the heterogeneity of behaviours that can be employed as means to self-control, in theory, *in the absence of barriers*, almost any agent, including agents with cognitive and learning disabilities, could possess a broad set of competences in self-control behaviours that are a good fit for them. However, given barriers in the social environment that make many such behaviours unfeasible and therefore unsuitable as means to self-control; epistemic barriers that prevent many agents from learning or being aware of suitable means to self-control; and the excessive effortfulness of some means of self-control, contingent features of the social and built environment result in aggravated individual differences in self-control.

5.2. When access is limited: the case of ADHD

The above discussion summarized my account of the access theory of self-control, gestured towards in article 2 and described in full in article 3. Here, I consider the implications of the access theory of self-control for one case where self-control is often limited. As mentioned above, many neurodevelopmental and mental health conditions modulate what sort of self-control behaviours are available for an individual. Attention-Deficit / Hyperactivity Disorder (ADHD) is only one such condition, but I have chosen it as an example for two reasons: first, it is a fairly common disorder, and second, its symptom presentation, which spans self-control-associated difficulties across various domains from the occupational to the interpersonal, is sufficiently similar to nonclinical self-control problems for it to serve as an illustrative example (by contrast, addictive behaviour, often used as an example in philosophical examinations of self-control, is typically described as involving self-control problems in relation to the addictive behaviour only). Second, ADHD is associated with deficits on all measures of self-control, from inhibitory control and delayed gratification to trait self-control, and its symptomology corresponds to difficulties in self-control on any theory of self-control (Barkley 1997, 2015).

As mentioned above, ADHD can make forms of self-control that heavily rely on the executive functions, such as brute inhibition, disproportionately effortful or even impossible. This,

[30] However, this allows that not all self-control failures are due to lack of access. Some failures may be simply blunders.

however, does not mean people with ADHD could not reliably succeed in self-control. The access theory is well placed to explain both why it is that people with ADHD usually are less self-controlled than others, and why it is that some people with ADHD are very self-controlled, which helps account for high achievers with ADHD.

There are many self-control behaviours that ADHD does not impact, or on which its impact is negligible: these include many environmental strategies, as well as such intramental strategies like construal that are more reliant on imagination than executive function. The question then is whether an individual with ADHD is well placed to enact such feasible self-control behaviours. For example, SES and the presence of social support modulate the feasibility to many environmental means to self-control.

A major barrier to self-control for persons with ADHD, however, is awareness. In particular, lay and professional discourse of self-control as "willpower" can obscure the multiple realizability of self-control and prevent people with ADHD from becoming aware of self-control means that are well suited for them.

These matters are made that much more urgent by the interconnected nature of socioeconomic disadvantage and ADHD (Miller et al. 2016). Due to the adverse impact of ADHD symptoms on education and employment, adults with ADHD earn less than peers. Due to the familiality of both ADHD and SES, this results in many persons with ADHD being also low-income. Low-SES people with ADHD are doubly disadvantaged with regard to self-control: it is harder for them to enact intramental self-control behaviours, as well as such environmental self-control means that require expenses. As a result, these people face multiple barriers to means to self-control. Low-SES persons with ADHD are also at a higher risk of many of the social risks associated with ADHD, such as substance use problems and problems with the law. All of this highlights the need for dismantling epistemic and other barriers to self-control for such people, in order to enable them to discover and enact suitable means of self-control. Steps towards this can be taken, e.g., by emphasizing the heterogeneity of effective self-control practices and by lifting practical barriers to situational strategies at schools and workplaces. If the access theory is true, then individual differences in self-control are not a matter of inherent character alone: they are a matter of justice. A fuller discussion of the above characterization of the double binds resulting from SES and ADHD can be found in article 3.

The self-control struggles connected with ADHD also have given rise to puzzles concerning moral responsibility. This is the topic of article 2 of this dissertation. People with ADHD sometimes act in morally and prudentially unacceptable ways from forgetfulness to speaking out of turn, procrastinating, or mismanaging their money. The question, then, is whether adults with ADHD are

fully responsible for this behaviour or whether it ought to be excused on the basis of their ADHD. This is not merely a question for armchair philosophers, but rather, one that persons with ADHD themselves struggle with: it can be difficult to take a clear stance on whether one is responsible for some action despite ADHD, or whether one should be excused (see, e.g., Honkasilta, Vehmas & Vehkakoski 2016).

In article 2, I divide appeals to responsibility mitigation into two groups: agential pleas and situational pleas. For agential pleas, the agent is excused or responsibility is mitigated due to some persistent feature of the agent, such as their ADHD. For example, Arpaly (2003, 2006) argues that responsibility ought to be mitigated when an action is caused by a disorder such as ADHD. The trouble with agential pleas is that reliance on agents' features suggests that we ought to similarly mitigate responsibility for all of the agent's actions (or, at least, all on which that feature causally contributed). But this seems heavy-handed: persons with ADHD seemingly sometimes act in ways that ought to be excused, and sometimes in ways that ought not to be. A further trouble is that it is hard to estimate whether the global level of mitigation ought to be set to match the actions of least mitigation or the acts of most mitigation. In other words, if an agent is fully responsible for action A, then should we treat her as equally responsible for all her actions? (This is, incidentally, an inference often drawn in popular discourse from anecdotal evidence concerning high achievers with ADHD: that persons with ADHD are fully responsible and ought to "get a grip" of themselves.) Or, if an agent is excused for action A, then should we treat her as equally not responsible for all her actions? This seems false given that persons with ADHD can and do take responsibility for many of their actions.

Instead of agential pleas, in article 2, I recommend situational pleas, in which the agent is excused or not based on whether situational factors caused their acting in a certain way. This enables us to treat different actions of the same agent differently, if they have occurred in different circumstances. The access theory of self-control enables us to view self-control as sensitive to the agent's situation rather than as resulting from their character alone. As I discuss in article 2, he access theory and situational pleas, combined, allow us to see how situations that enable many agents to act in certain morally or prudentially desirable ways may yet bar persons with ADHD from doing so. In effect, when assessing whether we should hold an agent with ADHD responsible for her failure of self-control, we need to look at whether that agent in fact had access to self-control in that situation.

The discussion in article 2 is also a response to a debate concerning the role of genetic explanations for assessments of moral responsibility. Here, research into behaviour genetics is informative. Today, it is universally acknowledged that both genetic and environmental causes

factor into the causation of behaviour. It is also not meaningful to ask which part has the "larger" portion of a given causal pathway, since the effects of genetics and environment are not additive: many genetic effects are produced by gene-environment interactions, some of which are active, some passive. Against this background, however, genetic explanations of diagnoses of ADHD do have a well-attested, even if misguided, effect on responsibility assessments. My discussion aims to demonstrate that this is a red herring, and that *if* responsibility is tied to self-control, as many hold it is, then responsibility assessments should track access to self-control.

5.3. Notes on self-control neuroenhancement and stimulant medication

The means by which agents can attempt to improve their self-control are many. One such means, however, is medical. Currently, some agents use stimulant medication such as methylphenidate or modafinil to help improve their self-control, either because they have been diagnosed with ADHD and the medication has been prescribed to help alleviate their self-control-related symptoms or because they wish to improve their self-control by medical means despite not having ADHD. In this section, I first discuss the feasibility and plausibility of self-control neuroenhancement, and then outline some moral constraints to it.

Within the articles of this dissertation, neuroenhancement is discussed in articles 1 and 4. In article 1, I and co-authors discuss the prospects of self-control neuroenhancement for purposes of enhancing responsibility within the criminal justice system, noting difficulties that stem from the use of neuroenhancement in that very particular context. Article 4 discusses the use of stimulant medication for purposes of cognitive enhancement. Note that while article 4 is framed in terms of cognitive enhancement, the cognitive processes that stimulant medication purports to enhance include such processes that contribute to self-control, such as executive functioning processes. Indeed, narrower targeting of self-control, per se, without at all affecting other functions, is not feasible.

Due to the various contexts in which self-control can be enacted, its enhancement can be desired for a variety of reasons and in a variety of contexts. Self-control can be desired, e.g., because it is seen as pivotal for academic and professional success, or for successfully changing one's lifestyle and habits. For those who view moral and criminal responsibility as hinging, in part, on the dual conditions of control and understanding (such as Vincent 2010, 2014 and Shaw 2014, 2016), self-control can be desired simply for its role in enabling responsibility. And it seems clearly true that many crimes would have been prevented if the offenders' self-control had not failed.

Indeed, a prominent theory in criminology, dubbed the General Theory of Crime (Gottfredson & Hirschi 1990), suggests all criminality is caused by individual differences in self-control. Self-control interventions have significant potential for criminal rehabilitation given that, even if the general theory of crime is not correct, it seems that many crimes are due to failures of self-control.

Stimulant medication is the most typical form of self-control neuroenhancement, but other means, such as transcranial stimulation, have also been suggested. Optimism towards these forms of self-control neuroenhancement has been garnered, e.g., from a study indicating that different forms of self-control involve activity in the same brain region, namely the inferior frontal cortex (see, e.g., Tabibnia et al. 2011, Aron, Robbins and Poldrack 2014). These findings support treating inhibition across domains as connected, but do not account for processes such as environmental selection and manipulation. They suggest that targeted interventions stimulating the function of this area may help improve inhibitory control across various domains in the future.

For stimulant medication, credence in its efficacy for self-control neuroenhancement arises from its use in the management of ADHD and related conditions. The impact of stimulant medication on ADHD symptoms is small but clinically significant (Storebø et al. 2015). Some studies indicated that stimulant medication improved cognitive (especially executive) function also in persons without ADHD, giving rise to early optimism about the prospects of neuropharmacologically improving self-control. However, the effects of stimulant medication on healthy subjects were found to be minor to none in meta-analyses (see, e.g., Repantis et al. 2010). For studies and anecdotal evidence reporting significant effects, the mechanism may be that the improved performance is caused by improved mood, confidence, and motivation resulting from misplaced optimism about the efficacy of the stimulant (Petersen, Enghoff, & Demant 2019). It can safely be concluded that the effect of current medical means on any means of self-control is small.

Given the non-dedicated nature of any process subserving self-control, it is important to acknowledge that any current or future pharmaceutical that is effective in modulating such processes will also have effects beyond self-control. Whether these effects are all things considered desirable depends, to a great extent, on the agent's specific circumstances, health, and preferences. In article 4, I discuss this issue at length, arguing that user decisions about stimulant medication are very difficult to make in an informed manner.

Do these arguments concerning pharmaceutical cognitive enhancement extend to all means of cognitive enhancement? To an extent, they do. For example, administering a series of sessions of transcranial direct current stimulation (tDCS), in a clinical setting, particularly in conjunction with cognitive training tasks, can have a long-term beneficial impact on cognitive function including for healthy participants; it is well tolerated, with itchiness of the scalp at the location where the anode

was applied as the most common side effect. However, these benefits vary with individual differences in a manner that is not currently comprehensively understood (Katz et al. 2019). Concerns have been raised about the safety of self-administered tDCS, including the concern that the intervention may alter more of the brain and its functioning than just the targeted area and processes, alterations that may be difficult to identify in the absence of direct clinical supervision (Wurzman et al. 2016). It should also be borne in mind that even for procedures that are safe, given individual differences in their efficacy and in the life circumstances of each user, the cost (financial, effort, time, etc.) of acquiring access to the intervention may outweigh its benefits for some users, which is an issue for those with little resources to spare. However, the relevant ethical considerations may not be fully analogous with those pertaining to stimulant medication, which is something that further research on this topic could help clarify.

A further feasibility worry is raised in article 1, which addresses the topic of self-control enhancement as a 'neurocorrective', i.e. its use as a measure of rehabilitation within the criminal justice system. While the access theory supports the claim that many of the persons who end up within the criminal justice system may be less successful in self-control (whether due to difficulties in inhibitory control or due to barriers to situational strategies of self-control), it urges caution about undue optimism that improved inhibition alone – if feasible – would result in improved self-control. Many persons within the criminal justice system are subject to marked and systemic disadvantage, as a result of which their access to self-control is limited. Stimulating the function of any neural system alone does not generate access to self-control: in order to improve self-control, all three criteria of means, awareness and non-excessive effort need to be accounted for. In article 1, I and co-authors argue for one practical upshot of the access theory (and of any externalist theory of self-control), namely that given the limited access to situational self-control strategies inmates have, any neuroenhancement measures that would work on the well-to-do may fail to similarly promote self-control in inmates and in subjects of the criminal justice system living in poverty. For a fuller discussion of these issues, see article 1, where I also discuss the context of responsibility at more length.

The reader of the two articles can discern that in article 1, my outlook on the prospects of stimulant medication as a cognitive enhancer is more optimistic than it is in article 4. One reason behind an increasing degree of caution is acknowledging the epistemic difficulties introduced by the phenomenon of folk pharmacology, i.e., that socially and culturally transmitted ideas concerning pharmaceuticals impact how users interpret the response to that pharmaceutical. As a result of folk pharmacology, both users and non-users overestimate the impact of stimulant medication. This is a point that ought to be emphasized: having personal experience with a pharmaceutical does not yield

accurate insight into that pharmaceutical's impact on one's functioning because the sociocultural associations of that pharmaceutical shape user experience in significant ways (Petersen, Enghoff, & Demant 2019). Folk pharmacology is not the same issue as the placebo effect: the placebo effect is a real effect, but due to folk pharmacology and other epistemic difficulties, users can be poorly placed to assess the overall effects (including placebo or nocebo effects) of the intervention they are undertaking.

The ethical upshots of these feasibility concerns, as argued for in articles 1 and 4, can be summarized as follows. First, these concerns suggest that the efficacy of self-control neuroenhancement – even more effective future neuroenhancement – is limited if not paired with identifying and tackling epistemic and environmental barriers to self-control. Neuroenhancement should therefore only be used in conjunction with promoting awareness of environmental strategies of self-control. Second, they suggest that the liberal approach to self-control neuroenhancement, according to which there is no reason to prohibit patients from enhancing given that the patient knows best whether they benefit, is false as the relevant cost-benefit analyses require nuanced, personalized information to conduct. Third, due to the level of nuance and individualization required, even if self-control enhancement should be prescribed as part of a clinical relationship, it should never be court-mandated: that system is not equipped to tackle the various considerations required for an individualized assessment of the effectiveness and desirability of the intervention. Further discussion on these issues can be found in articles 1 and 4.

6. Conclusion

In the introduction, I laid out the three research questions that have guided my philosophical explorations throughout the process of this dissertation.

1) What is self-control?
2) What explains individual differences in self-control?
3) What can we learn about self-control by looking at the case of ADHD?

The heterogeneity within the categories under scrutiny here – both the category of self-control and the category of ADHD – has made this investigation both more complicated than may have been projected, involving detours into topics like biological causation – and, I believe, that much more meaningful. Even as there is confusion about what, exactly, self-control entails, we depend increasingly on it: more and more of work, study, and leisure is self-guided, and success in our various life projects thus hinges on access to means of self-control that we can enact.

In sections 2 and 3, I have reviewed the existing literature concerning self-control, motivating why I believe there to be room for development in theorizing self-control. Having endorsed the externalist conception of self-control, I suggest that while it correctly describes self-control as involving wide scope of behaviours of environmental interaction, the definition of self-control suggested therein should be rejected and replaced with another definition that gives the externalist what they need.

I have argued that behaviours need not relate to value hierarchies in any specific way in order to count as self-controlled: self-control can occur in the service of that which is less valuable (by whichever means of ranking one prefers), and it can also occur when an agent has formed her intention without having judged which course of action is most valuable. I thus defend a new value-deflationist definition of self-control in section 4. Value hierarchies are not a defining part of self-control; instead, values figure in our *attitudes towards* self-control.

Another way to frame this argument is as a claim about the proper location of values for self-control. While my approach involves removing motivational value hierarchy from the definition of self-control, it does not make self-control a topic devoid of value. Rather, even as values are not part of an 'essence' of self-control, success and failure in self-control occurs in value-rich social contexts. We value self-control instrumentally and sometimes, intrinsically.

I have explored individual differences in self-control in articles 2 and 3 and in section 5, paying particular attention to ADHD as a case of clinically impaired self-control. In my analysis, I ended up claiming that individual differences in self-control are a result of interaction between the agent and their environment, and that they hinge on access to self-control behaviours. This does not mean that individual neurobiology would be meaningless: as explained in articles 2 and 1, neural functioning remains relevant for many forms of self-control. However, understanding how physical and socioenvironmental barriers, including epistemic barriers, preclude self-control success in persons with difficulties in executive functioning is instrumental for helping these agents access self-control behaviours that are a better fit for their individual biology.

If the access theory of self-control is correct, then individual differences in self-control are produced, in large part, by societal practices that create barriers to self-control for disadvantaged groups, and are therefore subject to societal intervention. As a result, self-control is not exhausted by individual character, but is subject to social distribution and arguably, distributive justice.

When we succeed or fail in self-control in significant ways, the question of whether and how we ought to hold each other responsible for these successes and failures arises. In article 2, I have argued that the deeply environment-sensitive character of self-control reflects on how we treat the sort of plea often leveraged in cases of moral shortcoming, namely that the agent could not control herself: my suggestion is that if we wish to verify whether the agent could have in fact controlled herself we need to look into her circumstances and investigate whether she had access to self-control behaviours. In other words, situational pleas are more appropriate than agentive pleas are for self-control failures.

I have also investigated the prospects of self-control neuroenhancement. In my analysis, the effects of existing means of neuroenhancement on self-control are small, and the presence of their other effects, some of which are harmful for some users, suggest that cost-benefit analyses ought to be made in a highly individualized manner. In any case, if stimulant medication can enhance self-control, my approach predicts that it can only successfully do so in contexts where the subject has access to self-control practices. Sometimes, self-control neuroenhancement is suggested as a rehabilitation intervention for criminal offenders. I suggest that for this population, their lived environments often prevent access to self-control behaviours even if their neural functioning is successfully amended by medical means.

I do not presume to have had the final say on any of the questions at hand. But I hope that some of the questions here probed will help move philosophical as well as empirical investigation of self-control forward in a manner that helps us make self-control more accessible for all.

Bibliography[31]

Ainslie, G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th edition.* Arlington, VA: American Psychiatric Publishing.

Aristotle (2000). *Nicomachean Ethics*. Crisp, R. (ed. / transl.) Cambridge: Cambridge University Press.

Aron, A.R., Robbins, T.W., & Poldrack, R.A. (2014). Right inferior frontal cortex: addressing the rebuttals. *Frontiers in Human Neuroscience* 8:905. doi: 10.3389/fnhum.2014.00905

Arpaly, Nomy (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.

Arpaly, Nomy (2006). *Merit, Meaning, and Human Bondage: An Essay on Free Will.* Princeton: Princeton University Press.

Audi, R. (1993). *Action, Intention, and Reason.* Ithaca: Cornell University Press.

Barkley, R. A. (1997). *ADHD and the Nature of Self-Control.* New York: Guilford Press.

Barkley, R. A. (2014). *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment, Fourth Edition*. New York: Guilford Press.

Barnes, E. (2016). *The Minority Body: A Theory of Disability*. Oxford: Oxford University Press.

Baumeister, R. F. & Exline, J. J. (1999). Virtue, Personality, and Social Relations: Self-Control as the Moral Muscle. *Journal of Personality* 67 (6): 1165–1194

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science* 16(6): 351–355.

Beauchamp, K. G., Kahn, L. E., & Berkman, E. T. (2016). Does inhibitory control training transfer?: behavioral and neural effects on an untrained emotion regulation task. *Social cognitive and affective neuroscience*, *11*(9), 1374–1382. https://doi.org/10.1093/scan/nsw061

Beebee, H. & Sabbarton-Leary, N. (eds.) (2010). *The semantics and metaphysics of natural kinds*. New York: Routledge.

Bermudez, J. L. (2018). Frames, rationality, and self-control. In J. L. Bermudez (ed.), *Self-Control, Decision Theory, and Rationality: New Essays.* Cambridge: Cambridge University Press.

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

[31] This is a bibliography for this summary section of the dissertation. Bibliographies for each article can be found within those articles.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences* 3: 122–129.

Chang, H. (2019). Operationalism. In Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2019 Edition)*. Url: https://plato.stanford.edu/archives/win2019/entries/operationalism/

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis* 58 (1): 7–19.

Clarkson, J.J., Otto, A.S., Hirt, E.R., & Egan, P.M. (2016). The malleable efficacy of willpower theories. *Personality and Social Psychology Bulletin* 42 (11): 1490–1504.

Cohen, J. R., Berkman, E. T., & Lieberman, M. D. (2013). Intentional and incidental self-control in ventrolateral PFC. In Stuss, D. T. & Knight, R. T. (eds.), *Principles of Frontal Lobe Function: Second Edition.* Oxford: Oxford University Press.

Connor, T. D. (2014). Self-control, willpower and the problem of diminished motivation. *Philosophical studies* 168: 783–796.

Cortese, S. & Castellanos, F. X. (2015). Attention Deficit/Hyperactivity Disorder. In Zigmond, M. J., Rowland, L. P. & Coyle, J. T. (eds.), *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders.* Cambridge, MA: Academic Press.

Dar-Nimrod, I. & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin* 137 (5): 800-818.

Davenport, A. (2006). *Descartes's Theory of Action.* Leiden: Brill.

Davidson, D. (1969). How is weakness of the will possible? In *Essays on Actions and Events.* Oxford: Oxford University Press.

Duckworth, A. & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* 45 (3): 259–268.

Duckworth, A., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science* 11 (1). 35–55.

Duckworth, A., Milkman, K.L., & Laibson, D. (2019). Beyond willpower: Strategies for reducing failures of self-control. *Psychological Science in the Public Interest* 19 (3): 102–129.

Egré, P. & O'Madagain, C. (2019). Concept utility. *Journal of Philosophy* 116 (10): 525–554.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy* 68 (1): 5–20.

Friedman, N.P., Miyake, A., Young, S.E., DeFries, J.C., Corley, R.P., & Hewitt, J.K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General* 137(2): 201–225.

Friedman, N.P., Miyake, A., Altamirano, L.J., Corley, R.P., Young, S.E., Rhea, S.A., & Hewitt, J.K. (2016). Stability and change in executive function abilities from late adolescence to early

adulthood: a longitudinal twin study. *Developmental Psychology* 52 (2): 326–340.
Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Personality and Social Psychology Review* 15 (4): 352–366.
Fujita, K., Carnevale, J. J. & Trope, Y. (2018). Understanding self-control as a whole vs. part dynamic. *Neuroethics* 11 (3): 283–296.
Gallagher, S. (2018). The extended mind: State of the question. *Southern Journal of Philosophy* 56 (4): 421–447.
Gillebaart, M. (2018). The 'operational' definition of self-control. *Frontiers in Psychology* 9: 1231. doi: 10.3389/fpsyg.2018.01231
Gottfredson, M.R., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford: Stanford University Press.
Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science* 11 (4): 546–573.
Harris, K. (2020). Why the self does not extend. *Erkenntnis*. https://doi.org/10.1007/s10670-020-00320-6
Haslam, N. (2002). Kinds of kinds: A conceptual taxonomy of psychiatric categories. *Philosophy, Psychiatry & Psychology* 9 (3): 203–217.
Henden, E. (2008). What is self-control? *Philosophical Psychology* 21 (1): 69–90.
Herdova, M. (2017). Self-control and mechanisms of behavior: Why self-control is not a natural mental kind. *Philosophical Psychology* 30 (6): 731–762.
Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science* 4 (2): 162–176.
Holton, R. (2003). How is strength of will possible? In Stroud, S. & Tappolet, C. (eds.), *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press.
Honkasilta, J., Vehmas, J, & Vehkakoski, T. 2016. Self- pathologizing, self-condemning, self-liberating: Youths' accounts of their ADHD-related behavior. *Social Science & Medicine* 150: 248–255.
Horstkötter, D. (2015). Self-control and normativity: Theories in social psychology revisited. *Theory & Psychology* 25(1): 25–44.
Hsu, T., Tseng, L., Yu, J., Kuo W., Hung D.L., Tzeng O.J.L., Walsh, V., Muggleton, N.G., & Juan, C. (2011). Modulating inhibitory control with direct current stimulation of the superior medial frontal cortex. *NeuroImage* 56(4): 2249–2257.
Ilieva, I., Boland, J., & Farah, M.J. (2013). Objective and subjective cognitive enhancing effects of mixed amphetamine salts in healthy people. *Neuropharmacology* 64: 496–505.

Inzlicht, M., & Schmeichel, B.J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science* 7 (5): 450–463.
Inzlicht, M., & Friese, M. (2019). The past, present, and future of ego depletion. *Social Psychology* 50 (5–6): 370–378.
Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion — Is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science* 21: 1686-1693.
Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
Kalis, A. (2018). Self-control as a normative capacity. *Ratio* 31 (S1): 65–80.
Katz, B., Au, J., Buschkuehl, M., Abagis, T., Zabel, C., Jaeggi, S. M., & Jonides, J. (2017). Individual differences and long-term consequences of tDCS-augmented cognitive training. *Journal of Cognitive Neuroscience* 29 (9): 1498–1508.
Kendler, K. S., Zachar, P. & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine* 41: 1143–1150.
Kennett, J. & Smith, M. (1996). Frog and toad lose control. *Analysis* 56(2): 63–73.
Koch, U., & Cratsley, K. (2020). Psychological mechanisms. In V. Zeigler-Hill, T. K. Shackelford (eds.), *Encyclopedia of Personality and Individual Differences.* New York: Springer.
Koi, P. (forthcoming). Towards informed user decisions about pharmacological cognitive enhancement. *Cambridge Quarterly of Healthcare Ethics.*
Koi, P. (2020). Born which way? ADHD, situational self-control, and responsibility. *Neuroethics*, online first. https://doi.org/10.1007/s12152-020-09439-3
Koi, P., Uusitalo, S. & Tuominen, J. (2018). Self-control in responsibility enhancement and criminal rehabilitation. *Criminal Law and Philosophy* 12 (2): 227–244.
Kotabe, H. & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science* 10 (5): 618–638.
Levi, I. (1986). *Hard Choices: Decision Making under Unresolved Conflict.* Cambridge: Cambridge University Press.
Levy, N. (2007). *Neuroethics: Challenges for the 21st Century.* Cambridge: Cambridge University Press.
Levy, N. (2011). Resisting weakness of will. *Philosophy and Phenomenological Research* 82 (1): 134–155.
Levy, N. (2013) (ed.). *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*. Oxford: Oxford University Press.
Levy, N. (2016). The sweetness of surrender: Glucose enhances self-control by signaling environmental richness. *Philosophical Psychology* 29 (6): 813–825.

Levy, N. (2017). Of marshmallows and moderation. In Sinnott-Armstrong, W. & Miller, C. B. (eds.), *Moral Psychology Volume 5: Virtue and Character.* Cambridge, MA: MIT Press.
Levy, N. (2017b). Trading in the trait? Response to Masicampo and Sripada. In Sinnott-Armstrong, W. & Miller, C. B. (eds.), *Moral Psychology Volume 5: Virtue and Character.* Cambridge, MA: MIT Press.
MacKenzie, C. (2008). Relational autonomy, normative authority and perfectionism. *Journal of Social Philosophy* 39: 512–533.
Magen, E. & Gross, J. (2010). Getting our act together: Toward a general model of self-control. In
McGuire, J. T. & Kable, J. W. (2013). Rational temporal predictions can underlie apparent failures of delay gratification. *Psychological Review* 120 (2): 395–410.
Mele, A. R. (1987). *Irrationality: An essay on akrasia, self-deception, and self-control.* Oxford: Oxford University Press.
Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy.* Oxford: Oxford University Press.
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.
Miller, G. E., Yu, T., Chen, E., & Brody, G. H. (2015). Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. *PNAS* 112 (33): 10325-10330.
Mischel, W. & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology* 16 (2): 329–337.
Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science* 244: 933–938.
Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., Kross, E., Teslovich, T., Wilson, N. L., Zayas, V., & Shoda, Y. (2011). 'Willpower' over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience* 6 (2): 252–256.
Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., Sears, M.R., Thomson, W.M., & Caspi, A. (2011): A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America* 108(7): 2693–2698.
Muraven, M., & Baumeister, R. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin* 126 (2): 247–59.
Muraven, M., Shmueli, D., & Burkley, E. (2006). Conserving self-control strength. *Journal of*

Personality and Social Psychology 91(3), 524–537.
Oliver, M. (1996). *Understanding Disability: From Theory to Practice.* New York: St Martin's Press.
Paul, L.A. (2015). What you can't expect when you're expecting. *Res Philosophica* 92 (2): 149–170.
Petersen, M.A., Enghoff, O., & Demant, J. The uncertainties of enhancement: A mixed-methods study on the use of substances for cognitive enhancement and its unintended consequences. *Performance Enhancement and Health* 2019:6:111–20.
Plato (1987). *The Republic*. Lee, D. (transl.). London: Penguin Books.
Rachlin, H. (2000). *The Science of Self-Control*. Cambridge, Mass.: Harvard University Press.
Radoilska, L. (2012). Autonomy and Ulysses arrangements. In Radoilska, L. (ed.), *Autonomy and Mental Disorder.* Oxford: Oxford University Press.
Rawls, J. (1999). *A Theory of Justice: Revised Edition.* Cambridge, MA: Harvard University Press.
Ryle, G. (1949). *The Concept of Mind*. Chicago: University of Chicago Press.
Sen, A. K. (1973). Behaviour and the Concept of Preference*. Economica* XI, (159): 241–259.
Shakespeare, T. (2006). *Disability Rights and Wrongs.* London & New York: Routledge.
Shaw, E. (2014). Direct Brain Interventions and Responsibility Enhancement. *Criminal Law and Philosophy* 8(1): 1–20.
Shaw, E. (2016). The Right to Bodily Integrity and the Rehabilitation of Offenders Through Medical Interventions: A Reply to Thomas Douglas. *Neuroethics*. doi:10.1007/s12152-016-9277-4.
Shepherd, J. (2017). Halfhearted action and control. *Ergo* 4 (9): 259–276.
Shoemaker, D. (2015). *Responsibility from the Margins*. Oxford: Oxford University Press.
Snoek, A., Levy, N., & Kennett, J. (2016). Strong-willed but not successful: The importance of strategies in recovery from addiction*. Addictive Behaviors Reports* 4: 102–107.
Sripada, C. S. (2014). How is willpower possible? The puzzle of synchronic self-control and the divided mind. *Noûs* 48 (1): 41–47.
Sripada, C.S., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science* 25: 1227–1234.
Sripada, C. S. (2017). Frankfurt's unwilling and willing addicts. *Mind* 126 (503): 781–815.
Sripada, C.S. (2020). The atoms of self-control. *Noûs,* online first. DOI: 10.1111/nous.12332
Storebø, O.J., Krogh, H.B., Ramstad, E., Moreira-Maia, C.R., Holmskov, M., Skoog, M., Nilausen, T.D., Magnusson, F.L., Zwi, M., Gillies, D., Rosendal, S., Groth, C., Rasmussen, K.B., Gauci, D., Kirubakaran, R., Forsbøl, B., Simonsen, E., & Gluud, C. (2015). Methylphenidate for attention-deficit/ hyperactivity disorder in children and adolescents: Cochrane systematic review with meta-

analyses and trial sequential analyses of randomised clinical trials. *BMJ* 351:h5203.

Strawson, P. (1962). Freedom and resentment. In G. Watson (ed.), *Proceedings of the British Academy, Volume 48: 1962.* Oxford: Oxford University Press.

Tabibnia, G., Monterosso, J. R., Baicy, K., Aron, A. R., Poldrack, R. A., Chakrapani, S., Lee, B., & London, E. D. (2011). Different forms of self-control share a neurocognitive substrate. *Journal of Neuroscience* 31 (13): 4805–4810.

Thoma, J. (2020). In defence of revealed preference theory. *Economics & Philosophy*, online first. DOI: https://doi.org/10.1017/S0266267120000073

Tremain, S. (2017). *Foucault and Feminist Philosophy of Disability.* Ann Arbor: University of Michigan Press.

Uusitalo, S. (2015). *Addiction in action*. Turku: University of Turku.

Varela, F., Thompson, E. & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Varelius, J. (2009). Defining mental disorder in terms of our goals for demarcating mental disorder. *Philosophy, Psychiatry & Psychology* 16 (1): 35–52.

Vierkant, T. (2014). Mental muscles and the extended will. *Topoi* 33 (1): 57-65.

Vierkant, T. (2015). Is willpower just another way of tying oneself to the mast? *Review of Philosophy and Psychology*, 6: 779–790.

Vincent, N. (2010). Madness, Badness, and Neuroimaging-based responsibility assessments. In Freeman, M. (Ed.), *Law and Neuroscience, Current Legal Issues (Vol. 13*). Oxford: Oxford University Press, 79–95.

Vincent, N. (2014). Restoring Responsibility: Promoting Justice, Therapy and Reform Through Direct Brain Interventions. *Criminal Law and Philosophy* 8(1): 21–42.

Vohs, K. D., Schmeichel, B. J., Gronau, Q. F., Finley, A., Wagenmakers, E. J., and Albarracín, D. A. (2021). Multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science.* [Preprint].

Weber, E.U., Johnson, E.J., Milch, K.F., Chang, H., Brodscholl, J.C., & Goldstein, D.G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science* 18: 516–523.

Wurzman, R., Hamilton, R.H., Pascual-Leone A., & Fox, M.D. (2016). An open letter concerning do-it-yourself users of transcranial direct current stimulation. *Annals of Neurology* 80 (1): 1–4.

Zachar, P. (2014). *A Metaphysics of Psychopathology.* Cambridge, MA: The MIT Press.

Summary of the original articles

Article 1, “Self-control in responsibility enhancement and criminal rehabilitation” (with co-authors Susanne Uusitalo and Jarno Tuominen) critically examines the prospects of improving convicts’ capacity responsibility by means of self-control neuroenhancement. It argues that in addition to the fact that evidence on the efficacy of self-control neuroenhancement is sparse, special characteristics of the target population generate further difficulties in efficacy, which also reflects on the permissibility of such an intervention. Mandating neurointerventions of self-control therefore falls outside the scope of legitimate court rulings.

Article 2, “Born which way? ADHD, situational self-control, and responsibility”, discusses agents’ responsibility for failures of self-control. When failures of self-control are connected to disorders such as ADHD, responsibility mitigation is often warranted. Often, however, arguments in favour of responsibility mitigation rest on false claims about the role of the genetic and neural aetiology of ADHD on those failures of self-control. I discuss the role of these notions about the genetics and neurobiology of ADHD at length. I then argue that rather than embrace the essentialist notion that ADHD directly diminishes self-control and, therefore, responsibility, we ought to think of ADHD as constraining only some self-control practices. In particular, situational self-control strategies remain feasible for people with ADHD. However, not all individuals have *access* to these strategies. I suggest a way to evaluate responsibility in terms of situational rather than agential pleas, which tracks whether the individual had access to self-control behaviours.

Article 3, ”Accessing self-control”, argues that differences in self-control are analogous to differences in mobility: they are modulated by inherent traits and environmental supports and constraints in interaction. This joint effect of individual (neuro)biology and environmental factors is best understood in terms of access to self-control behaviours. I sketch an account of access as including the three criteria of means, awareness, and non-excessive effort. Using the example of ADHD, I demonstrate that difficulties in executive functioning do not necessarily cause poor self-control but rather, it is a social contingency that persons with such difficulties stand at a disadvantage with regard to self-control.

Article 4," Towards informed user decisions about pharmacological cognitive enhancement", claims that making informed decisions about cognitive enhancement requires a nuanced risk-benefit analysis that is not accessible for many users. Presently, the enhancement use of stimulant medications such as methylphenidate and modafinil is widespread, but most commonly happens without medical supervision. Direct and indirect barriers generate a situation where the risks and benefits of neuroenhancement are inequitably distributed; as a result, pharmacological neuroenhancement is sometimes not in the user's best interest. As a result, even if these pharmaceuticals were equitably distributed, its associated risks and benefits would not be. I also discuss possible strategies for amending this state of affairs.

Crim Law and Philos (2018) 12:227–244
https://doi.org/10.1007/s11572-017-9423-z

ORIGINAL PAPER

Self-Control in Responsibility Enhancement and Criminal Rehabilitation

Polaris Koi[1] · **Susanne Uusitalo**[2] · **Jarno Tuominen**[3,4]

Published online: 6 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract Ethicists have for the past 20 years debated the possibility of using neurointerventions to improve intelligence and even moral capacities, and thereby create a safer society. Contributing to a recent debate concerning neurointerventions in criminal rehabilitation, Nicole Vincent and Elizabeth Shaw have separately discussed the possibility of responsibility enhancement. In their ethical analyses, enhancing a convict's capacity responsibility may be permissible. Both Vincent and Shaw consider self-control to be one of the constituent mental capacities of capacity responsibility. In this paper, we critically examine the promise of improving convicts' capacity responsibility by neuroenhancements of self-control to see whether the special characteristics of the inmate population make a difference in the analyses. As improving self-control by means of neurointerventions seems plausible, we then ask whether it is or could be a justified measure in court rulings. We conclude that, even if there are cases in which neurointerventions were warranted in the context of the stated goals of the criminal court, i.e., decreasing recidivism and rehabilitating the offenders to the society, due to the range of individual variability in the constitution of self-control, the prescription of specific neurointerventions of self-control falls outside the scope of legitimate court rulings.

Keywords Responsibility · Enhancement · Rehabilitation · Criminal rehabilitation · Self-control · Neurolaw

✉ Polaris Koi
polaris.koi@utu.fi

[1] Philosophy, Department of Philosophy, Contemporary History, and Political Science, University of Turku, Turku, Finland

[2] Social and Moral Philosophy, Department of Political and Economic Studies, University of Helsinki, Helsinki, Finland

[3] Centre for Cognitive Neuroscience, Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

[4] Brain and Mind Center, University of Turku, Turku, Finland

1 Introduction

If future neuroscience could alter the moral and other capacities of humans, could these neuroenhancement technologies be used on convicted felons in order to promote rehabilitation and decrease recidivism rates? Some ethicists, such as Persson and Savulescu (2008) and Harris (2011), have debated the possibility of improving our moral faculties to either above average, or above what is currently humanly possible, thereby creating a safer society. A related prospect is the use of neurointerventions in criminal rehabilitation. While the use of what Douglas (2014, 2016) terms "neurocorrectives" may or may not overlap with moral neuroenhancement, both share the aim of a safer society and the approach of character improvement. Some courts already prescribe certain medical interventions, such as libido suppressants for sexual offenders, as part of criminal sentences.[1] Echoing the character-improvement agenda of moral enhancement, philosophers including Vincent (2014), Shaw (2013, 2014), Pugh and Douglas (2016, 2017), have all entertained and debated the prospect that future neuroscience would make it possible to improve the characters and capabilities of convicted felons. According to both Vincent and Shaw, one such capability to be modulated is responsibility.[2] The denotations of responsibility range from legal culpability to reliability of character but, in the context of criminal rehabilitation powered by neuroscience, two denotations have garnered the most interest. These are virtue responsibility, which refers to maturity and reliability of character, and capacity responsibility, which refers to the presence of mental capabilities necessary for responsible agency. To use Vincent's example, a well-behaved eight-year-old can be described as "virtue responsible," but not yet "capacity responsible"; likewise, a wayward teenager may already have the necessary mental abilities to count as capacity responsible, while he may not be virtue responsible (Vincent 2009, 2010).

While neurointerventions in criminal rehabilitation in general have been under somewhat wider discussion (see, e.g., Pugh and Douglas 2017; Greely 2008; Ryberg 2012), court-mandated improvement of a convict's capacity responsibility, specifically, has been discussed by Vincent (2014) and Shaw (2014). Resting on the idea that responsibility tracks mental capacity, both Vincent and Shaw suggest that the improvement of mental abilities necessary for responsible agency improves capacity responsibility. However, they each say very little about what neural or mental states and processes precisely would be targeted by capacity responsibility-enhancing interventions. The mental abilities they each list as underlying capacity responsibility are *understanding* and *self-control* (Shaw 2014; Vincent 2009). Understanding and self-control are broad, yet distinct, phenomena. Hence, we believe that the neuroenhancement of each merits a distinct account. Enhancing understanding has been explored in Shaw (2013). In this paper, we critically examine the promise of improving convicts' capacity responsibility by neuroenhancements of self-control[3] and whether it is or could be a justified measure in court rulings.

[1] For a review of neurointerventions in criminal rehabilitation, see Pugh and Douglas (2017).

[2] Others have also been suggested: for example, Pugh and Douglas (2017) point out that "recent research has suggested that pharmacological agents could be used to affect certain traits that are linked to criminal behavior, such as aggression, impulsivity, and the willingness to inflict harm on others.".

[3] We choose to tackle self-control in this paper, because it has recently been under extensive study (see, e.g., Tabibnia et al. 2011; Kotabe and Hofmann 2015; Moffitt et al. 2011), which we can draw on in order to have an empirically informed conception of what self-control enhancement is, and would be, like; furthermore, we do this because problems in self-control have widely been associated with criminal propensity, an association we will elaborate on in Sect. 3.

Various definitions and conceptions of self-control are present in the discussion at hand. The two standard accounts of self-control define it as inhibition control (e.g., Muraven and Baumeister 2000), and as acting on a larger, later reward over a smaller, sooner one (e.g., Mischel and Ebbesen 1970; Fujita 2011). Our argument is compatible with any standard definition of self-control; see Sect. 3.2 for a broader discussion of the concept of self-control. By "neuroenhancements" of self-control, we refer to medical interventions that target the neurophysiological basis of self-control. The most obvious forms of such enhancements are psycho-active pharmaceuticals, including stimulants such as methylphenidate (Sripada et al. 2014), but neuroenhancements could also take the form of, for instance, transcranial stimulation (Hsu et al. 2011) and other brain intervention methods.

In what follows, we first review the current discussion concerning responsibility neuroenhancements in criminal rehabilitation. Neurointervention-based criminal rehabilitation has faced a variety of objections based on autonomy (cf. Caplan 2006), authenticity (Vincent 2014), and bodily integrity (Shaw 2016). In contrast to these objections, stemming from concerns regarding the possible agency-undermining consequences of the medical intervention on the convicted individual, Vincent (2014) and Shaw (2014) have discussed these interventions in terms of the aims, and justified scopes of action, of justice systems. We will address whether such enhancements can help further the aims of the state, namely decreasing recidivism and aiding individual rehabilitation, and whether their use for those aims is justified, as well as account for our focus on self-control enhancements specifically by discussing both the concept of self-control and the prospects of enhancing it. In this paper, we adopt a similar strategy to that of Vincent and Shaw, ultimately arguing that the prescription of specific neurointerventions of self-control falls outside the scope of legitimate court rulings.

2 Responsibility Neuroenhancements in Criminal Rehabilitation

As outlined above, and as Vincent (2009, 2010, 2014) has repeatedly demonstrated, the concept of responsibility has a number of different denotations. Vincent (2014) argues various senses of responsibility (e.g., blameworthiness and criminal responsibility) necessitate capacity responsibility, which in turn relies on the presence certain mental capacities, including self-control.

For an agent to be responsible for some specific action, the agent needs to be capable of the kind of responsibility at stake, which requires a range of mental capacities. While there are obvious differences in scope and demandingness between, say, characterizing someone as fully capacity responsible or holding them blameworthy for some specific, mundane action, responsibility assessments rely on the evaluation of whether certain mental capacities are present.

Noting that neuroscience has managed to outline the neural basis of a variety of mental capacities, Vincent (2010) argues that neuroimaging technologies may, in the future, help courts conduct responsibility assessments to evaluate whether a defendant is capacity responsible to the degree that they[4] can held accountable for their act. She then goes on to suggest that future neuroscience may help restore or enhance a convict's capacity responsibility, thereby helping the state prevent recidivism and furthering rehabilitation (Vincent 2014). However, she also voices a concern that neurointerventions intended to

[4] Throughout this article, we use the singular "they" to refer to persons of unknown gender, instead of the more cumbersome and less inclusive "he or she.".

enhance responsibility may fail to do so because responsibility is also undergirded by a sense of authenticity, mental capacity ownership, and personal identity, which interventions may undermine. Due to this, she argues, *restored* mental capacities may fail to promote capacity responsibility.

Shaw (2014), echoing Vincent, holds enhancing capacity responsibility more likely to be permissible than the use of neurointerventions to enhance virtue responsibility, which she argues would amount to instilling the state's preferred values in the offenders in question, which would objectify them, treating them as means to an end—the end of decreased recidivism rates—instead of ends in themselves. Instead, Shaw holds that the enhancement of a convict's capacity responsibility via improving their understanding and self-control can help *enable* offenders to participate in rational moral dialogue, which in turn aids rehabilitation and reintegration. Furthermore, it can rehabilitate criminals whose criminal behavior is caused either by difficulties with delayed gratification, or by lapses in responding to reasons against breaking the law. Improving these capacities, Shaw argues, does not bind the subject to the state's preferred values, nor does it restrict their freedom of will by foreclosing the option of reoffending.

Shaw's approach draws on Vincent's (2009) distinction between virtue responsibility and capacity responsibility. Vincent defines virtue responsibility as being willing to "do the right thing" and to having a history of behavior that testifies to the fact that agents are willing to act conscientiously in general. Vincent uses the example of Jane, a "responsible little girl." "She gets up in the morning all by herself, she washes and gets dressed and even makes her own lunch, she doesn't fight with other kids at school, after school she does her homework, she cleans up after herself, she helps me make dinner, and she even looks out for her older brother John," all testifying to the virtue responsibility of Jane, despite that, being an eight-year-old child, she does not yet possess capacity responsibility (Vincent 2009, p. 115). By contrast, capacity responsibility entails having the prerequisite mental capacities, but does not necessarily entail virtue responsibility: someone fully capable of such responsible behavior depicted above may or may not behave conscientiously. In Shaw's view, whether one behaves virtue responsibly needs to be, ultimately, up to the agent in question. Improving capacity responsibility, however, may make it easier for an agent to choose to act conscientiously.

Shaw (2014) holds that the offender's informed consent is required for capacity responsibility enhancements. However, a number of concerns have been raised in the literature concerning whether informed consent is possible in the circumstances in question (see, e.g., Greely 2008). Responding to these concerns, Douglas (2014, 2016) has offered a way to circumvent this problem by arguing that requirements for consent are not invoked in the use of neurointerventions in criminal rehabilitation, given that they are not invoked in the case of incarceration, and that there is no compelling reason to require consent in the case of the former but not the latter. Douglas points out that attempts at such, e.g., invoking a right to bodily integrity, rely on factors that are not unique to brain interventions but are also present in the case of incarceration.

Concurring with Douglas, we hold that similar ethical concerns arise with respect to both neurointerventions and incarceration. One reason that many look into neurointerventions for solutions in criminal rehabilitation is that incarceration is typically not very effective in preventing recidivism or facilitating individual rehabilitation (cf. Criminal Sanctions Agency 2015). We worry that court-mandated self-control enhancements may have the same shortcoming.

3 Self-Control In and Out of Criminology

3.1 Self-Control in the General Theory of Crime

A correlation between poor self-control and criminal behavior has been observed in both psychology and criminology (see, e.g., Moffitt et al. 2011; Beaver et al. 2010; Wikström and Treiber 2007). In their seminal work, *A General Theory of Crime* (1990), Michael Gottfredson and Travis Hirschi have even proposed low self-control to be the source of all crime. Empirical evidence has since confirmed that low self-control, as defined in the General Theory of Crime, does correlate with and predict criminal behavior—although this effect has been found weaker than expected, especially in longitudinal studies (Pratt and Cullen 2000; Burt et al. 2006). Whether self-control fulfills its promise of explaining all criminal activity, evidence clearly suggests it to play an important role in the aetiology of criminal action.[5]

However, the precise role of self-control in the aetiology of criminal action is ambiguous. This is, in part, because of the various notions of "self-control": it is often studied as either impulse control (e.g., Muraven and Baumeister 2000) or delayed gratification (e.g., Mischel and Ebbesen 1970), while others rely on more robust character trait conceptions of self-control. Gottfredson and Hirschi's definition of "self-control" *in terms of* criminal propensity and delinquency is an examplar of the latter notion. They start out describing "self-control" as the ability to defer gratification: "A major characteristic of people with low self-control is therefore a tendency to respond to tangible stimuli in the immediate environment …. People with high self-control, in contrast, tend to defer gratification." (1990: p. 89.) However, they then proceed to elaborate on how the tendency to pursue immediate gratification is composed of six behavioral elements, out of which impulsivity is only one. The rest are described as temper, risk taking, self-centeredness, preference for simple solutions to problems, and preference for physical activity over more abstract, such as verbal, forms of action; these characteristics are best detected from the person's history of behavior (Gottfredson and Hirschi 1990; Piquero 2008). While not identical, the picture of self-control painted by Gottfredson and Hirschi bears some resemblance to Vincent's characterization of virtue responsibility. Self-control as defined by Gottfredson and Hirschi is a complex construct of characteristics rather than a specific capacity: all deviant behavior testifies to the poor self-control of the agent in question. Both are assessed, at least partly, by examining the past record of deviant or norm-conforming behavior; and both submit that an agent with good self-control, or a virtue responsible agent, will conform to a broad range of normative expectations. While this synergy between the General Theory of Crime (and the empirical studies affirming it) and Vincent's conception of virtue responsibility strengthens the claim that responsibility enhancement may succeed in decreasing crime, in order to examine the role of self-control as a more defined, value-neutral capacity that factors into capacity responsibility, we need to conceive of self-control in a different way.

[5] It is, nevertheless, unclear what kind of role self-control actually plays in bringing about the desired kind of action and preventing others. Is criminal action in general assumed to be so tempting that self-control is required to refrain from that action? If it is, is self-control considered something that is sufficient for prevention of criminal action? Is it, for instance, a muscle-like faculty that motivates the agent to a better life or ability to contemplate the feasible courses of action in a reasoned way?

3.2 Self-Control as a Capacity

Definitions and characterizations of self-control abound in the philosophical and empirical literature. They vary in terms of longitudinality, ranging from accounts describing self-control as a relatively stable character trait, virtue, or capacity (e.g., Gottfredson and Hirschi 1990; Baumeister and Exline 1999) to situational self-control, in which self-control is analyzed as a situational concept, referring to part of the decision-making process (Wikström and Treiber 2007). They vary in terms of the precise nature of the capacities or traits invoked, from accounts stressing inhibition control (e.g., Muraven and Baumeister 2000) to those in which the delay of gratification and the advancement of distal over proximal motivations plays a central role (e.g., Mischel and Ebbesen 1970; Fujita 2011). Finally, some accounts stress the environmental, context-bound elements of self-control by emphasizing the importance of environmental manipulation (e.g., Levy 2017; Duckworth et al. 2016; Pratt et al. 2004). Recently, Kotabe and Hofmann (2015) have proposed a model integrating these approaches to, or "components of," self-control into a construct describing a complex deliberative process that involves both the neurobiology of the agent and the environmental constraints at play.

Finally, characterizations of self-control range from the non-normative (such as self-control as the inhibition of otherwise imminent impulses, regardless of the content of these impulses; see, e.g., Tabibnia et al. 2011; Cohen et al. 2013) to descriptions placing it firmly in the sphere of morality (e.g., Baumeister and Exline 1999). Most fall somewhere in between, displaying terminology intended as neutral, yet laced with normative notions (Horstkötter 2009, 2015).

Discussing self-control as something that can be a target of neuroenhancement only makes sense if self-control is a trait or capacity that can be neurobiologically manipulated, and that is relatively stable so that enhancements thereof are not too fleeting to be useful. Situational definitions of self-control, such as that of Wikström and Treiber, pose a challenge to this approach. Wikström and Treiber argue that, instead of being a trait of the agent, self-control emerges from the situation it is used in. However, even Wikström and Treiber (2007) allow that, while self-control itself is situational in character, it relies on executive functioning, which can in turn be improved in order to improve the subject's chances at success in situational self-control.[6]

In the well-known Stanford marshmallow experiment, Walter Mischel and Ebbe B. Ebbesen approached the research of self-control via studying delayed-gratification behavior. Out of Mischel and Ebbesen's test subjects, many of the children who successfully refrained from eating the marshmallow accomplished that task by turning their attention elsewhere, such as by singing, turning their backs to the marshmallow, or attempting to go to sleep (Mischel and Ebbesen 1970)—"These children seemed to

[6] The reader may question whether this approach leads us to a similar situation regarding self-control enhancement as we had regarding responsibility enhancement in the first place, that is, whether self-control is a construct so complex that we need to examine enhancing its underlying capacities or traits one-by-one in order to understand what enhancing self-control would actually be like. As regards self-control, however, we argue that whether we adopt a situational or a character trait analysis of the concept of self-control, our endeavors to enhance self-control neurobiologically will involve the modulation of the executive capacities. In fact, that the executive functions are the underlying processes in self-control appears uncontroversial (see, e.g., Tabibnia et al. 2011; Corbett et al. 2009), whereas responsibility, even capacity responsibility, is a broader construct and involves a number of differently operating mental capacities, including the executive functions. We submit that, even if one conceptualizes self-control in terms of circumstances, enhancing the executive functions is likely to yield better self-control outcomes as long as the circumstances are conducive to self-control.

facilitate their waiting by converting the aversive waiting situation into a more pleasant non-waiting one" (ibid., p. 335), in other words, into one where exercising self-control was no longer necessary. The importance of environment and situation management in self-control situations has since been affirmed by a number of studies, including a recent study by Ent et al. (2015) showing that high trait self-control significantly predicts temptation avoidance, both in an experimental setting and in self-reports. Drawing on these findings, among others, Levy (2017) has pointed out that people with high trait self-control appear not to be self-controlled but rather to exhibit a range of skills or practices enabling them to avoid using their self-control altogether, along with opportunities for employing these skills. This analysis relies on a distinction between self-controlled behavior and willpower as the related (neuro)psychological trait of the agent, in which the frequency of the former may not track the strength of the latter. This distinction is further supported by the role of targeted, skills-based behavioral interventions (Knouse and Safren 2010) and environmental manipulation in the context of managing disorders that impair self-control, such as Attention Deficit Hyperactive Disorder (ADHD) (see, e.g., Barkley 1997). More modestly, the many self-help techniques available for people willing to improve their self-control also contribute towards showing that self-control is a product of the interaction of neural processes with the environment: it involves a set of practices, and can be both aided and hampered by environmental and social factors.[7]

Despite the importance of the environment, the empirical research does not imply that self-control could not be neurobiologically enhanced—quite the contrary. There is evidence that, while forms of self-control differ, they share a neurobiological substrate (Tabibnia et al. 2011; Casey et al. 2011)—a substrate that may be modulated in order to improve self-control. Furthermore, Berman et al. (2013) were able to predict whether a person was in the low or high self-control group with 71% accuracy, based on the dimensionality analyses of their brain networks while undergoing a working memory task. To give an example of self-control enhancement already in use, methylphenidate and other substances used to alleviate the symptoms of ADHD have a strong track record in improving the self-control abilities of people with ADHD, a disorder that has been characterized as a disorder *of* self-control (Barkley 1997).[8] The neuroscientific advances in understanding the neural mechanisms involved in self-control, together with the strong

[7] While self-control has often been accounted for as the control of the self by the self, it is not dependent on the absence of control by others. Someone or something can control the agent, causing the agent to control herself. We may exercise self-control under coercion, brainwash, and other kinds of pressure.

[8] Relevant to our discussion, see a large Swedish study that found ADHD medication to reduce criminality associated with male ADHD (Lichtenstein et al. 2012). It should be noted, however, that recently the exact benefits and long-term efficacy of the medication in ADHD children and youth have been questioned. Some studies indicate the effects seemingly to focus on teacher or parent evaluations of perceived academic performance, not on performance or symptoms themselves (for a recent review, see Storebø et al. 2015). Other studies have found the effects of ADHD medication to either disappear over time (Jensen et al. 2007; Molina et al. 2009; Langberg and Becker 2012) or, in one study, even deteriorate academic performance (Currie et al. 2014). This discussion could, however, also have other societal or cultural factors in the educational system, which may explain it being somewhat at odds with current clinical practice, and the perceived effects of stimulant medication (for a broad review, see Baroni and Castellanos 2015). When considered as a possible cognitive enhancer, methylphenidate has been found to prevent self-control depletion in a simple inhibition task in healthy volunteers (Sripada et al. 2014). Another ADHD stimulant, mixed amphetamine salts (Adderall), on the other hand, was found to have only a small effect on various cognitive measures, including no effect on inhibitory control, despite the subjects' self-perceived enhancements (Ilieva et al. 2013). While both medications target the epinephrine and dopamine systems, the differences in effects can be due to either different psychopharmacological properties or study design, and clearly calls for further research (for a critical look at cognitive enhancers in general, see Zohny 2015).

track record of existing medical interventions on self-control, point towards the idea that, regardless of self-control's involvement of environmental management and situational factors, it is a promising candidate for an ability to be enhanced neurobiologically with both existing and future neuroscientific means. We grant Vincent's and Shaw's assessments that self-control contributes towards capacity responsibility. This suggests that it might well be possible to improve an agent's capacity responsibility by neurobiologically enhancing their self-control. This raises questions as to whether these interventions can help further the aims of the state, that is, to decrease recidivism and aid individual rehabilitation, and whether they can permissibly be used to do so.

4 The Goals of the State

The two goals of the state evoked in the debate on neurointerventions in criminal rehabilitation are, first, decreasing recidivism rates; and second, rehabilitating individual criminals into society. These tasks have been found challenging. Shaw expresses concern about "a pressing need to develop more effective ways of re-integrating offenders back into society" (2014: p. 1). Gottfredson and Hirschi "see little hope for important reductions in crime through modification of the criminal justice system," such as by means of rehabilitation programs (1990: p. XVI, 268–269). Could neurointerventions offer a novel means to help the state accomplish these two goals?

It is important to notice that, while rehabilitation may contribute towards decreased recidivism rates, these are fundamentally different goals. The goal of decreased recidivism works in the sphere of policy. It aims to decrease the statistical likelihood of recidivism, thereby decreasing frequency of crime, and improving the safety of the society. Individual rehabilitation, by contrast, is concerned with the well-being of the criminal offender in question, treating the offender's quality of life as an end in itself. We can imagine a society where offenders have not been rehabilitated, but they are nevertheless prevented from reoffending, say, by physically restraining them for life; likewise, an offender who has been fully rehabilitated and happily leads the life of a model citizen may still reoffend if placed in circumstances where no other viable courses of action are present. In the first of these two imagined scenarios, a safer society has been successfully created, but at the cost of treating the convicts as means to that end. In the second scenario, the goal of rehabilitating the individual has successfully been promoted, whereas recidivism has not been prevented. While these often go hand-in-hand, the distinction is worthwhile in ethical discussions of criminal rehabilitation methods. We therefore address self-control neuroenhancements' prospects in promoting each goal in turn.

4.1 Decreasing Recidivism Rates via Self-control Neuroenhancements: Feasibility Concerns

The promise of self-control enhancements in decreasing recidivism rates can be formulated in two, complementary ways.[9] The first way is to tackle self-control as a component of capacity responsibility, to echo both Shaw and Vincent:

[9] We present the reader with both the argumentative line stemming directly from self-control, and with the argument following Vincent's and Shaw's discussions of capacity responsibility because while the appeal directly to self-control is more attractive in its simplicity, we find the concept of capacity responsibility draws out a more nuanced conception of agency in the subjects.

(1) Self-control can be improved by means of neuroenhancements.
(2) Improving individuals' self-control improves their capacity responsibility.
(3) Improving criminal offenders' capacity responsibility decreases the likelihood of their reoffending.

Therefore, by using neuroenhancements that improve the self-control of criminal offenders, we improve their capacity responsibility and decrease recidivism rates.

Another possible formulation postulates a direct relationship between self-control and recidivism:

(1) Self-control can be improved by means of neuroenhancements.
(2b) Improving criminal offenders' self-control decreases the likelihood of their reoffending.

Therefore, by using neuroenhancements that improve self-control on criminal offenders, we decrease recidivism rates.

We accept premise (2), agreeing with both Shaw and Vincent that self-control is one of the abilities on which capacity responsibility is reliant.[10] It is premises (1) and (3)/(2b) about which we are concerned.

Premise (1) appears uncontroversial. Existing neurological means, as well as the growing body of knowledge about the neurobiological foundation of self-control, testify to the malleability of self-control by means of neurointerventions. However, due to the importance of environmental factors in self-control, the results of neurointerventions of self-control are uncertain. Before returning to this point, we will discuss premises (3) and (2b).

Concerning premises (3) and (2b), it appears statistically correct that low self-control is a criminogenic factor (Pratt and Cullen 2000; Moffitt et al. 2011). However, we would like to present some concerns arising from the fact that certain crimes are not committed due to low self-control. Rawn and Vohs (2011) outline how expected social rewards may motivate agents to employ their self-control and engage in risky, harmful, and criminal behaviors. Certain crimes are calculated, or products of rational behavior, from the shoplifter who could not otherwise afford a balanced diet to politically motivated murders. It appears that, in many cases, ranging from civil disobedience and activism, to crime done for pressing prudential reasons, and politically motivated crimes, improving capacity responsibility would be unlikely to decrease the likelihood of reoffending.[11] We therefore stress that, insofar as we aim to decrease recidivism, the improvement of self-control or of capacity responsibility in criminals is applicable only to those criminals for whom poor self-control or poor capacity responsibility *does* underlie their criminal behavior.

However, this may be a significant enough majority to render these concerns minor, especially if we are able to detect which criminals fall into the group where self-control

[10] There may be certain constraints to the capacity responsibility-improving effect of self-control enhancement. These include upper and lower thresholds, above or below which changes in self-control produce no change in capacity responsibility; and prerequisites, including certain criteria for the subject's understanding (recall Vincent's characterization of capacity responsibility as based on understanding and self-control) in order for changes in self-control to be effective in altering their capacity responsibility. However, these constraints are minor. It is standard for a medical procedure, which neuroenhancements are, to have its set of cases where it is inapplicable. This does not detract from its value.

[11] Some may even worry that, by improving criminals' self-control, we would end up with more successful criminals, i.e., criminals who are better at committing their crimes while avoiding detection.

improvement is a viable intervention, and which do not.[12] For this reason, we consider this concern supplementary to our main concern, which has to do with premise (1), and to which we now turn.

Above, we discussed existing medical means for modulating self-control, and found self-control a promising target for neuroenhancement. However, we have reservations about whether this is so when the target group is criminal offenders. Namely, we wish to point out that there are factors that must be given due consideration in assessing whether self-control neuroenhancements, even if they were largely feasible and effective in the general population, are likely to be as effective in criminal offenders.

For many criminals, their lives are not going well. Neurodiversity and mental health diagnoses are common in convicted criminals. According to the Finnish Criminal Sanctions Agency, 25% of Finnish inmates have ADHD, and 90% have been diagnosed with some mental or neurological condition (Criminal Sanctions Agency 2015). Among these, learning disabilities such as dyslexia are common: in Finland, 33% of inmates have dyslexia and even more have poor literacy skills (ibid.). Mental health disorders are also overrepresented in the prison population in the United Kingdom (Ministry of Justice 2013), and 20–30% of all UK inmates have learning disabilities (Loucks 2007). Many mental disorders diminish self-control ability to some degree (cf. Cohen et al. 2013), and, in the case of ADHD, self-control is significantly impaired (cf. Barkley 1997). From this, we can infer that a large number of inmates have deficiencies in self-control.

In addition, many convicts may not have access to the basic prerequisites of human flourishing, such as nutrition, safety, and shelter, and have little control over their living circumstances. Homelessness is overrepresented in prison populations: in Finland, 12–14% of inmates have no fixed address, while in the UK, 15% of new convicts report being homeless prior to incarceration (Criminal Sanctions Agency 2015; Ministry of Justice 2012). In summary, we can safely say that a large number of convicts face a challenging combination of economic and social disadvantage, including mental health problems and neurodiversity.

This is a markedly challenging context for neuroenhancements of self-control because, as discussed above, self-control does not stop in neuroanatomy: it also depends on a variety of practices to a great extent. Consider, for example, people with ADHD, for whom neurointerventions in the form of stimulant medication often help improve their self-control. The medication alone does not do the trick: the patient also needs to set up practices that help them structure their environment (cf. Levy 2017; Kotabe and Hofmann 2015). The increase in dopamine levels caused by stimulant medication facilitates this behavior change, which may aid the patient in the task of structuring their environment, such as by setting in place behavioral constraints and nudges. By contrast, if we did modify convicts' neuroanatomy in a similar way, would those living in hardship have access to establishing and maintaining the kinds of practices on which self-control relies?

Again, some of them are likely to succeed; additionally, those convicts whose living conditions are stable enough to facilitate these practices would not face this problem. However, the prevalence of conditions that may hamper self-control practices in the criminal population raises concerns about whether self-control neuroenhancements, even if predominantly effective in the general population, would have a high likelihood of success in improving self-control in the criminal population. Furthermore, would the circumstances

[12] Naturally, offenders cannot be neatly divided into two groups: in many cases, the aetiology of crime may involve both deficiencies in self-control and/or capacity responsibility, and environmental factors, such as social pressure towards crime.

of successfully responsibility-enhanced offenders make it likely that they reoffend despite having improved self-control [given our concerns regarding premises (3) and (2b)]? In light of these concerns, we remain skeptical about the prospects of self-control enhancements for decreasing recidivism rates. Recall that what is attractive about court-mandated self-control neuroenhancements is that it would enable the application of enhancements to a large number of convicts, hopefully resulting in a substantially safer society.[13] In the case of a program with a widespread positive impact on public safety, the presence of some instances of ineffective, enforced intervention may be considered an acceptable adverse effect (see Pugh and Douglas 2016). We will further discuss the ethics of mandatory ineffective interventions in Sect. 5. While the intervention may succeed in improving the self-control of some individuals, there is no indication that this would have fairly consistent results in decreasing recidivism. The resulting improvement in public safety may therefore remain too small to warrant instituting a program where court-mandated neurointerventions become part of criminal sentencing by default.

4.2 Individual Rehabilitation via Self-control Neuroenhancements: Feasibility Concerns

Now that we have discussed our concerns regarding whether self-control enhancements would succeed in decreasing recidivism rates, we can move on to the state's other goal: criminal rehabilitation. This is the individual-centered goal of the justice system, where the subject is not discussed merely as statistics or in terms of their contribution to public safety; instead, the aim is to help each individual find a way to lead a satisfactory life while avoiding crime. The argument for using self-control neuroenhancements for criminal rehabilitation can be summed up as follows:

(1) Self-control can be improved by means of neuroenhancements.
(2) Improving individuals' self-control improves their capacity responsibility.
(3) Improving some criminal offenders' capacity responsibility helps their rehabilitation, reintegrating them into society and decreasing the likelihood of future crime.

Therefore, by applying neurointerventions that improve self-control on some criminal offenders, we help their rehabilitation, reintegrate them into society, and decrease the odds for reoffending.

Alternatively, echoing Gottfredson and Hirschi's General Theory of Crime, a more direct link between self-control improvement and rehabilitation can be made, omitting the role of capacity responsibility:

(1) Self-control can be improved by means of neuroenhancements.
(2b) Improving some criminal offenders' self-control helps their rehabilitation, reintegrating them into society, and decreasing the likelihood of future crime.

Therefore, by applying neurointerventions that improve self-control on some criminal offenders, we help their rehabilitation, reintegrate them into society, and decrease the odds for reoffending.

[13] To be clear, we do not argue that, if these interventions were largely successful, they should be prescribed by courts; rather, we have a number of concerns about this, many of which fall outside the scope of this paper, and have been dealt with elsewhere in the literature. We only stress feasibility over arguments regarding concerns, such as personal autonomy and treating people as mere means to ends because we choose to restrict ourselves to points arising specifically with regard to self-control enhancement—although similar points may arise with regard to other kinds of court-mandated medical interventions.

In this approach, no statistically large difference in recidivism rates is aimed at. Rather, the promise is that courts could include self-control enhancements as part of some criminal sentences, although how it would be discerned which criminal offenders on which to apply self-control neurointerventions is a point up for debate.

Using medical interventions as part of criminal sentences is not a new idea. For example, in some legislations, chemical castration is used as part of criminal sentences for sex offenders. Along the same lines, self-control neuroenhancements might, in the future, constitute part of the criminal sentence in crimes that are often thought to be due to self-control failure.

Reminding ourselves of the above discussion on the relationship between self-control and crime, it is fair to assume that there is a large number of criminals whose reintegration process would benefit from the improvement of self-control and of capacity responsibility. But, again, we are faced with the same concerns regarding the first premise as with the scenario regarding decreasing recidivism rates. The fact that many convicts would, due to their adverse life circumstances, be hampered in their ability to carry out the behavior changes necessary for the improvement of self-control in practice raises concerns that self-control neurointerventions may be unsuccessful in the context of criminal rehabilitation, save for some cases.

This is a good place to tackle one more argument for the use of neuroenhancements in a scenario like this. Recently, Ray (2016) has proposed that stimulants could be used for opportunity maintenance on healthy but socially disadvantaged subjects, in order to "level the playing ground" and giving underprivileged students opportunities they would otherwise lack. Applied to the context of self-control enhancements in criminal rehabilitation, given that many offenders come from underprivileged backgrounds, could self-control enhancements be used to remedy part of the effect of those backgrounds, thereby helping these individuals access life options that would otherwise be beyond their reach? This novel approach has also raised concerns, such as ones concerning feasibility (Erler 2016) and self-pathologization (Stevenson 2016). While framing neuroenhancements as a way to alleviate the burden of disadvantage appears to increase the attractiveness of self-control enhancements in criminal rehabilitation, it does not alleviate our concerns stemming from an understanding of self-control as requiring environmental scaffolding. In what follows, we propose that the benefits that self-control enhancement may have for some individuals could plausibly be reaped in the context of individual rehabilitation—but not in the criminal court.

5 Legitimate Self-Control Neuroenhancements in Criminal Rehabilitation: An Ethical Argument

The above discussion so far appears to leave us tied. On one hand, the robust connections between self-control and crime suggest that self-control enhancement, if successful in improving convicts' self-control, could be a promising means of rehabilitation; on the other, the prospects of neuroenhancement in large-scale criminal rehabilitation appear gloomy. The mandated use of neurocorrectives should be limited to those offenders who can be expected to benefit from the intervention. We argue that these measures are best exercised by placing them outside the domain of criminal justice, and inside that of medical care.

A source of much debate in the use of medical interventions as part of criminal sentences is that the question lies at the intersection of medical ethics and criminal justice ethics. Medical ethics stresses the importance of beneficence, non-maleficence, and informed consent and holds any intervention that risks harming the patient without obvious benefits to the patient ethically suspect, and doubly so if done without consent. Meanwhile, in the sphere of criminal justice ethics, it is standard to believe that we may be justified in inflicting some harm on criminals; doing some things that are not beneficent to them in order to benefit others; and doing things to them without their consent, such as forcibly imprisoning them. This friction is apparent in a number of debates, such as the debate on whether convicts can give valid consent to neurointerventions, and whether consent is indeed even required (cf. Shaw 2014; Douglas 2014; Pugh and Douglas 2016). It would greatly clarify matters if we could place self-control neurointerventions firmly in the sphere of either medical or criminal justice ethics.

We assume that, out of the schoolbook purposes of the criminal justice system—punishment, deterrence, rehabilitation, and decreasing recidivism—rehabilitation is the one that self-control neuroenhancements stand a chance of furthering.[14] Therefore, whether self-control neuroenhancements can be considered to promote any of the aims of criminal justice hinges on whether they are applicable for criminal rehabilitation by criminal courts. We have reservations about this, and argue for these reservations as follows:

(1) Medical interventions need to be reasonably effective in benefitting the patient in order to be considered criminal rehabilitation.
(2) In order for a neurointervention of self-control to be likely to succeed in improving the patient's self-control, the patient's individual circumstances need to be such that they facilitate setting up practices that allow for self-control.
(3) Criminal courts are ill equipped to assess whether the convict's individual circumstances facilitate setting up practices that allow for self-control, or to change those circumstances to that end.

Therefore, neurointerventions of self-control, if prescribed by criminal courts, cannot be considered part of criminal rehabilitation.

However, neurointerventions of self-control can be very useful when the subject's circumstances are such that a self-control neurointervention is likely to succeed. Consider, for example, stimulant medication for ADHD: while it does not succeed in helping all people with ADHD, its success rate is high enough that prescription is considered legitimate.[15] This is likely to be largely so because patients live in circumstances that enable them to develop practices of self-control. In these circumstances, the neurointervention can make a vast difference for the subject, making the subject better equipped to pursue any life plan they have. This does not foreclose the option of reoffending, but it can decrease its likelihood by improving the subject's chances in pursuing other courses of action. We

[14] This relies on the widespread idea that the purposes of criminal justice are rehabilitation (divided, in our analysis, to decreasing recidivism and individual rehabilitation), retribution, incapacitation, and crime deterrence (Alschuler 2003). We argued above that this is not a promising way to further the goal of decreasing recidivism rates. In our view, if retribution is necessary, neurointerventions are not needed for it, given that there exist more than enough means for inflicting suffering. We also assume that self-control neuroenhancements are a poor incapacitator and a poor deterrent of crime.

[15] Recently, however, a meta-analytical Cochrane Review on the efficacy of methylphenidate on children and adolescents with ADHD has cast light on the quality and implications of this research, highlighting that the effectiveness of methylphenidate treatment in children and adolescents remains uncertain; see Storebø et al. (2015).

therefore need a policy that would allow us to find out when the requirements for a successful self-control neurointervention are met. One possibility, in the sphere of mental health care, would be that, if properly trained, a team of medical professionals and social workers could be equipped to assess, with a reasonably high success rate, whether a subject's circumstances facilitate self-control; depending on available resources, the subject's living circumstances may furthermore be improved upon in order to facilitate this. The intervention can furthermore be supported with behavioral therapy in order to help establish the practices involved in self-control.

Many criminal justice systems currently use medical practitioners and social workers to inform at least some of their decision-making. For example, in the UK, expert witnesses may be called upon by the police, the prosecution, and the defense (Milroy 2003). This raises a question: would it not suffice that a multidisciplinary group of medical and social work experts be employed by the court to assess whether an individual meets criteria for intervention, and to lay out the parameters of the intervention prescribed by the court?[16] Giving this multidisciplinary group sufficient autonomy in making decisions about subsequent interventions would alleviate our concern about the inability of criminal courts to engage in tasks that require medical and social work expertise. However, we have two concerns about this arrangement. First, we worry about whether the context of court interaction would facilitate a therapeutic relationship between the convict and the medical and social work professionals involved. Second, and more importantly, court-mandated interventions are remarkably rigid. Should a criminal sentence require an intervention, say, by means of pharmaceuticals for the duration of five years, if the intervention were to be found unwarranted after one year, whether due to its remarkable efficacy, due to finding out it is ineffective in this case, or due to the circumstances of the convict no longer facilitating self-control, the convict would nevertheless be required to continue consuming the pharmaceuticals for the remainder of the sentence. This is especially problematic if adverse side effects are involved but, even if not, it already involves harms by virtue of undermining autonomy without facilitating either rehabilitation or public safety—harms that the beneficial outcomes of criminal rehabilitation may render acceptable, but that are indefensible if none of the goals of the criminal justice system end up being promoted. These considerations constrain the capacity of criminal courts to effectively further its own aims by prescribing neurocorrectives, even if expert advice is employed.

We submit that all a criminal court can legitimately include as part of the judgment is a requirement to visit a mental health practitioner for assessment. However, the form, content, and duration of any subsequent mental health care should be up to the patient and the medical professionals involved. Assessing whether this treatment may permissibly be involuntary should be conducted in accordance with the same ethical guidelines as in involuntary mental health care for non-offenders.[17] Given that criminal courts are poorly equipped to conduct tasks that require medical and social work expertise, for them to prescribe the specific form or duration of mental health care, such as a self-control

[16] We thank an anonymous reviewer for this point.

[17] While these guidelines vary on a regional basis, and are subject to debate and improvement, a good example of such a guideline is listed in the objectives of The European Mental Health Action Plan (WHO 2013): "All steps should be taken to promote voluntary admission and treatment, and avoid coercion, while guaranteeing protection in accordance with international and national human rights instruments. Strong safeguards need to be in place if involuntary admission and treatment are deemed necessary, including independent reviews, inspection of the conditions under which people are detained and access to complaints procedures, independent legal advice and other relevant support.".

neuroenhancement, runs the risk of overstepping the purpose of the criminal justice system, into the domain of illegitimate use of power.

6 Conclusion

Capacity responsibility tracks certain mental capacities, as Vincent and Shaw submit in their discussions of responsibility enhancement. Therefore, it is the relevant mental capacities that we need to examine in order to determine whether responsibility can indeed be enhanced. It is highly plausible that self-control is indeed one of the key capacities on which capacity responsibility hinges.

Self-control enhancement appears both feasible and attractive in the context of those with disorders of self-control as well as the general population. Insofar as improved self-control yields improved capacity responsibility, and capacity responsibility inversely correlates with criminality, its availability may even provide decreased crime rates. However, self-control neuroenhancements may be unsuccessful in improving self-control in criminal offenders who may lack the skills required to establish the kinds of practices of environmental manipulation that self-control behaviors necessitate, or whose environments may not be conducive to self-control practices.

Given the concerns that self-control neuroenhancement raises, as well as its attractiveness when successful, it both merits and requires further empirical study. Neuroscientific research has mostly looked at simple inhibition or delayed gratification tasks, without engaging with other aspects of self-control. The psychopharmaceuticals currently used, for example, as cognitive enhancers are likely to have more to do with maintaining vigilance and attention than with more complex issues related to self-control, such as future-oriented thinking and causal reasoning. These effects are still little studied, and conclusions drawn from them should be cautious (Zohny 2015). Furthermore, current techniques such as brain stimulation or medication usually have effects that are short-lived and could not as such be considered to reliably cause a lasting enough change to present an alternative to other measures of the criminal justice system. Our contribution has mainly served to highlight this need. The idea of enhancing criminal offenders' capacity for being responsible for their actions sounds like a desirable goal but, as things currently stand, it remains an idea that still necessitates more empirical and conceptual work. As our understanding of the neurobiology of self-control increases, the medical means of modulating self-control may improve in efficacy. But, however thoroughly we understand their neural correlates, self-control practices remain decidedly context-dependent and environmental in character. An individual-centric approach where self-control neuroenhancements are viewed in terms of health care rather than in terms of the justice system thus appears both more effective and more ethically sustainable.

Acknowledgements This paper stems from the authors' participation in "Improving Ourselves as Moral Agents: The Ethics and Neuroscience of Self-Control" cluster group, which was part of the John Templeton-funded Philosophy and Science of Self-Control research project. The authors would like to thank the research project and everyone who participated in the activities of our cluster group. This paper has benefited from comments from Jaakko Kuorikoski, Jukka Varelius, Juha Räikkä, and from participants in the Oxford Applied Ethics Discussion Group and in the Turku Practical Philosophers' Club (Klubi). Finally, Polaris Koi would like to thank the Finnish Cultural Foundation for research funding.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Alschuler, A.W. (2003). The Changing Purposes of Criminal Punishment: A Retrospective on the Last Century and Some Thoughts about the Next. *University of Chicago Law Review* 70(1): 1–22.

Barkley, R.A. (1997). *ADHD and the Nature of Self-Control.* New York: The Guilford Press.

Baroni, A., & Castellanos, F.X. (2015). Stimulants, cognitions and ADHD. *Current Opinion in Behavioral Sciences* 4: 109–114.

Baumeister, R.F., & Exline, J.J. (1999). Virtue, Personality and Social Relations: Self-Control as the Moral Muscle. *Journal of Personality* 67(6): 1165–1194.

Beaver, K.M., DeLisi, M., Vaughn, M.G., & Wright, J.P. (2010). The Intersection of Genes and Neuropsychological Deficits in the Prediction of Adolescent Delinquency and Low Self-Control. *International Journal of Offender Therapy and Comparative Criminology* 54(1): 22–42.

Berman, M.G., Yourganov, G., Askren, M.K., Ayduk, O., Case, B.J., Gotlib, I.H., Kross, E., McIntosh, A.R., Strother, S., Wilson, N.L., Zayas, V., Mischel, W., Shoda, Y., & Jonides, J. (2013). Dimensionality of brain networks linked to life-long individual differences in self-control. *Nature Communications* 4(1373): 1–7. doi:10.1038/ncomms2374.

Burt, C.H., Simons, R.L., & Simons, L.G. (2006). A Longitudinal Test of the Effects of Parenting and the Stability of Self-Control: Negative Evidence for the General Theory of Crime. *Criminology* 44 (2): 353–396.

Caplan, A.L. (2006). Ethical issues surrounding forced, mandated, or coerced treatment. *Journal of Substance Abuse Treatment* 31(2): 117–120.

Casey, B.J., Somerville, L.H., Gotlib, I.H., Ayduk, O., Franklin, N.T., Askren, M.K., Jonides, J., Berman, M.G., Wilson, N.L., Teslovich, T., Glover, G., Zayas, V., Mischel, W., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences of the United States of America* 108(36): 14998–15003.

Cohen, J.R., Berkman, E.T., & Lieberman, M.D. (2013). Intentional and Incidental Self-Control in Ventrolateral Prefrontal Cortex. In D.T. Stuss & R.T. Knight (Eds.), *Principles of Frontal Lobe Function, 2nd Edition.* Oxford: Oxford University Press, 417–440.

Corbett, B.A., Constantine, L.J., Hendren, R., Rocke, D., & Ozonoff, S. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. *Psychiatry Research* 166(2–3): 210–222.

Criminal Sanctions Agency (2015). *Statistical Yearbook 2015 of the Criminal Sanctions Agency.* Helsinki: Rikosseuraamuslaitos.

Currie, J., Stabile, M., & Jones, L. (2014). Do stimulant medications improve educational and behavioral outcomes for children with ADHD? *Journal of Health Economics* 37: 58–69.

Douglas, T. (2014). Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily integrity. *Journal of Ethics* 18(2): 101–122.

Douglas, T. (2016). Nonconsensual Neurocorrectives and Bodily Integrity: A Reply to Shaw and Barn. *Neuroethics.*. doi:10.1007/s12152-016-9275-6.

Duckworth, A.L., Gendler, T.S., & Gross, J.J. (2016). Situational Strategies for Self-Control. *Perspectives on Psychological Science* 11(1): 35–55.

Ent, M.R., Baumeister, R.F., & Tice, D.M. (2015). Trait self-control and the avoidance of temptation. *Personality and Individual Differences* 74: 12–15.

Erler, A. (2016). Using Stimulants to Tackle Social Disadvantages: Interesting in Theory, Problematic in Practice. *American Journal of Bioethics* 16(6): 48–50.

Fujita, K. (2011). On Conceptualizing Self-Control as More Than the Effortful Inhibition of Impulses. *Personality and Social Psychology Review* 15(4): 352–366.

Gottfredson, M.R., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford: Stanford University Press.

Greely, H.T. (2008). Neuroscience and Criminal Justice: Not Responsibility but Treatment. *University of Kansas Law Review* 56(5): 1103–1138.

Harris, J. (2011). Moral enhancement and freedom. *Bioethics* 25(2): 102–111.

Horstkötter, D. (2009). *Self-Control Revisited: Varieties of Normative Agency.* Nijmegen: Radboud University Nijmegen.

Horstkötter, D. (2015). Self-control and normativity: Theories in social psychology revisited. *Theory & Psychology* 25(1): 25–44.

Hsu, T., Tseng, L., Yu, J., Kuo W., Hung D.L., Tzeng O.J.L., Walsh, V., Muggleton, N.G., & Juan, C. (2011). Modulating inhibitory control with direct current stimulation of the superior medial frontal cortex. *NeuroImage* 56(4): 2249–2257.

Ilieva, I., Boland, J., & Farah, M.J. (2013). Objective and subjective cognitive enhancing effects of mixed amphetamine salts in healthy people. *Neuropharmacology* 64: 496–505.

Jensen, P.S., Arnold, L.E., Swanson, J.M, Vitiello, B., Abikoff, H.B., Greenhill, L.L., Hechtam, L., Hinshaw, S.P., Pelham, W.E., Wells, K.C., Conners, C.K., Elliott, G.R., Epstein, J.N., Hoza, B., March, J.S., Molina, B.S.G., Newcorn, J.H., Severe, J.B., Wigal, T., Gibbons, R.D., & Hur, K. (2007). 3-Year Follow-up of the NIMH MTA Study. *Journal of the American Academy of Child & Adolescent Psychiatry* 46(8): 989–1002.

Knouse, L.E., & Safren, S.A. (2010). Current Status of Cognitive Behavioral Therapy for Adult Attention-Deficit Hyperactivity Disorder. *Psychiatric Clinics of North America* 33(3): 497–509.

Kotabe, H.P., & Hofmann, W. (2015). On Integrating the Components of Self-Control. *Perspectives on Psychological Science* 10(5): 618–638.

Langberg, J.M., & Becker, S.P. (2012). Does Long-Term Medication Use Improve the Academic Outcomes of Youth with Attention-Deficit/Hyperactivity Disorder. *Clinical Child and Family Psychology Review* 15(3): 215–233.

Levy, N. (2017). Of Marshmallows and Moderation. In W. Sinnott-Armstrong & C.B. Miller (Eds.), *Moral Psychology, Volume 5: Virtue and Happiness.* Cambridge: MIT Press, 197–214.

Lichtenstein, P., Halldner, L. Zetterqvist, J., Sjolander, A., Serlachius, E., Fazel, S., Langstrom N., & Larsson, H. (2012) Medication for Attention Deficit-Hyperactivity Disorder and Criminality. *New England Journal of Medicine* 367(21): 2006–2014.

Loucks, N. (2007). *No One Knows: offenders with learning difficulties and learning disabilities – review of prevalence and associated needs.* London: Prison Reform Trust.

Milroy, C.M. (2003). Medical experts and the criminal courts: all relevant material must be disclosed, including facts detrimental to the opinion. *BMJ* 326(7384): 294–295.

Ministry of Justice (2012). *Accommodation, homelessness and reoffending of prisoners: Results from the Surveying Prisoner Crime Reduction (SPCR) survey.* London: Ministry of Justice.

Ministry of Justice (2013). *Gender differences in substance misuse and mental health amongst prisoners.* London: Ministry of Justice.

Mischel, W., & Ebbesen, E.B. (1970). Attention in Delay of Gratification. *Journal of Personality and Social Psychology* 16(2): 329–337.

Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., Houts, R., Poulton, R., Roberts, B.W., Ross, S., Sears, M.R., Thomson, W.M., & Caspi, A. (2011): A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America* 108(7): 2693–2698.

Molina, B.S.G., Hinshaw, S.P., Swanson, J.M., Arnold, L.E., Vitiello, B., Jensen, P.S., Epstein, J.N., Hoza, B., Hechtman, L., Abikoff, H.B., Elliott, G.R., Greenhill, L.L., Newcorn, J.H., Wells, K.C., Wigal, T., Gibbons, R.D., Hur, K., & Houck, P.R., and the MTA Cooperative Group (2009). The MTA at 8 years: Prospective Follow-up of Children Treated for Combined-Type ADHD in a Multisite Study. *Journal of the American Academy of Child & Adolescent Psychiatry* 48(5): 484–500.

Muraven, M., & Baumeister, R.F. (2000). Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle? *Psychological Bulletin* 126(2), 247–259.

Persson, I., & Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy* 25(3): 162–177.

Piquero, A.R. (2008). Measuring Self-Control. In E. Goode (Ed.), *Out of Control: Assessing the General Theory of Crime*. Stanford: Stanford University Press, 26–37.

Pratt, T.C., & Cullen, F.T. (2000). The Empirical Status of Gottfredssons and Hirschi's General Theory of Crime: A Meta-Analysis. *Criminology* 38(3): 931–964.

Pratt, T.C., Turner, M.G., & Piquero, A.R. (2004). Parental Socialization and Community Context: A Longitudinal Analysis of the Structural Sources of Low Self-Control. *Journal of Research in Crime and Delinquency* 41(3): 219–243.

Pugh, J., & Douglas, T. (2017). Neurointerventions as criminal rehabilitation: An ethical review. In J. Jacobs & J. Jackson (Eds.), *The Routledge Handbook of Criminal Justice Ethics.* Abingdon: Routledge, 95–109.

Pugh, J., & Douglas, T. (2016). Justifications for Non-Consensual Medical Intervention: From Infectious Disease Control to Criminal Rehabilitation. *Criminal Justice Ethics* 35(3): 205–229.

Rawn, C.D., & Vohs, K.D. (2011). People Use Self-Control to Risk Personal Harm: An Intra-Interpersonal Dilemma. *Personality and Social Psychology Review* 15(3): 267–289.

Ray, K.S. (2016). Not Just "Study Drugs" for the Rich: Stimulants as Moral Tools for Creating Opportunities for Socially Disadvantaged Students. *American Journal of Bioethics* 16(6): 29–38.

Ryberg, J. (2012). Punishment, Pharmacological Treatment, and Early Release. *International Journal of Applied Philosophy* 26(2): 231–244.

Shaw, E. (2013). Cognitive Enhancement and Criminal Behavior. In E. Hildt & A.G. Franke (Eds.), *Cognitive Enhancement: An Interdisciplinary Perspective*. Dordrecht: Springer, 265–282.

Shaw, E. (2014). Direct Brain Interventions and Responsibility Enhancement. *Criminal Law and Philosophy* 8(1): 1–20.

Shaw, E. (2016). The Right to Bodily Integrity and the Rehabilitation of Offenders Through Medical Interventions: A Reply to Thomas Douglas. *Neuroethics*. doi:10.1007/s12152-016-9277-4.

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate Blocks Effort-Induced Depletion of Regulatory Control in Healthy Volunteers. *Psychological Science* 25(6): 1227–1234.

Stevenson, C. (2016). Self-Pathologizing and the Perception of Necessity: Two Major Risks of Providing Stimulants to Educationally Underprivileged Students. *American Journal of Bioethics* 16(6): 54–56.

Storebø, O.J., Krogh, H.B., Ramstad, E., Moreira-Maia, C.R., Holmskov, M., Skoog, M., Nilausen, T.D., Magnusson, F.L., Zwi, M., Gillies, D., Rosendal, S., Groth, C., Rasmussen, K.B., Gauci, D., Kirubakaran, R., Forsbøl, B., Simonsen, E., & Gluud, C. (2015). Methylphenidate for attention-deficit/hyperactivity disorder in children and adolescents: Cochrane systematic review with meta-analyses and trial sequential analyses of randomised clinical trials. *BMJ* 351:h5203.

Tabibnia, G., Monterosso, J.R., Baicy, K., Aron, A.R., Poldrack, R.A., Chakrapani, S., Lee, B., & London, E.D. (2011). Different Forms of Self-Control Share a Neurocognitive Substrate. *Journal of Neuroscience* 31(13): 4805–4810.

Vincent, N. (2009). Responsibility: Distinguishing Virtue from Capacity. *Polish Journal of Philosophy* 3(1): 111–126.

Vincent, N. (2010). Madness, Badness, and Neuroimaging-based responsibility assessments. In M. Freeman (Ed.), *Law and Neuroscience, Current Legal Issues (Vol. 13)*, Oxford: Oxford University Press, 79–95.

Vincent, N. (2014). Restoring Responsibility: Promoting Justice, Therapy and Reform Through Direct Brain Interventions. *Criminal Law and Philosophy* 8(1): 21–42.

Wikström, P.H., & Treiber, K. (2007). The Role of Self-Control in Crime Causation: Beyond Gottfredson & Hirschi's General Theory of Crime. *European Journal of Criminology* 4(2): 237–264.

World Health Organization Regional Office for Europe (2013). *The European Mental Health Action Plan: Regional Committee for Europe Sixty-Third Session.* Copenhagen: World Health Organization Regional Office for Europe.

Zohny, H. (2015). The Myth of Cognitive Enhancement Drugs. *Neuroethics* 8(3): 257–269.

Neuroethics
https://doi.org/10.1007/s12152-020-09439-3

ORIGINAL PAPER

Born which Way? ADHD, Situational Self-Control, and Responsibility

Polaris Koi

Received: 8 December 2019 / Accepted: 15 June 2020
© The Author(s) 2020

Abstract Debates concerning whether Attention Deficit/Hyperactivity Disorder (ADHD) mitigates responsibility often involve recourse to its genetic and neurodevelopmental etiology. For such arguments, individuals with ADHD have diminished self-control, and hence do not fully satisfy the control condition for responsibility, when there is a genetic or neurodevelopmental etiology for this diminished capacity. In this article, I argue that the role of genetic and neurobiological explanations has been overstated in evaluations of responsibility. While ADHD has genetic and neurobiological causes, rather than embrace the essentialistic notion that it directly diminishes self-control and, therefore, responsibility, we ought to think of ADHD as constraining only some self-control practices. In particular, situational self-control strategies remain feasible for people with ADHD. However, not all individuals have *access* to these strategies. I suggest a way to evaluate responsibility in terms of situational rather than agential pleas, which tracks whether the individual had access to self-control behaviors. While I restrict my discussion to ADHD, the access-based approach is also relevant for assessments of responsibility for other cases where self-control failures are at stake.

Keywords ADHD · Responsibility · Moral responsibility · Self-control · Genetic essentialism

P. Koi (✉)
Polaris Koi, University of Turku, Turku, Finland
e-mail: polaris.koi@utu.fi

Introduction

Attention Deficit/Hyperactivity Disorder (ADHD) is often[1] seen as decreasing the degree to which one can be held responsible.[2] A standard conception of moral responsibility holds that it involves a control condition, i.e., that agents can only be held responsible to the extent that they were in control of their behavior. The responsibility of persons with ADHD is mitigated, it seems, because genetic and neurodevelopmental causes have contributed to their having a diminished capacity for self-control, thereby undermining the extent to which these agents are in control of their behavior. As a result, agents with ADHD sometimes act and cognize in ways that they do not fully control.

It seems that individuals with ADHD sometimes fail in responding adequately to reasons simply due to their struggles with self-control. This notion is also embraced by critics of the standard conception, such as Nomy Arpaly [6, 7], who illustrates this with the story of John, an adult with Attention Deficit/Hyperactivity Disorder

[1] For views arguing that ADHD should be a mitigating factor, e.g., in criminal sentencing, See [1]. For research on lay thinking about ADHD and responsibility, see [2–4].

[2] Recent debate has highlighted the ambiguity of the concept of responsibility. In this paper, the focus is on what Shoemaker terms the accountability sense of responsibility [5], which is compatible with both the standard conception of responsibility and reactive attitudes-based accounts. An agent is accountability-responsible for x when she has acted in such a way that it is appropriate for us to blame or praise her for so acting. In what follows, unless stated otherwise, 'responsible' refers to the accountability sort of responsibility, and 'blame' refers to blame warranted by these actions. Accountability-blame is gradated: blame may be mitigated, or the agent may be wholly excused.

Published online: 25 June 2020

(ADHD). John, upon seeing an attractive house from his car, buys it within minutes. In making this purchase, he does not consult his wife, despite that they were not looking to buy a house. His impulsive purchase causes financial distress to his family and is an act on grounds of which we would ordinarily find it appropriate to blame him.

> If John were normal – that is, if we could assume that in John, the same relationship between urge and action holds as in an ordinary person – his behavior, especially his failure to consult his wife, would indicate moral indifference to us. However, John's neurological disorder results in an astounding loss of self-control – however much he wants to, he simply does not have, in some circumstances, the ability to stop before acting, think, and implement his practical conclusion. [6: 22.]

Arpaly then argues that John is not an appropriate target for blame because due to his disorder, the buying of the house is not a sign of moral indifference or ill will. For Arpaly's account, the agent is excused or blame is mitigated due to features of the individual agent – not because they infringe on the control condition, but rather, because they indicate that the agent's behavior may not be an accurate reflection of their quality of will [6: 21]. For John, that feature is his self-control impairment. John has behaved in a reckless manner in making such a major purchase without consulting his wife or the state of family finances. Arpaly suggests that upon learning about "the fact that John has no self-control […] given his organic deficiency of self-control" [7: 152], we accept that John's failure to engage in these activities was not due to moral indifference and conclude that John is not to be held (fully) responsible for his purchase.

Pleas for excusing or mitigating responsibility can be divided into two rough groups. Let us label these *situational pleas* and *agential pleas*. Situational pleas excuse the agent by appeal to a feature of the situation that limited the agent's capability to control their behavior in accordance with our moral demands, such as that they were misinformed about what they were doing, were strong-armed into it, or were tied to a chair. Agential pleas excuse the agent by characterizing the agent as not fully capable of moral behavior, in that the agent is, in Strawson's [8] words, "warped or deranged, neurotic or just a child". In such a case, the person's agency is either not fully developed, or is limited, either globally or locally,[3] by a biological condition beyond the agent's control. The mitigating or excusing factor is not a feature of the situation the agent was in, but rather, a feature of the agent. As long as that feature remains stable, the agent is considered less than fully responsible and may be considered to be outside the moral community.[4]

When genetic and neurobiological explanations apply to an agent's atypical behavior, agential pleas are often invoked.[5] An agential plea is also what Arpaly appears to have in mind for John[6]: due to genes, neurobiology, or other inexorable features of his person (that we associate with his ADHD), John is simply uncapable of self-control in the face of the house he fancies. When Arpaly writes, "His lack of self-control exempts him from blame (if we assume that he has not, for example, knowingly refused to take his medication)" [6: 21], she seems to suggest that due to John's trait – his self-control impairment – we assess him as less than fully responsible, and the genetic and neurobiological explanations of that trait are seen as supporting the idea that his poor self-control is indeed a relatively stable trait of the agent. Medication is seen as 'canceling out' some, but not all, of the behavioral effects of these traits.

[3] Global agential pleas are such as are applied, e.g., to toddlers who are not held to be full agents in the relevant sense with respect to any behavior, and are placed outside the moral community: a helpful discussion of disabled agents who may fall just inside or just outside the moral community is provided by Shoemaker [5]. Local agential pleas are such as Arpaly applies to John: since behavior caused by John's poor self-control is taken to not reflect poorly on his quality of will, John is excused for behavior caused by a failure of self-control, but not for other behavior.

[4] All behavior, of course, is done by individuals in situations. There is no metaphysical bright line between agential and situational pleas. Rather, this is a pragmatic distinction meant to characterize most appeals to mitigate or excuse responsibility.

[5] This phenomenon is further characterized in sections 2 and 3.

[6] In the passage quoted above, Arpaly does qualify John's incapacity to "some circumstances". Is this to be interpreted as suggesting that something like situational pleas is to be intended, after all? I think that an unlikely reading, given the repeated emphasis Arpaly places on her characterization of John as persistently, biologically abnormal. Rather, the qualification can be interpreted as acknowledging that John sometimes succeeds in self-control out of luck. Arpaly [7:151–153] draws an analogy to someone with Tourette's syndrome who says "fuckers!" in a family gathering. Of course, someone with coprolalia (a relatively rare vocal tic, present in a minority of Tourette's patients) will not *always* utter obscenities. But the uttering of obscenities, for Arpaly, is not something the patient can control. The coprolalic patient is not to be blamed for saying "fuckers!" any more than he is to be praised for not doing so, because neither reflects on the patient's quality of will. Likewise, on Arpaly's analysis, for self-control impairments in ADHD.

As I will argue, medication is not the only agent-external factor that significantly contributes to self-control successes and failures in persons with ADHD. In addition to genetic, neurobiological and medical factors, many social, educational, psychological, and environmental phenomena can act as constituents of self-control (or lack thereof). Self-control is a broad construct encompassing a wide range of behaviors, only some of which are made significantly harder by ADHD. Whether the individual's circumstances facilitate self-control for that person will vary on a situational basis. This is doubly important for persons whose self-control is particularly vulnerable to environmental variance, such as persons with ADHD, for whom the management of their disorder involves finding situational strategies to support their self-control and avoiding environmental factors that may undermine it. As a result, I argue that situational pleas ought to be preferred to agential pleas when discussing moral and criminal responsibility for failures of self-control, including in persons with ADHD.

In order to analyze how it is that ADHD impacts one's capacity for self-control and, hence, responsibility, I also discuss two other themes, addressing which is helpful for resolving the question at hand and for making my case in favor of situational pleas. First, what sort of role should explanations involving recourse to the genetic etiology of ADHD play in assessing their responsibility? Second, how should self-control be understood in order to effectively assess whether, and to what degree their responsibility ought to be mitigated? I argue that in the context of responsibility assessments, the genetic and neurodevelopmental etiology of ADHD does not show that agential pleas would be appropriate in the context of self-control impairments. Rather, they highlight a need for an account of self-control that describes it in terms of access to behaviors, where self-control is generated in interaction with the environment. An access-based approach enables assessing and mitigating responsibility in a way that is sensitive to the heterogeneity among people with ADHD and the context-dependence of such assessments diachronically for the same individual. As I will try to show, situational pleas are best suited for this task.

This paper will proceed as follows. An overview of ADHD and the current debate on its impact on responsibility is provided in section 2, while section 3 expands on that debate with particular attention on how our thinking about genes and brains shapes our responses towards ADHD. In section 4, I describe an externalist stance on self-control, arguing that self-control failures are best explained in externalist terms. The heart of the present paper can be found in section 5, which brings together the preceding discussions of ADHD, responsibility, and self-control, describing how it is that individuals with ADHD are at a disadvantage with regard to self-control and how it is that the situation could be amended.

ADHD and Responsibility

ADHD is often described as the most common childhood neurobehavioral disorder. It is typically diagnosed in childhood or adolescence, although the diagnostic criteria have recently been updated to better accommodate diagnosis in adulthood [9]. Persons with ADHD struggle with behavioral and attentional control. They may act impulsively, have trouble actively listening to instructions or feedback, or have trouble planning for the future. As a result, persons with ADHD are often described in normatively loaded terms: as problem children and reckless teenagers [2, 3, 10]. These persons are often held to be less responsible for their misbehavior due to their ADHD. Not only do others excuse individuals on grounds of their ADHD, individuals with ADHD also externalize problematic behaviors to their ADHD, which can help alleviate feelings of guilt or self-blame [11, 12].

DSM-V lists eighteen core symptoms of ADHD which are divided into the subgroups of inattention and hyperactivity-impulsivity. The two symptom dimensions are distinct but correlated. Examples of inattentive symptoms include trouble holding attention on a task, trouble completing a task, forgetfulness in daily activities, and losing items necessary for tasks and activities. Examples of hyperactive-impulsive symptoms include motor fidgeting, excessive talking, trouble waiting for one's turn, and frequent interrupting or intruding on others. [9].

The symptoms of ADHD correspond to impairments in various dimensions of self-control, such as inhibitory control, delayed gratification, and temporal discounting. As a result, Barkley [13] has characterized ADHD as a disorder *of* self-control.[7] For Barkley [13, 14], a neurodevelopmental impairment of the executive

[7] I take for granted that difficulties in self-control are a crucial aspect of ADHD but remain agnostic about whether ADHD is reducible to impaired self-control.

functions causes the persistent self-control difficulties characteristic of ADHD. When the disorder is seen as (neuro)biological and genetic, it is often also seen as an exculpating or mitigating influence on the individual's behavior: we cannot help the way we are born.

ADHD was formerly believed to be a transient childhood disorder, but we presently know that symptoms often persist into adulthood [14]. The most common clinical response to ADHD is stimulant medication, although other medications and behavioral interventions are also used for managing the disorder. Various etiologies for ADHD have been suggested, but most such suggestions, such as exposure to sugar or poor parenting, have not been substantiated. The consensus is that ADHD is a highly heritable neurodevelopmental disorder, and that while the environment has an impact on it, social factors alone cannot bring about ADHD. While the heritability of ADHD has long been estimated to be at roughly 70–80% based on decades of twin studies [15], specific risk genes for ADHD were first identified very recently, in a genome-wide association study published in Nature Genetics [16]. The study found twelve risk loci for ADHD, and furthermore noted that the loci were in concordance not just with whether or not an individual was diagnosed with ADHD but also with a spectrum of subclinical ADHD symptoms, supporting the idea that ADHD and subclinical impairments in self-control and related behaviors lie on a continuum [15, 16].

ADHD is also associated with differences in neurobiology. These differences are connected with impairments in the executive functions [13, 14], as documented using both EF scales and EF tests.[8] The pathophysiology of ADHD[9] is by no means homogenous: each physiological correlate of ADHD may be relevant only to a subset of persons with ADHD. Neuroimaging is not a plausible substitute for clinical interview in diagnosis. However, while the differences in brain anatomy and functioning in persons with ADHD compared to controls are varied, they are consistent with the clinical and behavioral finding that differences in executive functioning are characteristic of ADHD.

The self-control of individuals with ADHD is subject to individual variance. Some individuals with ADHD seem to succeed in tasks, careers and projects that place robust demands on their self-control,[10] lending support to portrayals of ADHD as a strength. Such success stories can be motivational for many people with ADHD and their loved ones, who wonder about how their diagnosis will impact their life. Yet while these success stories highlight that persons with ADHD lead a variety of lives, a one-sided description of ADHD as a strength can feel like erasure for persons whose relationship with their ADHD is an uphill battle.

Many adults with ADHD face considerable difficulty in the educational, occupational, dating and marital and financial spheres [14]. Barkley describes the adverse effects of ADHD on educational achievement as "chronic and cumulative" (ibid., p. 314), and reports a correlation between ADHD and lower work performance (ibid.). Fiscally, adults with ADHD are more likely than controls to face challenges in managing their finances, such as, to engage in impulsive shopping, to have trouble paying rent and bills on time (ibid.) Barkley also draws attention to an aspect of ADHD that has received less attention: its impact on social functioning, such as in the context of relationships. Barkley notes that the impairments in top-down inhibitory control that are a central feature of ADHD also extend into the domain of emotional regulation, and that impaired emotional self-control commonly manifests in "temper outbursts and more unstable personal relationships; and […] difficulty maintaining friendships" (ibid., 327).

For persons facing struggles described above, ADHD may feel like a persistent handicap where their impaired

[8] Executive functioning deficits are characteristic of both children and adults with ADHD, although they are subject to individual variance. Out of the executive functions, poor performance in planning and inhibitory control tasks is common, but not ubiquitous, in subjects with ADHD; deficits in working memory appear more prevalent [14].

[9] Functional neuroimaging studies support the idea that ADHD relates to impaired executive functioning: a meta-analysis relating FMRI studies to the seven large-scale brain circuits found hypoactivation in the frontoparietal network, involved in executive functioning, in both children and adults with ADHD [17]. By contrast, hyperactivation was observed in both children and adults in the default network, which underlies processes that are not focused on the outside world, such as self-referential processes, planning and daydreaming (ibid.) Structural neuroimaging studies have furthermore found ADHD to correlate with differences in gray matter volume in various areas of the brain, including the basal ganglia and occipital lobes; stimulant medication is associated with the normalization of these structural differences (ibid.).

[10] For example, an article in the Forbes magazine argues that ADHD is "an entrepreneur's superpower" [18]. The portrayal is based on the observation that some extremely professionally successful individuals, such as some celebrities, have ADHD, and on the postulation that this success is due to these individuals' symptoms, which in the right context turn to strengths. That some individuals with ADHD outperform others, e.g. in terms of educational achievement, is also noted in the academic literature [11]; it has been suggested that this is representative of compensation for the impairment related to ADHD rather than less severe symptoms.

self-control adversely impacts their capacity to function in multiple areas of life, from work to relationships. Yet while adults with ADHD are more likely than controls to face adversity in all these domains of functioning, some individuals with ADHD experience these challenges to a significantly smaller extent. There appears to be considerable variance in *how*, and *to what extent*, the self-control of persons with ADHD is impaired. An account of how responsibility should be assessed and mitigated in persons with ADHD should be able to account for this variance.

In addition, there is documented variance in the self-reports of individuals with ADHD assessing their *own* capacity for responsibility and to what extent it has been impaired by ADHD (see, e.g., 5, 2, 12]. Individuals with ADHD are often seen as, or experience themselves as, having trouble following their own normative convictions as well as the moral codes of society [1–3, 19]. One aim for medical interventions for ADHD, then, is the restoration of moral agency. Based on research by Ilina Singh [2, 3], there is some indication that this aim is realized: Singh analyses an interview study of 151 children with ADHD in the UK and US, where the children's self-reports suggested that in the children's own experience, "stimulant drugs enable, rather than threaten, moral reasoning in the face of conflicting impulses" [3] – although the overall efficacy of common stimulant medications, such as methylphenidate, continues to be debated [20].

ADHD can result in grave impairments in an individual's self-esteem, social and occupational life [14, 21]. It is furthermore a risk factor for criminal activity, drug and alcohol abuse disorders, anxiety and depression [1, 21]. While appeals to ADHD in courtrooms are uncommon and successful appeals even more so, individuals with ADHD are overrepresented in the criminal justice system [22]. International studies show that up to half of the adult prison population, and two thirds of juvenile offenders, screen positive for childhood ADHD. Many continue to be symptomatic [21].

Genes, Brains, and Agential Pleas

Within academic debates surrounding criminal responsibility, discussions around the moral and criminal responsibility of people with ADHD often propose a neural and genetic etiology and ontology of ADHD. Robert Eme [1], for example, argues that ADHD ought to be treated as a mitigating factor in criminal sentencing, because "the diminished capacity for self-control in ADHD is caused (at least in part) by a dysfunction in the [dual systems]. […] in the case of ADHD […] the dysfunction is due primarily to genetic factors" (ibid., 352–353). That is, ADHD entails a dysfunction on the level of systems neuroscience, a dysfunction that is genetically caused, and is an exculpating or mitigating factor *because* of its biological etiology and ontology: we cannot help the way we are born.[11]

Nita Farahany [23] calls such stances as Eme's 'behavioral morality'. On the behavioral moralist view as described by Farahany, biological causes of behavior are exculpating or mitigating factors for criminal sentencing because they render individual behavior less subject to volition. According to such views, the agent's behavior fails to reflect their volition in ways that matter for ascriptions of responsibility due to being determined, either partially or wholly, by their biology. Behavioral moralist views are, in essence, agential pleas. They involve the stipulation that some features of the individual – in this case, individual biology – prevent that individual from being a fully-fledged moral agent.

However, as Farahany notes, it is very hard to make direct inferences from evidence in brain physiology to evaluations of individual responsibility. Neuroscience and related disciplines work in a way that is often both more basic and more general than the sort of information that could, in practice, inform assessments of the physiology of specific behavioral causation for individual moral agents in the relevant ways [23]. As a result, individual behavior cannot be neatly divided into behaviors resulting from biological causes, such as genes or atypical neural functioning, on one hand and social causes and/or full-fledged agency on the other. Rather, the empirical question in the study of behavior is to find out how biology and the environment together shape individual agency.

In a study by Lebowitz, Rosenthal & Ahn [4], lay responses toward the behavior of children with ADHD were gauged using vignettes that attributed their behavior either to biological or to psychosocial causes. Attribution to biological causes, such as genetics, decreased attributions of both blame and control, and diminished

[11] Eme would likely agree with Arpaly's account of John (although note that Arpaly takes no stance as to whether John's lack of self-control is biologically caused or constituted, simply stating that it is biological in character).

stigmatizing attitudes towards these children. But it also prompted perceptions of ADHD as relatively untreatable. This effect has been called 'the double-edged sword' and has also been noticed in other contexts assessing genetic explanations of agency. For example, genetic accounts of criminality reduce perception of the perpetrator's control and ascriptions of blame, but they also prompt perceptions of this behavior as characteristic and expectations that the perpetrator will reoffend [24–26]. In short, genetic explanations result in perceptions of the moral responsibility of persons with ADHD on agential rather than situational terms.

This double-edged sword is often termed 'genetic essentialism'. The term 'genetic essentialism' refers to a set of cognitive biases related to interpreting genetic information about the causation of behavior. Dar-Nimrod & Heine describe the phenomenon as follows:

> Learning about genetic attributions for various human conditions leads to a particular set of thoughts regarding those conditions: they are more likely to be perceived as (a) immutable and determined, (b) having a specific etiology, (c) homogeneous and discrete, and (d) natural, which can lead to the naturalistic fallacy. [27: 800.]

All these notions demonstrably arise in response to information about the genetic causes of specific behaviors. Dar-Nimrod & Heine use the concept of *essence placeholder* to describe lay attitudes towards genetics: genes are assigned placeholder status with respect to a person's immutable core traits. In this view, genetic determinism – the view that genes direct behavior in a way that undermines genuine agency – is an aspect of genetic essentialism [27].[12]

The reality of genetic causation of behavior is far from the stark effect that postulated genetic explanations have on lay ideas of responsibility. Some traits are indeed produced by specific, discrete genes. This is true, e.g., for some genetic diseases. However, the vast majority of phenotypes are produced by multiple genetic and gene-environment influences [19, 28]. Following behavior geneticist Eric Turkheimer [19], the prospects of any project matching behavioral phenotypes with discrete genes are gloomy. While behavioral traits, in general, are heritable – some more so than others – the relationship of behavioral phenotypes to genes is many to many, not one to one. Likewise, the causal pathways leading to these phenotypes involve complex gene-gene and gene-environment interactions in addition to environmental causes. As Dar-Nimrod & Heine [27] remark, the more complex the genotype-phenotype relationship, the less rational a genetic essentialist response appears. As a result, despite the widespread association of genetic explanations with agential pleas, agential pleas are not supported by empirical evidence in behavioral genetics.

Finally, a tendency towards agential pleas has also been documented in qualitative research. Honkasilta, Vehmas & Vehkakoski [12] have documented that ambivalence is present in the narratives of individuals with ADHD about their moral agency: individuals with ADHD sometimes use the label to externalize or excuse misbehavior whereas sometimes, they assert that ADHD does not hinder their agency or self-control, committing themselves to norms of behavior and holding themselves responsible for any failures to follow those norms (ibid.).

This ambivalence reflects a difficulty in agential pleas: while they may enable us to excuse agents in situations where they ought to be excused, they also carry with themselves the baggage of describing the agent as persistently incapable of full-fledged agency. Because ADHD is persistent, any associated diminishment of self-control is often assumed persistent. The qualitative research reflects that agents who may have acted in wayward ways due to self-control failures that are associated with their ADHD may prefer assuming responsibility to describing themselves as less than full moral agents. In these dilemmas, the third option – the situational plea – appears obscured by a focus on the persistence of the agent's ADHD. I hypothesize that these agents may assume that when behavior is explained by reference to a persistent trait, to admit to a failure of control is to admit to a persistent failure of control. Agential pleas may thus seem as if they were the only sort of plea available. Yet if self-control (or lack thereof) is subject to situational variance, situational pleas regarding moral and criminal responsibility become a salient option.

Consider responsibility mitigation for agents like John, bearing in mind that the self-control difficulties experienced by people with ADHD are variable not just from one individual to another but also diachronically within the same person, from one situation to another. Because his impulsive real estate purchase seems

[12] See [26] for a view distinguishing the effects of genetic determinism from those of genetic essentialism.

extreme to us, the example may lead us unduly towards thinking that John is entirely uncapable of self-control. Yet it is plausible that John frequently does things that most would find rather more demanding on their self-control: he might, e.g., mow the lawn on a hot day, refrain from shouting at his children when he finds them sneaking to the cookie jar way past bedtime, and hold his tongue when his boss assigns him to a client he despises. An agential plea based on John's real estate impulsivity runs the risk of not assigning John due responsibility for the many other things he does. Yet an assessment of John's responsibility scaled to his success in other tasks would run the risk of also holding him fully responsible for his self-control failure in the real estate purchase. Situational pleas would be best suited to properly dealing with this heterogeneity.

In summary, neither neuroscience nor behavior genetics support agential pleas over situational pleas. While both are powerful methods of scientific inquiry, their applicability to the assessment of individual epistemic or moral responsibility is limited. Ideas about the immutability of self-control in persons with ADHD may stem from genetic essentialist and behavioral moralist thinking that is at best a misunderstanding of the science. This is not to dispute that disorders like ADHD are grounds for special considerations when it comes to self-control. But as I will argue below, because our social and environmental arrangements currently do an erratic job at enabling people to enact self-control, those considerations are better employed as situational pleas.

The Environment in Self-Control

Due to its genetic etiology, ADHD has been seen as a fairly immutable trait, although symptoms often decrease in severity as the subject matures [14]. Ascriptions of responsibility for people with ADHD have been attached to their ADHD. An agential plea, such as given by the behavioral moralist, would scale the mitigation of responsibility to the degree of the relevant impairment. In such an account, the more severe a person's ADHD, and the more their self-control is compromised, the more their responsibility is mitigated. For the behavioral moralist, successful medication decreases the severity of ADHD, restoring responsibility.

Situational pleas present themselves as a viable alternative to agential pleas in the case of ADHD if we remark how our responsibility assessments do not seem to be mitigated just with respect to diagnoses, genetics, or persistent neurobiological traits. They seem to be mitigated (or aggravated) also with respect to the way the environment constrains and scaffolds agency, such as with respect to how these traits are managed. Yet allowing that properly managing an innate trait, e.g., with medication, is relevant for assessments of responsibility is only part of the equation.

Arpaly is right that insofar as responsibility is compromised in particular ways for persons with ADHD, it is due to the difficulties they have with self-control. Self-control, however, is a heterogeneous construct under much current debate in both philosophy and the empirical sciences. In philosophy as well as in ordinary parlance, self-control refers, broadly, to the capacity to align behavior with intentions and better judgment. Within psychology and related disciplines, it is operationalized in various ways, most commonly as trait self-control, inhibitory control, delayed gratification, and temporal discounting. These constructs are distinct yet interrelated and consistent [29, 30]. The heterogeneity of various processes and behaviors associated with self-control suggests that self-control is not a natural kind [31].

The idea that self-control is not fully agent-internal is not novel to the present paper. Neil Levy [32, 33] has persuasively argued that rather than being a display of an innate capacity for inhibitory control, self-control makes use of various environmental practices. For example, someone striving to diet avoids the street where the bakery is, thus keeping clear from the delicious smell of baked goods. In Levy's analysis, multiple experiments have demonstrated that persons high in trait self-control or in delayed gratification do not necessarily employ inhibitory control, but rather, either avoid environments where behavioral inhibition is required or modulate their interaction with those environments. An example of such research is Mischel's [34] famous 'marshmallow test'. In the experiment, children were instructed to wait for the researcher in a room with an edible treat, such as a marshmallow, without consuming it; if they succeeded, they got two treats instead of one. In this experiment (and subsequent variants), the children who succeeded in not eating the marshmallow appeared to do so not by means of iron inhibition but rather by directing their attention elsewhere, such as by singing, turning their back to the marshmallow, imagining the marshmallow is a tiny cloud, or attempting to fall asleep. Mischel and colleagues [35] furthermore

found that these heterogeneous behaviors that helped children succeed in not eating the marshmallow are teachable.

Mischel et al.'s research suggests that the role of the environment in the production of self-control is not merely something where the individual takes a passive role. Individuals consciously shape and choose their surroundings, and how they relate to these surroundings, in order to be more self-controlled. Duckworth, Gendler & Gross [36] argue that situational strategies of self-control are not just a complement to what they term intrapsychic strategies of self-control: based on their research, situational strategies are the more potent means for self-control because they can be used to diffuse an oncoming impulse before the impulse gets too strong.

Not only does the environment figure in situational strategies of self-control, intrapsychic strategies, such as rehearsing one's reasons for doing what one judges best, or construing of the marshmallow as a tiny cloud, often also depend on the environment in various ways. For example, intrapsychic strategies can be learned and taught, and they are subject to interference from external distractions.

In addition to strategies of self-control, self-control can manifest in seemingly mechanistic ways. For example, an agent may be presented with an impulse, e.g., to yell at a child, but simply inhibit that impulse without (conscious) recourse to any internal strategy. As a result, there is a wide range of behaviors, including mechanisms as well as strategies and practices, that we label self-control because of their function rather than their structure.

Which behavior or behaviors of self-control are employed in a given instance of self-control depends on the individual at hand and their immediate and past environment. The individual's genes, neurobiology, and personal interests and preferences play a role in what strategy or mechanism of self-control to employ. Past environment has a shaping role in what sorts of self-control practices and behaviors the individual has learned; the immediate environment regulates what sorts of self-control practices are currently available. For example, a prisoner is unable to simply walk away when a fellow inmate begins to heckle them, removing one self-control behavior that is available to most others upon noticing an impulse towards aggressive behavior.

On the other hand, environments can enable behaviors of self-control otherwise unavailable. Ulysses arrangements, where an agent enables refraining from a given activity by physically preventing it are a good example of practices of self-control that environments facilitate. For example, a student wishing to stay off social media in order to get a paper written might install an application that prevents them from accessing certain websites and applications for a given period. In an environment where such applications would be in standard usage much like the most common word processors are, the environment would not just enable but foster and encourage such self-control behaviors. Environments can support or undermine responsibility by virtue of their impact on self-control.

The heterogeneity of self-control behaviors shows that self-control is multiply realizable. Persons who have aggressive impulses upon being heckled refrain from aggressive behavior and procrastinators coax themselves to write papers all the time, but how they accomplish such instances of self-control varies tremendously, ranging from simply inhibiting conflicting impulses without much thought to it, to rehearsing one's reasons to act in a preferred way, to walking away or establishing Ulysses arrangements. In assessing the responsibility of persons with ADHD, our question then becomes whether, and in what sorts of situations, persons with ADHD can realize self-control behaviors. In the following section, I will argue ADHD does not directly decrease self-control. Rather, due to it, certain self-control behaviors may become impossible or difficult to enact. However, due to the tremendous heterogeneity of self-control behaviors, other strategies remain available to persons with ADHD if those can be accessed in a given situation. Compare this idea to another capability, the capability to make music. Various illnesses and physical trauma may impair an individual's ability to sing, but that does not mean they would be prevented from making music as they could also play an instrument – at least if the individual can access instruments and instruction, is not prevented from practicing, and so forth. While these two capabilities are not perfectly analogous, they are alike in that both are subject to multiple realizability and situation-sensitivity. Due to the multiple realizability and situation-sensitivity of self-control, an inference directly from the presence of a genetic condition such as ADHD to responsibility mitigation is mistaken. Rather, responsibility ought to be mitigated or the agent excused only in situations where self-control behaviors were not accessible for the agent.

The reader may here object that given the heritability of ADHD, even if situational strategies are an important aspect of self-control, they are not what is at stake for persons with ADHD: persons with ADHD have impaired self-control because of their genes or because of their neurobiology, not because of their environment. After all, they share their environments with persons who do not have impaired self-control. This objection merits a response before proceeding.

It is true that the robust role of the environment in self-control behaviors may appear contradictory to what we know about disorders like ADHD. The strong heritability of ADHD gives rise to various intuitions about it being immutable, and it is deceptively easy to extend the same immutability to self-control impairments associated with ADHD. However, these intuitions do not match what we know about the genetic causation of complex behavior. Instead, findings in behavior genetics support thinking about self-control in persons with ADHD as involving environmental causes and as including situational strategies of self-control.

Decades of twin studies have produced heritability estimates for ADHD at roughly 70–80% [15]. that entails that 20–30% of variance is explained by gene-independent environmental factors. Heritability estimates however include gene-environment interactions. Gene-environment interactions are often thought of in passive terms [37]: a genotype, when placed in an environment, produces a trait in a way that responds to that environment. However, gene by environment interactions (GxE for short) can also be active. Passive gene-environment interaction refers to how genes code for how the individual organism responds to specific environmental influences, but individuals also actively seek out and create environments that are a "good fit" for their genotype, which is referred to as active GxE. In other words, while GxE interactions encompass such 'passive' interactions as epigenetic transmission in utero, some such interactions are not beyond the individual's control. Agents can select environments that foster desirable behaviors. GxE interactions are an important form of genetic causation, and active GxE interactions are a form that agents participate in. Even as genes code for our traits, the way we select, shape and navigate our environments also has a role in the production of those traits [37, 38].

The genotype of persons with ADHD codes for differences in their self-control [15]. But that should not be interpreted in essentialist terms: the evidence does not indicate that self-control, in people with ADHD or otherwise, would be any more inherent or immutable than any other complex heritable trait is. Above, we saw that the environment has a robust role in the constitution of self-control. This is also true for self-control in persons with ADHD. The complex etiology of self-control impairments in ADHD does not foreclose externalism about self-control. On the contrary, findings in behavioral genetics about the genetic production of complex behavioral phenotypes lend support to highlighting the role of the environment in it. Of course, the behavior of people with ADHD is partly explainable by their genes and brains, in addition to their environments. But since this is a trivial fact in that it pertains to *all* human behavior, it is not particularly useful for probing whether and how people with ADHD ought to be held responsible.

Situational pleas are appropriate for self-control failures in ADHD because even as ADHD may be persistent, self-control failures are not a stable feature of the agent. Rather, as discussed in more detail in section 5, agents with ADHD are perfectly capable of a wide range of self-control practices, unless those practices are prevented by a situation-dependent lack of feasibility or awareness. In assessing whether the agent ought to be excused for her self-control failure, we ultimately need to assess whether self-control behaviors were accessible for the agent in that situation. While genetic and neurobiological information can help highlight how not all self-control strategies are equally effective for all agents, taken alone it is of limited value in assessing to what extent agents with ADHD ought to be held responsible.

Is Self-Control Accessible for Persons with ADHD?

Above, we have seen that the environment has a robust role in self-control behaviors, but the argument for why that entails that pleas for mitigating moral responsibility for persons with ADHD should be situational pleas remains to be made. In a nutshell, that argument is as follows: persons with ADHD have impoverished access to self-control behaviors. This is because an environment that adequately fosters self-control for persons without the disorder does not foster it to the same extent for persons with ADHD. Not only may an environment that fosters self-control for someone without ADHD fail to do so for someone with ADHD, an environment that is harmless for the self-control of someone without

ADHD may severely undermine the self-control of someone with the disorder. For this reason, situational pleas are available for persons with ADHD even in contexts where they are not available for persons without the disorder.

As discussed above, there is convergent evidence that persons with ADHD have different executive functioning when compared to controls. Due to these differences, persons with ADHD are more likely to have difficulty enacting behaviors strongly reliant on executive functioning. Many self-control behaviors, such as that of inhibiting a prepotent impulse by intrapsychic or mechanistic means, are prime examples of behaviors firmly dependent on the executive functions. However, it must be borne in mind that many self-control behaviors crucially involve the environment.

Interactions between the genetics and neurobiology of persons with ADHD and their environments in the production of self-control behaviors take varied forms. For persons with ADHD, their different neurobiology yields special considerations as to what sort of self-control behaviors they can enact. Explaining differences in self-control in terms of access to self-control may help solve the mystery of heterogeneity in the extent to which individuals with ADHD succeed in self-control. Based on the above empirical evidence, it is highly plausible that some individuals with ADHD have better access to self-control behaviors than others; and it is also plausible that it is easier to access such behaviors in some situations than it is in others.

Environmental factors that have been used to foster self-control in persons with ADHD include environmental self-control arrangements, such as arranging for frequent feedback, increased monitoring, and Ulysses arrangements where the action to be avoided is made physically unavailable, and medication. However, not all individuals live in environments that they can readily structure in order to use them for self-control purposes. Feedback and monitoring are likewise not available for everyone or everywhere. This uneven availability also pertains to medication, which is not always available to individuals with ADHD, whether due to the medication not being suited to this individual due to side effects, financial reasons, a parental decision not to medicate, or a lack of information. Individuals with ADHD vary greatly in terms of to what extent they can access any of the abovementioned sorts of environmental self-control support: for example, Ulysses arrangements may be easier to arrange when there is financial flexibility. Furthermore, the extent to which individuals with ADHD can access such environmental self-control supports varies from one situation to another. For example, Ulysses arrangements may be more available at home than they are at school or workplace contexts.

In the absence of available self-control behaviors, individuals with ADHD may act in ways that result in prudential and moral failures. They may act impulsively, or fail to implement their chosen course of action and procrastinate. In assessing whether the individual ought to be held fully responsible for these failures, or whether excusing the individual or mitigating responsibility is appropriate, we ought to ask whether the individual had access to self-control behaviors. Arpaly [6: 23–24] gestures towards a similar concern. She remarks that while we may expect agents to take a deep breath and count to ten rather than give in to an outburst of rage, or to delete the games folder on their computers rather than give in to procrastination, sometimes such "'counting to ten' measures" – Arpaly's phrase for self-control strategies – are not available. The access approach to self-control within this paper can be taken as building on that remark.

For a self-control behavior to be accessible for an individual, requires, at minimum, that the behavior is feasible and that the individual is aware of the availability and efficacy of this behavior. The latter, epistemic, constraint to self-control seems commonplace: information about the various ways in which self-control can be enacted is rarely available.

Despite the growing attention given to externalist strategies for self-control in the scientific literature, discussed above, these are often not being addressed in teaching self-control, whether at home or school. Both children and adults are often simply advised to "just (not) do it". Consider Laura, a little girl, who hits her younger brother. When her father scolds her for it, Laura breaks into tears, saying that she knows it's wrong to hit someone, but the brother was so annoying, she just couldn't help herself. "Even if it's hard, you need to control yourself", the father responds. This furnishes the individual with a further impetus to self-control in the form of social pressure yet is unhelpful in helping Laura understand *how* to enact self-control.

Situations like Laura's are a challenge for all children, yet children are typically expected to, over time, come up with their own behavioral strategies for self-control. For children with ADHD, this process can be harder, take a longer time, and have less reliable results.

Furthermore, situations where there is a social expectation for self-control success but no practical support is available can be considerably stressful for individuals with the disorder.

Some common approaches in education provide advice on how to succeed in self-control, but that advice is not always equally suited for everyone. For example, advice to students (and professors) struggling to write a paper often begins with the sage advice of (figuratively) gluing one's behind to a chair. While such advice can be helpful for some, for individuals with ADHD, advice focused on staying still and focused for prolonged periods of time can be markedly hard to follow. When alternative strategies are not taught, persons with ADHD are set at a disadvantage in terms of having salient strategies for self-control that they are able to enact.

In brief, even as our knowledge of how self-control behaviors are enacted is growing, it is not reflected in our practices as educators.[13] While this impacts the quality of teaching self-control for all students, it has especially undesirable consequences for students with ADHD who may be less able to enact self-control behaviors without such overt advice.

In short, the way we teach self-control is often unhelpful and is it especially ill-suited to persons with disorders such as ADHD. As a result, in addition to barriers to self-control from the unavailability of self-control practices, epistemic barriers arise when individuals with ADHD are not aware of such self-control practices that would be suited for them. *Access* to self-control practices hinges on *awareness and availability* of such practices that one can enact.

Thinking of self-control in terms of access likens access to self-control behaviors to physical access. A wheelchair user is not inherently less able to get around, but the presence of environmental barriers significantly modulates their access to some spaces, and it modulates that in ways that do not always prevent access to some spaces for all people. For example, for a wheelchair user to access a meeting room on the 3rd floor, the building needs to have ramps and/or elevators, and the individual needs to be aware of their existence and of their location. Similarly for self-control in individuals with ADHD: while persons with ADHD are capable of many self-control strategies and behaviors, there are significant environmental barriers that modulate whether a person with ADHD can enact self-control. In some situations, the epistemically and physically available self-control practices may not be suited for the different neural phenotype of persons with ADHD, resulting in an interaction with the environment that forecloses rather than enables self-control behaviors. By contrast, in other situations, individuals with ADHD may be either presented with self-control behaviors they can enact, or they may be able to actively change their environment in order to enable such behaviors. Much, if not all, of the individual variance in the self-control functioning of persons with ADHD can be explained by reference to active and passive environmental interactions.

Due to the genetic, neurobiological and behavioral heterogeneity of the disorder, we can expect there to be some variance in what environmental features best foster self-control for each individual with ADHD. Based on what we know about the range of biological and behavioral features of ADHD, however, it can be postulated that strategies relying heavily on the executive functions, such as on working memory and on inhibitory control mechanisms, are less successful for persons with ADHD. Environmental, situational strategies should be emphasized instead. Access to self-control for persons with ADHD can be fostered by spreading awareness of and teaching environmental self-control strategies, and by encouraging environmental self-control practices in spaces where they are needed but presently often not allowed, such as schools and workplaces. These measures are useful beyond people with ADHD, as they also furnish people who have a hard time using executive functioning-heavy self-control strategies due to e.g. a poor night's sleep, a long day at work, or due to being tipsy, with strategies they too can enact; and, furthermore, they furnish everyone with methods of self-control that Duckworth, Gendler and Gross [36] suggest are more efficient than their executive functioning-heavy, intramental counterparts.

[13] Information on how self-control is taught at school settings is scarce, even as educators often wish to impart self-control on their students. I consulted a textbook for educators wishing to teach their students beneficial character traits such as self-control and grit [39]. The textbook suggested mindfulness practice, giving students opportunities for making choices and facing the consequences, having students set goals and monitor their progress on those goals, and asking peers to remind each other to focus, each of these being recommended as a practice that is effective in teaching self-control (ibid.). These are all laudable practices, but they vary considerably in terms of the extent to which they effectively support the development of self-control. Notably, few of these methods make overt use of the environment, and none involve practical strategies in how to avoid or inhibit unwanted impulses.

Conclusion: Situational Pleas Are best Suited for Failures of Self-Control

In the beginning of this paper, I described lay and professional views that consider the genetic etiology of ADHD, and its status as a neurobiological disorder, to inform responsibility assessments in a way that appears to support agential pleas. However, when we begin to appreciate the robust role of the environment for self-control and the complexity of its biological causation, agential pleas turn out to be unhelpful.

I have argued that situational pleas are most appropriate for cases where responsibility is to be mitigated or the agent excused due to failures in self-control. Pleas to mitigate or excuse responsibility for agents with ADHD ought to be similar in structure to pleas that say the individual was strong-armed into it or was tied to a chair, rather than similar in structure to pleas that say the individual is partially outside the moral community due to their innate traits. This is because the trait on grounds of which epistemic and moral responsibility would be mitigated for individuals with ADHD is poor self-control. But the self-control in persons with ADHD is not inherently or immutably poor, nor is their self-control difficulty any more 'biological' or 'organic' than that of any agent. Rather, they are at a disadvantage in accessing self-control behaviors. It is only in conjunction with environmental barriers to self-control that ADHD thwarts self-control success: the disorder itself is not sufficient for success in self-control to be prevented. Therefore, situational pleas have more explanatory force: accounting for responsibility in people with ADHD based on whether they have situational access to self-control practices explains heterogeneity in capacity responsibility between various individuals with ADHD as well as diachronic variation in self-control for the same individual. In determining whether responsibility ought to be mitigated on grounds of self-control failures, the presence of epistemic and practical barriers to self-control behaviors should be assessed. It should be determined whether the individual had full access to self-control in that situation.

Individuals with ADHD may be born with an inherited risk towards ADHD, but they are not born less capable of self-control nor less morally responsible. If the access theory of self-control is correct, their genotype simply modulates which self-control practices they can access – just as it does for people without the disorder. If there are no barriers to individuals' access to fitting self-control practices, ADHD need not impair self-control for that individual even if it does impact their executive functioning.

Self-control – the aligning of one's behavior with one's intentions and better judgment – is fundamentally a set of practices, not an innate trait. Innate traits, as well as environmental supports and constraints, modulate access to these practices. When assessing whether responsibility ought to be mitigated, instead of looking merely into the genotype or the brain, we ought to look into whether the individual had access to self-control practices.

While this article has focused on the case of ADHD, there is good reason to believe that the argument generalizes to any case where pleas for responsibility mitigation or exculpation are made on grounds of self-control failures. In such cases, we ought to direct our attention to whether the individual had access to self-control practices, keeping in mind that individual differences, such as neurodivergence, may modulate *which* self-control practices are feasible for the agent.

Returning now to Arpaly's [6, 7] description of John, I think John's failure in self-control is not just due to his neurological disorder. It is also due to environmental factors. Many such environmental factors are ones John can actively impact: unless epistemic or other barriers prevent it, John can set in place environmental supports and constraints that enable self-control, including Ulysses arrangements. If John indeed cannot enact self-control upon coming across the beautiful house, it is not because of "the fact that John has no self-control […] given his organic deficiency of self-control", as Arpaly puts it [7: 152]. Rather, it is because situational factors prevent him from enacting such practices that constitute that ability, such as Ulysses arrangements and other environmental strategies. This matters not only because it might turn out that John, despite his ADHD, could have indeed enacted self-control; it also matters because if John indeed was not able to enact any such strategies, that is plausibly caused in part by our societal failure to effectively and inclusively promote self-control.

Put briefly, we do not respond differently to John because of any persistent trait of his, such as his ADHD, but because we assume there were no self-control practices he could have enacted in that situation. To assume that the former necessarily leads to the latter is to tread frightfully close to essentialism, and to overlook the importance of social and environmental structures for

awareness and feasibility of self-control practices. Whether or not John had access to such practices should ultimately be the deciding factor in holding John responsible.

Acknowledgments This research has been funded by the John Templeton Foundation (via a Genetics and Human Agency research project Junior Investigator award) and by the Finnish Cultural Foundation. The author would like to thank participants in the Genetics and Human Agency research project, Kaisa Kärki, Juha Räikkä, and Jukka Varelius, as well as audiences at the European Society for Philosophy and Psychology and Association for the Advancement of Philosophy and Psychiatry annual meetings for helpful comments on earlier drafts of this paper.

Funding Information Open access funding provided by University of Turku (UTU) including Turku University Central Hospital.

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

References

1. Eme, R. 2016. The neuroscience of ADHD, the paradigmatic disorder of self-control. *Journal of Law and the Biosciences* 3 (2): 350–354.
2. Singh, I. 2011. A disorder of anger and aggression: Children's perspectives on attention deficit/hyperactivity disorder in the UK. *Social Science and Medicine* 73: 889–896.
3. Singh, I. 2013. Not robots: children's perspectives on authenticity, moral agency and stimulant drug treatments. *Journal of Medical Ethics* 39: 359–366.
4. Lebowitz, Rosenthal, and Ahn. 2016. Effects of biological versus psychosocial explanations on stigmatization of children with ADHD. *Journal of Attention Disorders* 20 (3): 240–250.
5. Shoemaker, D. 2009. Responsibility and disability. *Metaphilosophy* 40 (3–4): 438–461.
6. Arpaly, Nomy. 2006. *Merit, meaning, and human bondage: An essay on free will*. Princeton: Princeton University Press.
7. Arpaly, Nomy. 2003. *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.
8. Strawson, P. 1962. Freedom and resentment. In *Proceedings of the British Academy, volume 48: 1962*, ed. G. Watson. Oxford: Oxford University Press.
9. American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing.
10. Horton-Salway, M. 2011. Repertoires of ADHD in UK newspaper media. *Health* 15 (5): 533–549.
11. Fleischmann, A., and E.C. Miller. 2013. Online narratives by adults with ADHD who were diagnosed in adulthood. *Learning Disability Quarterly* 36 (1): 47–60.
12. Honkasilta, J., S. Vehmas, and T. Vehkakoski. 2016. Self-pathologizing, self-condemning, self-liberating: Youths' accounts of their ADHD-related behavior. *Social Science & Medicine* 150: 248–255.
13. Barkley, R.A. 1997. *ADHD and the nature of self-control*. New York: Guilford Press.
14. Barkley, R.A. 2015. *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. 4th ed. The Guilford Press.
15. Faraone, S.V., and H. Larsson. 2019. Genetics of attention deficit hyperactivity disorder. *Molecular Psychiatry* 34 (4): 562–575.
16. Demontis, D., R.K. Walters, […] B. M. Neale (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics 51: 63–75.
17. Cortese, S., and F.X. Castellanos. 2015. Attention deficit/hyperactivity disorder. In *Neurobiology of brain disorders: Biological basis of neurological and psychiatric disorders*, ed. M.J. Zigmond, L.P. Rowland, and J.T. Coyle. Cambridge, MA: Academic Press.
18. Archer, D. 2014. *May 14*. ADHD: The Entrepreneur's Superpower. Forbes Magazine Retrieved from https://www.forbes.com/sites/dalearcher/2014/05/14/adhd-the-entrepreneurs-superpower/.
19. Turkheimer, E. 2000. Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science* 9 (5): 160–164.
20. Storebø, O.J., H.B. Krogh, E. Ramstad, C.R. Moreira-Maia, M. Holmskov, M. Skoog, T.D. Nilausen, F.L. Magnusson, M. Zwi, D. Gillies, S. Rosendal, C. Groth, K.B. Rasmussen, D. Gauci, R. Kirubakaran, B. Forsbøl, E. Simonsen, and C. Gluud. 2015. Methylphenidate for attention-deficit/ hyperactivity disorder in children and adolescents: Cochrane systematic review with meta-analyses and trial sequential analyses of randomised clinical trials. *BMJ* 351: h5203.
21. Young, S., and J. Thome. 2011. ADHD and offenders. *The World Journal of Biological Psychiatry* 12 (S1): 126–130.
22. Koi, P., S. Uusitalo, and J. Tuominen. 2018. Self-control in responsibility enhancement and criminal rehabilitation. *Criminal Law and Philosophy* 12 (2): 227–244.
23. Farahany, N. 2012. Law and behavioral morality. In *Evolution and morality: Nomos LII*, ed. J.E. Fleming and S. Levinson. New York: NYU Press.
24. Cheung, B.Y., and S.J. Heine. 2015. The double-edged sword of genetic accounts of criminality: Causal attributions from genetic ascriptions affect legal decision making. *Personality and Social Psychology Bulletin* 41 (12): 1723–1738.

25. Appelbaum, P.S. 2014. The double helix takes the witness stand: Behavioral and neuropsychiatric genetics in court. *Neuron* 82: 946–949.
26. Tabb, K., M.S. Lebowitz, and P.S. Appelbaum. 2019. Behavioral genetics and attributions of moral responsibility. *Behavior Genetics* 49: 128–135.
27. Dar-Nimrod, I., and S.J. Heine. 2011. Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin* 137 (5): 800–818.
28. Schaffner, K. 2016. *Behaving: What's genetic, What's not, and why should we care?* Oxford: Oxford University Press.
29. Duckworth, A., and M.L. Kern. 2011. A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* 45 (3): 259–268.
30. Kotabe, H., and W. Hofmann. 2015. On integrating the components of self-control. *Perspectives on Psychological Science* 10 (5): 618–638.
31. Herdova, M. 2017. Self-control and mechanisms of behavior: Why self-control is not a natural mental kind. *Philosophical Psychology* 30 (6): 731–762.
32. Levy, N. 2011. Resisting weakness of will. *Philosophy and Phenomenological Research* 82 (1): 134–155.
33. Levy, N. 2017. Of marshmallows and moderation. In *Moral psychology volume 5: Virtue and character*, ed. W. Sinnott-Armstrong and C.B. Miller. Cambridge, MA: MIT Press.
34. Mischel, W., and E.B. Ebbesen. 1970. Attention in delay of gratification. *Journal of Personality and Social Psychology* 16 (2): 329–337.
35. Mischel, W., O. Ayduk, M.G. Berman, B.J. Casey, I.H. Gotlib, J. Jonides, E. Kross, T. Teslovich, N.L. Wilson, V. Zayas, and Y. Shoda. 2011. 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience* 6 (2): 252–256.
36. Duckworth, A., T.S. Gendler, and J.J. Gross. 2016. Situational strategies for self-control. *Perspectives on Psychological Science* 11 (1): 35–55.
37. Kendler, K.S., and J.H. Baker. 2007. Genetic influences on measures of the environment: A systematic review. *Psychological Medicine* 37 (5): 615–626.
38. Jaffee, S.R., and T.S. Price. 2008. Genotype-environment correlations: Implications for determining the relationship between environmental exposures and psychiatric illness. *Psychiatry* 7 (12): 496–499.
39. Hoerr, T.R. 2017. *The formative five: Fostering grit, empathy, and other success skills every student needs*. Alexandria, VA: ASCD.

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

Polaris Koi

Accessing Self-Control

Abstract
Self-control is that which is enacted to align our behaviour with intentions, motives, or better judgment in the face of conflicting impulses of motives. In this paper, I ask, what explains interpersonal differences in self-control? After defending a functionalist conception of self-control, I argue that differences in self-control are analogous to differences in mobility: they are modulated by inherent traits and environmental supports and constraints in interaction. This joint effect of individual (neuro)biology and environmental factors is best understood in terms of access to self-control behaviours. I sketch an account of access as including the three criteria of means, awareness, and non-excessive effort. I further demonstrate that people with disorders such as ADHD have limited access to self-control behaviours and stand therefore at a disadvantage with regard to self-control.

Keywords: self-control; mental disorder; accessibility; externalism; functionalism; disability

1. Introduction

Self-control is that which is enacted[1] to align our behaviour with a motivational commitment, such as an intention or better judgment, in the face of conflicting impulses of motives. It is an umbrella concept, subject to various measures that are distinct but correlated (Duckworth & Kern 2011). We enact self-control, e.g., to quit intrusive habits and begin new ones, to put down a novel to do the dishes, and to refrain from shouting at a loved one when angry.

Some people seem to have more self-control than others: for example, Miguel seems to have an easier time aligning his actions with his better judgment than Rafael, who often finds himself doing something he would not endorse, such as shouting at his mother or putting off doing something he intends to do. Measures such as the trait self-control scale have been developed in psychology to study the various correlates of individual variation in self-control[2]. Yet Miguel, too, sometimes fails in self-control: for example, while he usually refrains from beef to protect the environment, he sometimes succumbs to the temptation of a freshly grilled steak. Vice versa, it is not unusual for Rafael to succeed in self-control.

---

[1] This choice of phrasing is not intended as a commitment to enactivism but rather is used in the sense of brought to pass, effected, executed, deployed, engaged in.

[2] The trait self-control scale (Tangney, Baumeister & Boone 2004) predicts success in multiple domains of self-control. It is correlated with but distinct from other measures, such as various measures of inhibitory control and delayed gratification (Duckworth & Kern 2011). See Levy (2017) for an account of trait self-control as crucially involving environmental practices.

Often, successes and failures in self-control are seen as personal successes and failures. We commend Miguel for his self-controlled character. We likewise think Rafael's persistent troubles in self-control are a personal failure of his and treat his successes in self-control as evidence that had he flexed his 'mental muscle'[3] a bit more, he could have done better at other occasions, as well. I argue that this is mistaken. Rather, self-control depends on whether one has access[4] to such behaviours that serve the function of self-control. There is interpersonal variation in the extent to which various processes and strategies are feasible to us, and in the extent to which we are aware of these forms of self-control. When an agent lacks means M to enact self-control, is not aware of M, or when M requires excessive effort, that agent does not have access to self-control. I have previously gestured towards this radically externalist view of self-control (AUTHOR, omitted for anonymization); here, I develop it more fully.

To make my case for an access-based account of individual differences in self-control, I must first argue for self-control as subject to multiple realizability, defined by its behavioural function rather than by the presence of any specific mental process or mechanism. However, as I will point out, those who are not persuaded by such a functionalist account of self-control can still adopt a restricted form of the access account.

In this paper, my first order of business is to establish that exercises of self-control are constituted by a heterogeneous set of behaviours (including both mental and overt behaviours), pooled together by their applicability to the direction of behaviour[5] to match intention. Given their heterogeneity, the environment has a central epistemic and constitutive role in these behaviours. In section 2, I argue for such a functionalist approach to self-control by showing that a 'process' view of self-control, such as is found in Sripada (2020), must posit a further feature in virtue of which those processes are self-control, and that a 'results' view such as the one defended here provides just that. Further evidence is provided by findings in social psychology, which indicate that skill in self-control tracks the deployment of a variety of strategies.

In section 3, I analyse the concept of access as consisting of means, awareness and non-excessive effort. I show that the presence and absence of environmental constraints and supports modulate the presence of feasible means of self-control, as well as awareness of them. As a result, interpersonal and diachronic differences in self-control are a result of differences in environmental access to self-control behaviours[6], rather than just a result of innate individual differences. To have a capacity for self-control, for this stance, is to have access to self-control behaviours. It follows that the extent to which one has the capacity for self-control critically hinges on the extent to which they have access to self-control behaviours, and that it is commonplace for access to self-control to be unevenly distributed in ways that favour agents with certain innate and social traits over others.

---

[3] The concept of self-control as a 'mental muscle' was popularized by Roy Baumeister (Muraven & Baumeister 2000, Baumeister & Exline 1999, Baumeister, Vohs & Tice 2007).

[4] Not to be confused with the concept of epistemic access.

[5] In this paper, I speak much of behaviour. This is not to be taken to imply a covert commitment to Skinnerian radical behaviourism. It is simply used as a term that refers to both intentional and unintentional actions, movements, reflexes, etc.

[6] Some failures of self-control are such that an agent chooses not to control themselves, i.e., that she chooses not to enact self-control behaviours. The causes of these motivational failures are beyond the scope of this article; in any case, it is a different question than what I am probing here, which has to do with self-control at the level of efficacy rather than intentionality.

However, this contingency can be significantly ameliorated. Individual differences in self-control thus become a matter of justice.

Finally, in section 4, I demonstrate the practical upshots of the access theory and how it can be applied to particular cases by analysing attention-deficit hyperactivity disorder (ADHD) as a case where access to self-control is limited.

2. Self-control is a broad set of behaviours

Accounts of self-control in both philosophy and social psychology are highly varied. If there is a consensus, it is that self-control is that which enables one to resolve a motivational conflict and act accordingly: to inhibit a prepotent impulse in order to act intentionally, to follow one's judgment rather than desire, or to forgo a smaller, sooner reward to pursue a larger, later goal[7]. Sometimes these motivational conflicts further map onto a conception of the mind as divided, e.g., between reason and appetite, or between hot and cool processes[8]. Another compatible conception holds self-control to be roughly equivalent with inhibitory control and describes it as an intramental mechanism, often reducible to the brain[9]; sometimes, this mechanism is associated with a mysterious faculty dubbed 'willpower'. Sripada (2020) paints the debate in terms of *process views* of self-control, which define self-control as a specific process, and *results views*, for which self-control is defined by its 'result' or behavioural output and can be realized by a variety of processes, skills, and strategies[10].

Within self-control research, the suppression of unwanted impulses with 'brute' inhibition is often treated as a benchmark form of self-control against which other forms of self-control are compared. However, inhibitory control should not be equated with self-control. This is because inhibitory control processes are not specific to self-control, but are also present for other tasks that do not involve motivational conflict. Instead, they undergird much of skilled action, such as baseball batting (Gray 2009). In other words, there must be a further feature in virtue of which some instances of inhibition count as self-control. And if there is something in virtue of which some instances of inhibition count as self-control, then could not other processes and mechanisms also count as self-control if they share that feature?

The view put forth in this section is that the feature in virtue of which some processes count as self-control is the behavioural function of these processes; that a range of processes and

---

[7] Examples include Mele (1987, 1995) and Kennett (2001), who each characterize self-control in terms of judgment versus desire; and Ainslie (2001) and Fujita (2011) who define it in terms of smaller sooner versus larger later rewards.

[8] For divided mind accounts, see Levy (2011), where self-control is defined in terms of 'hot' system 1 processes vs 'cool' system 2 processes, and Kalis (2014), who defends the rationality over appetites view. Fujita, Carnevale & Trope (2018) offer an alternative to the divided mind approach, suggesting a model where self-control succeeds when a unified 'whole' self gets its way.

[9] See, for example, Cohen, Berkman & Lieberman (2013), who define self-control as inhibitory impulse control, and describe self-control as a function of the right ventrolateral prefrontal cortex.

[10] By my lights, it would be more informative to speak of the function of self-control than of its results, as the heart of the results views, in my view, is to treat behaviours that are functionally equivalent as falling within the scope of the same concept; for the purposes of this paper however, it does no harm to follow Sripada's terminology.

behaviours share this function; and that a range of functionally equivalent behaviours and processes therefore count as self-control.

In this paper, I define self-control by its behavioural function, which is the regulation of output to match motivational commitments such as intentions, plans, and the like. In brief, self-control is that which is enacted to align behaviour with intention, better judgment, or other motivational commitment in the face of a concurrent or expected competing motivation.

For this account, then, self-control simply is whatever fulfils this behavioural function; it falls under results views in Sripada's taxonomy.

The regulation of output to match intention is arguably the prototypical domain of self-control for this account. However, on accounts of intention that distinguish it from other sorts of motivational commitments (e.g, Audi 1991), the behaviours that enable matching output to intention also enable matching output to other sorts of motivational commitments such as plans and (for those who view self-control in terms of evaluative processes) preference or better judgment. On some accounts of intention and the relationship between motivation and action, self-control may occur in service of a commitment other than intention. In what follows, I will speak of conformity to intention for simplicity, but allow for self-control to also occur in service of other sorts of motivational commitments.

Take Miguel, who upon dining out grapples with the temptation to order steak. After rehearsing his reasons to forgo beef, he orders the more climate friendly dish. For the view espoused here, this is a success in self-control because Miguel, upon observing or expecting a competing motivation (to eat steak) enacted something (rehearsed reasons) to modulate behavioural output so as to conform with Miguel's intention (to forgo beef). Rafael, by contrast, loses his patience at family dinner and shouts at his mother despite having resolved to keep his cool. This counts as a failure of self-control because the behavioural output has not conformed with Rafael's intention to stay calm, but rather, with a contrary motivation.

In addition to the present paper, results views have been defended by Levy (2017) and, in social psychology, gestured towards by Duckworth, Gendler & Gross (2016) with what could be called the argument from efficacy. For this line or argument, if processes such as inhibitory control and attention regulation count as self-control because they regulate behaviour in the face of a motivational conflict, presumably the processes that most reliably achieve this regulatory purpose should count as paradigmatic instances of self-control. If behaviours and processes other than inhibitory control turn out to be equally or more effective than inhibitory control in regulating behaviour in the face of a motivational conflict, then this raises the question why we should think of processes such as inhibitory control as the benchmarks of self-control in the first place.

To illustrate the breadth of behaviours that, for my account, are functionally equivalent, I follow Duckworth, Gendler & Gross's (2016) rough division of the pool of self-control behaviours into situational strategies and intrapsychic strategies. Situational strategies include situation selection and situation modification. Situation selection involves avoiding the presence of an unwelcome temptation altogether. For example, Miguel, scheduling a family dinner, might choose a vegetarian restaurant instead of a steakhouse in order to stick to his resolution to avoid beef. In that case, Miguel is anticipating a motivational conflict, and enacts a behaviour (reserving a table at the vegetarian restaurant) to align his future behaviour with his intention.

Situational modification refers to behaviours undertaken within a situation to alter its physical and social features in a way that facilitates resolving the anticipated or concurrent

motivational conflict in favour of intention rather than conflicting motivation. For example, if Miguel is having dinner at a steakhouse, he might ask his dinner party to order him a salad so that he does not need to look at the tempting menu; or a professor striving to write an article but being tempted to scroll social media might unplug their wi-fi router. In these cases, again, a motivational conflict is experienced or anticipated, and a behaviour (asking others to order, unplugging the router) is enacted to align subsequent behaviour with intention.

Situational strategies are not always readily available; for example, in the Stanford experiments in delayed gratification (Mischel & Ebbesen 1970), children placed in a room with a treat did not have the option of leaving the room or otherwise altering the physical or social setup of the experiment. This leaves room for intrapsychic strategies of self-control (Duckworth, Gendler & Gross 2016). These are comprised of attention deployment (such as self-distraction), cognitive change (such as via construal or rehearsing one's reasons), and response modulation (including the suppression of impulses).

Intrapsychic strategies have been documented to be effective in delayed gratification tasks. For example, Mischel & Ebbesen (1970) observed that self-distraction was a predictor of success in delayed gratification. In addition to self-distraction, later work suggested that construal plays a similar key role. While children can spontaneously come up with these strategies, these can also be taught, which improves performance in subsequent delayed gratification tasks (Mischel et al. 2011).

In Levy's (2017) analysis, skills and strategies such as self-distraction, and construal, and situation selection count as self-control because they predict success in pursuing larger later rewards[11]. Noting that situation selection has been found to be positively correlated with measures on the trait self-control scale, but people high on trait self-control were not found to do better in ego depletion tasks[12], Levy (ibid.) suggests that the trait self-control scale may measure skill in environmental strategies of self-control. A similar conclusion can be found in Duckworth, Gendler and Gross (2016), in whose analysis intrapsychic strategies, while effective, are not as effective in producing the desired behaviour as environmental strategies[13].

At first brush, describing self-control as a heterogeneous set of behaviours may seem at odds with describing it as a capacity subject to interpersonal differences. It is certainly not a dedicated capacity if such a broad pool of behaviours is involved. Instead, it is a capacity in the same broad sense as mobility is a capacity: in the sense that there are individual differences in the extent to

---

[11] I have some reservations concerning thinking of self-control in terms of larger later rewards only as that introduces an evaluative dimension to self-control that in my view is unnecessary; however, this is a distinction that is of no consequence to the argument presented in this article.

[12] See Imhoff, Schmidt & Gerstenberg 2014 and Baumeister, Ent & Tice 2015 for the experiments, and Levy 2017 for an analysis of their implications for self-control.

[13] The distinction between situational and intrapsychic strategies is not clear-cut: for example, it is not clear whether attention deployment strategies deployed in the Stanford experiments, such as turning one's back to the marshmallow or singing, fall within the situational or intrapsychic group. The distinction is simply meant to illustrate the heterogeneity of functionally equivalent self-control practices. These blurred lines are not a problem for the functionalist view of self-control, as the distinction does no work for that conception. However, the fact that attentional shifting and construal are often enacted in such an embodied manner does generate some difficulty for those who wish to maintain intrapsychic heterogeneity of self-control practices as including not just brute inhibition but also processes like attentional shifting and construal, while excluding situational strategies from the scope of the concept of self-control.

which people can realize the sorts of behaviours that serve the function demarcated by the self-control construct. The availability of these behaviours is also subject to diachronic variation, which explains diachronic differences in self-control: Miguel's self-control may fail when he finds himself in an environment where his ordinary self-control behaviours cannot be readily enacted, and Rafael may be more successful in self-control in certain situations where salient means for self-control are present.[14]

The functionalist approach defended here is controversial in that it allows for any process that fulfils the behavioural function of self-control to count as self-control. By contrast, for some views, such as willpower theories and for the mechanistic process view described by Sripada (2020), self-control is either a single dedicated mental faculty, process, or mechanism, or else a narrowly defined set of such. The behavioural output, for such accounts, is a downstream effect of such a mechanism or process.

Willpower theories of self-control[15] are perhaps the best-known contemporary process accounts of self-control. Inspired by Aristotelian accounts of akrasia and enkrateia on one hand and by theories in social psychology on the other, they posit that what distinguishes self-control from other faculties is that it relies on a dedicated mental faculty. This mental faculty has been associated with the executive functions. Some willpower theorists make the additional claims that willpower, far from being a mere mechanism, is a discrete faculty that can be developed, exercised, and consciously exerted; and that it fatigues or becomes depleted when used, much like a muscle[16].

Yet the willpower approach faces numerous problems. Chief among these is that willpower accounts struggle to explain diachronic variation in self-control: it is puzzling how people with high willpower fail at self-control, and vice versa, how people low in willpower manage to succeed in it. Failures of self-control have been accounted for, in some willpower theories, by reference to the ego depletion hypothesis according to which willpower involves a limited, dedicated resource that is depleted with use; this hypothesis has faced copious criticism, including due to replication issues[17].

In response to these problems, those unwilling to adopt a results view of self-control have taken pains to revise the notion that self-control refers to a dedicated mechanism or process by clarifying it, setting aside ego depletion theories in favour of a different theoretical approach.

---

[14] While I claim that all interpersonal variation in self-control is best explained by differences in access, I do not claim that that all single instances of self-control failure would be so explained. Some failures happen due to arbitrary factors, such as neuronal noise (Stein, Gossen & Jones 2005), which renders all processes that involve the nervous system subject to some variability that agents cannot control for. These processes may cause a blunder in self-control even in the ideal circumstances, although I am agnostic as to how common this is.

[15] In philosophy, willpower theories have been advocated, e.g., by Henden (2008) and Holton (2003).

[16] In social psychology, perhaps the most prominent willpower theorists have been Roy Baumeister and colleagues (e.g. Muraven & Baumeister 2000, Baumeister & Exline 1999, Baumeister, Vohs & Tice 2007).

[17] For a report of a preregistered multilaboratory replication attempt which showed no effect for ego depletion, see Hagger et al. (2016). A recent replication by Kathleen D. Vohs (one of the originators of the ego depletion hypothesis) and colleagues likewise showed no effect (Vohs et al. forthcoming). A prominent methodological conundrum for the study of ego depletion is that it is difficult, within experimental settings, to distinguish the proposed ego depletion effect from the effects of fatigue. For an overview of the current state of the debate, see Inzlicht & Friese (2019).

Sripada (2020) offers an admirably executed such revision. For Sripada (2020), self-control consists in a skilled, extended sequence of cognitive control aimed at regulating responses associated with an unwanted motivational state (or emotion-type state, to use Sripada's terminology).

However, Sripada's account, while nuanced, raises the question *why* some such extended sequences of control count as self-control. After all, extended sequences of cognitive control aimed at regulating responses are not sufficient for something to be self-control, given that such extended regulation is present in much of skilled action. There must therefore be something further in virtue of which it counts as self-control. Sripada would likely concur: he reminds us that "an exercise of self-control doesn't consist of just any arbitrary sequence of cognitive control actions. It is rather a sequence that manifests the appropriate sort of *knowing how* to block the actional upshots of an emotion-type state" (Ibid., p. 14, italics in the original). But if such a sequence counts as self-control because it skilfully blocks actional upshots, this raises the question why should we think of self-control as necessarily involving such a sequence, rather than thinking of such a sequence as an accidental property of many instances of self-control? One reason to prefer a results view, then, is that it explains *why* Sripada's formulation is an apt description of many instances of self-control, while allowing self-control to be subject to multiple realizability.

Results views are not without their critics. In the remainder of this section, I offer some responses to criticisms that target results views of self-control, including the functionalist conception defended here.

1. Results views are normative or arbitrary. The startling heterogeneity of behaviours and processes labelled self-control in the results view has given rise to charges that self-control would be either a nonsensical category, the contents of which are arbitrary, or else a normative construct held together by the normative valence we give to self-controlled behaviour[18]. However, pooling the various behaviours and processes together by their behavioural function is neither normative nor arbitrary.

2. Functionalist concepts are useless for science. For this objection, describing self-control as a heterogeneous set of behaviours makes it a concept well suited for lay and philosophical conversations but hard for empirical scientists to operationalize. I agree, but I do not believe this poses a problem: empirical scientists can focus their research on specific behaviours or their specific correlates, as they have to date. In terms of continued empirical research, I merely urge a cautious acknowledgement that these operational targets of research represent but some of the many possible mechanisms and behaviours that self-control may encompass.

3. Results views yield conclusions that are unintuitive or that do not match everyday language. Sripada's (2020) chief objection to results views is that according to them, some behaviours count as self-control that do not intuitively strike us as such, and that we do not routinely label as self-control in everyday conversation. Sripada illustrates this with a range of cases. One these is that of Bo, who has an itchy rash on his arm that he intends not to scratch. In order to squash the contrary motivation to scratch the itch, Bo applies lotion on the rash and the urge to scratch goes away. Sripada's pre-theoretical view is that this "is clearly not an exercise of self-control" (2020: p. 18), despite that for results views it clearly is such. Sripada's reasoning is

[18] The arbitrariness claim is made by Herdova (2017), whereas the normativity claim is argued for by Kalis (2017) and Horstkötter (2015).

that Bo gets the results too easy and too fast, and should have to struggle with the urge for an extended period of time for his behaviour to count as self-control. Here, one may counter with the charge that gratuitous cognitive effort is not, pretheoretically, self-control either. If, after Bo's rash naturally stops itching, Bo rubs vinegar on it, causing it to itch again so that he can continue to refrain from scratching, many of us would be unwilling to describe this gratuitous activity as self-control. In light of this, it is unclear why the gratuitous struggle caused by abstinence from lotion would be self-control, either.

In other words, it seems that neither process accounts nor results accounts fully track all everyday language usage, and that both can be mined for counterintuitive cases. This shouldn't be a problem for philosophy or psychological theory, however. While our analyses of concepts like self-control should have robust overlap with everyday usage (or else we risk analysing an altogether different construct), it should also be expected that philosophical analyses sometimes yield results that urge us to revise our pre-theoretical notions.

Sripada's most persuasive case is that of a self-control pill. By eating this pill, an agent thwarts an anticipated urge to smoke. (ibid.) For Sripada, this is counterintuitive because we intuitively expect self-control to always involve effort, and the agent has it too easy. While I agree that this case is counterintuitive, I am entirely willing to bite the bullet. Sometimes enacting self-control is easy; sometimes it is hard; sometimes it takes forms that, pre-theoretically, seem odd to us. However, not all medication aimed at regulating behaviour is self-control: for a behaviour, such as the ingestion of a pill, to count as self-control it needs to be aimed at a current or anticipated motivational conflict[19]. For this reason, e.g., medication used to treat ADHD, though helpful, is not self-control for the functionalist view.

In light of the above discussion, I maintain that self-control is best conceived of as a broad set of behaviours united by their behavioural function. While controversial, this view gives us a conception of self-control that is conceptually coherent and matches existing information in social psychology concerning how, in practice, agents regulate their behaviour in the face of motivational conflicts.

I next describe my account of individual differences in self-control as resulting from differences in access to self-control behaviours. While the functionalist view of self-control described above gives the access account the widest consequences, those who are not persuaded by the above discussion can still accept a weaker version of the access account described below.

3. Which self-control behaviours are accessible?

That there is a heterogeneous pool of self-control behaviours rather than one such behaviour is fortunate, because it enables substituting one behaviour for another based on which strategies of self-control are a good fit for a given agent and their context. For example, a student sitting in an exam and noticing that her attention is wandering to an attractive classmate cannot physically remove herself from the space without forfeiting the exam, but she may attempt intramental

[19] We should also remember that thought experiments do not always yield reliable intuitions: in a world where such pills as Sripada describes were readily available, our intuitions surrounding these might be very different.

strategies of self-control, as described above. In theory, almost any agent, including agents with cognitive and learning disabilities, could possess a robust 'toolkit' of self-control behaviours that are a good fit for them, and select a means out of that set based on contextual fit. In this section, I argue that individual differences in self-control are best conceived of in terms of access to self-control behaviours. Differences in self-control, for this account, arise from barriers in the social environment. These barriers make many such behaviours unfeasible and therefore unsuitable as means to self-control. Furthermore, epistemic barriers prevent many agents from learning or becoming aware of suitable means to self-control. Unjust structures of the social and built environment thus result in aggravated individual differences in self-control.

My discussion of access draws from recent philosophical work on disability[20], which has brought attention to the way many disabilities disable not merely because of physiological differences between, for example, wheelchair users and those who walk, but also due to the various barriers the environment places on the former. The social and built environment, for these accounts, generates differences in the extent to which spaces, services, activities and the like can be accessed, providing access for some but not for all.

For example, wheelchair users have a hard time getting around in a world designed for people who walk, but this is not an intrinsic feature of the person using the wheelchair: rather, the environment is structured in such a way that helps others get around, such as by providing stairs, while hindering the wheelchair user from doing so, such as by omitting ramps and elevators, by placing them inconveniently, and by making it hard to find out where these are located. In brief, the environment contingently favours people with certain physiological traits[21], granting some a capacity to get around while placing obstacles on others. From this, it has been inferred that physical disabilities are not fully inside our bodies. The environment also plays a constitutive role in them.

However, despite a focus on the social and built environment in the study of disability and a near-ubiquitous adoption of accessibility measures as policies of inclusion, the concept of access or accessibility has not been the topic of systematic philosophical analysis. For my purposes here, it is sufficient to sketch out access as having three central criteria. These criteria are:

1. *Means.* For an agent A to have access to a space, behaviour, or practice X, there must be a concrete means M by which A is able to X (given A's traits).

---

20 See, e.g., Oliver (1996), Barnes (2016), Shakespeare (2008), Tremain (2015). While there is disagreement among these philosophers about the precise nature of philosophy and the extent to which disability is socially constituted, they agree that the social and built environment plays a considerable role in disabling people with certain biological traits.

21 The social model of disability (Oliver 1996) distinguished between impairments and disabilities, the former being the innately disadvantageous, biological foundation on which social stigma and physical barriers build the latter. In what follows, I will refer to traits rather than impairments because in many cases, the traits at hand do not themselves impair: they may be mere differences. (This is not true for all traits – e.g., chronic pain is itself obstructive for many life plans.) See Tremain (2015) for an analysis of the concept of impairment.

For example, for many people, walking up stairs is such a means M in accessing a meeting room on the 3rd floor. But that is not such a means M for wheelchair users. Ramps and elevators would be such a means for a larger group of people.

2. *Awareness*. The agent must be aware of the concrete means M by which they can X.

For example, the location of the ramp or elevator must be known or immediately salient to A.

3. *Non-excessive effort.* The effort it takes for A to M needs to not be excessive.

For example, if the ramps are located at the other side of a large building complex, the effort a wheelchair user must expend to get to the 3rd floor is excessive. Likewise, many people with limited mobility are able to use stairs but not without excessive effort; in such a case, stairs do not provide access for them. Here, I take no stance on what exactly determines whether something is excessively effortful: for the purposes of this paper, it suffices to note that we there are clear cases that illustrate that non-excessive effort is a criterion for access.

Applying the above criteria to self-control, for A to access self-control, there must be some self-control behaviours that are feasible means M to self-control for agent A. The agent needs to be aware of the availability and efficacy of these means to self-control. And the effort required to enact this behaviour should not be excessive. The central claim in thinking of self-control in terms of access is that means, awareness and non-excessive effort each are modulated by barriers in the past and immediate environment in addition to differences individual biology. In what follows, I will describe how such barriers to self-control may arise.

Barriers to means and to non-excessive effort. Recall that, for the view espoused here, a large pool of processes and overt behaviours count as self-control and can sometimes be means M to self-control. However, many such behaviours are not equally feasible for all. Some self-control behaviours are not feasible for some agents, whereas others may require excessive effort for some to enact.

For example, people who struggle with executive functions such as working memory, attention shifting, or inhibitory control, whether due to a neurodevelopmental disorder such as attention-deficit hyperactivity disorder, a mental disorder such as major depressive disorder, or due to lack of sleep, can be expected to encounter pronounced difficulties with intrapsychic strategies of self-control that are heavily reliant on cognitive control. Depending on the scope of their impact on executive functioning, these strategies and processes may not constitute a means M, or else they may require excessive effort to enact.

Genetic and neurodevelopmental factors, together with complex environmental causes at multiple junctions during the developmental trajectory, together influence what sorts of behaviours are feasible for the agent from the neurobiological perspective. However, we are not passive with regard to our biological make-up. People with difficulties in executive functioning may gravitate away from attempting such intramental strategies of self-control. Depending on whether other strategies are readily available and detectable, they may gravitate towards those strategies that they can enact without excessive struggle. While environmental self-control strategies are helpful for all agents, they can be a lifeline for agents for whom enacting intramental strategies is an excessive struggle, or for whom such strategies are not feasible.

However, whether the strategies of environmental selection and environmental manipulation constitute such a feasible means M for an agent is modulated, to a great extent, by the agent's immediate circumstances, which can both enable and constrain the use of these strategies. Environmental strategies for self-control can include, for example, installing an application that blocks access to certain websites, avoiding a specific street, place or person, staying physically away from sweets when trying to cut down on sugar, setting visible reminders about one's aims, and relying on friends, colleagues, and family for reminders and support in sticking to a chosen behaviour. These strategies are apt means for agents living in some but not all circumstances: for example, agents who drive have more control over avoiding specific streets than do those who use public transport, and the presence of social strategies as means to self-control is contingent on the presence of supportive social relationships. Removing doughnuts from a company coffee room may be a good strategy for avoiding them at lunchtime, but whether making that change is feasible depends not on the intrinsic traits of the agent but on her position at the company and the company culture regarding such initiatives. When being heckled and facing an impulse to punch the heckler, walking away is a tried and tested environmental self-control strategy, but one that often is not available for prisoners.

In other words, the circumstances of the agent preclude some strategies of self-control and make others excessively effortful, but these barriers to situational strategies of self-control are not equitably distributed. Nor is their distribution fully arbitrary. Rather, systemic, structural, and historical factors have contributed to a world where the options and opportunities agents have in leading their lives, including the scope of situational self-control strategies that are at their disposal, are unjustly distributed. In other words, lived circumstances generate barriers on strategies of environmental selection and manipulation, and the presence of such barriers tracks disadvantage. I do not claim all such barriers arise from disadvantage; rather, my claim is that disadvantaged agents face more such barriers than non-disadvantaged agents do.

Socioeconomic status is one such determinant of individual variance in access to means to self-control. Some of this determination is fairly obvious: for example, higher-SES people have access to more environmental strategies of self-control. They may, e.g., hire personal trainers to incentivize and remind themselves of their fitness goals or pay for other services in order to avoid contact with people and situations that they know put them at risk of self-control failure. Middle-class workplaces are also often more welcoming of environmental manipulation in the workplace, such as putting up post-its as reminders or even working from home, whereas working-class positions tend to be less flexible.

Yet it is plausible that SES furthermore impacts the feasibility and effortfulness of intrapsychic strategies of self-control. For example, socioeconomic status varies with physical and mental health, as well as with various neurocognitive capacities, such as memory, executive function, and language (Farah 2017). This is due to known mechanisms by which environments affect the brain, including, e.g., the effects of chronic stress on physical and mental health and on neural functioning (ibid.). While the connection of executive functioning to self-control is the most straightforward owing to that inhibitory control is an executive function, all of the above neurocognitive capacities may modulate the extent to which an agent can enact various strategies of

self-control.[22] However, if fitting self-control strategies are accessible, i.e., if there are self-control strategies that satisfy the three criteria of accessibility sketched above, then differences in executive functioning need not result in diminished self-control.

Barriers to awareness. Access to self-control is often prevented by epistemic barriers. Despite the complex difficulties in feasibility and effortfulness that some agents face, described above, if we take self-control to encompass a very broad set of behaviours, in nearly all circumstances (outside of thought experiments), there factually are *some* such behaviours that would be feasible and non-excessively effortful for agents, and that they therefore could access *if only they were aware of them*. In other words, even if the conditions of means and non-excessive effort are satisfied, an agent may not be aware of the availability of the behaviours that would constitute such means, or she might not know that those behaviours constitute an effective form of self-control.

In part, these epistemic barriers arise from continued confusion surrounding self-control. When self-control is discussed, it is often a 'black box': what happens in agents when they enact self-control is not expanded on. As self-control is increasingly discussed, this black box is beginning to crack open. Even so, intramental strategies continue to dominate both academic and lay discussions and views concerning what self-control is and how it is enacted. Agents struggling with self-control are ordinarily advised to make more of a mental effort, or else to think of the reasons they have for their preferred behaviour. In brief, practical advice on specific means of self-control is rarely given. When it is, the advice typically highlights intramental strategies of self-control. By contrast, while environmental selection and manipulation are often advised, e.g., in the context of goal pursuit, the information that these are de facto strategies of self-control is not readily available.

In terms of the epistemic condition for access, both past and present environment shape the extent to which we learn about various self-control strategies. There are considerable epistemic barriers that hinder the conscious learning and development of self-control strategies, despite that these strategies can be learned and taught[23]. At other times, self-control strategies are learned but not consciously so, as the agent either spontaneously generates them or has slowly acclimated into them. Some self-control strategies may be such that they require some practice or habituation to be effective, in which case the role of the past environment is highlighted.

However, the presence of specific self-control behaviours can be also made transparent in the immediate environment. For example, consider a campus study area where applications that allow students to block any websites that compete for their attention are pre-installed on the computers, and where there are visually captivating posters explaining how the application is used and its effectivity as a self-control strategy. A student entering such a space may not have previously been aware of this self-control strategy but will still be able to immediately utilize it with relative ease. In that case, the structuring of the immediate environment helps satisfy all three criteria for access: means, awareness, and non-excessive effort.

---

[22] Furthermore, SES may modulate to what extent it is rational for an agent to enact self-control in the first place. A study by Miller et al. (2015) found high trait self-control to correlate with better psychosocial outcomes in youth regardless of SES, but to be correlated with faster immune cell ageing in low-SES youth, suggesting that self-control may be a double-edged sword for such youth.

[23] See Mischel, Ayduk, Berman et al. (2011), Duckworth, Gendler & Gross (2016).

A further modulator of awareness is the salience of given self-control behaviours. In many situations, while multiple behaviours are feasible, only some of them are salient. Salience is situational: a strategy that an agent has been aware of in the past may simply not occur to an agent in a given situation. Salience can, however, be modulated by environmental cues as well as by habituation[24]. The role of salience is highlighted in fast-paced situations where the agent may not have time to consider his various self-control options, such as a family dinner where the conversation is getting tense and he wishes to avoid an altercation. In such cases, habituation – i.e., which strategy the agent typically tries to utilize – can be a large determinant of salience. Salience can guide agents to feasible, non-excessively effortful means, thus enabling access to self-control. But it can also misguide. The most salient means is not always an easy one, especially if the agent has only learned about means to self-control that require excessive effort for them to enact. For that reason, agents whose lived environments have emphasized strategies that are a poor fit for them may be at heightened risk of self-control failure, and pronouncedly so in fast-paced situations.

The access approach has the broadest consequences when combined with a broad results view of self-control. However, since barriers to awareness modulate the extent to which agents can learn any skill, Sripada's (2020) process view, which describes self-control as a *skilled* sequence of cognitive control, is also compatible with the access approach. For a combined process and access view of self-control, barriers to awareness obstruct the learning of cognitive control skills and thus contribute to individual differences in self-control. Proponents of the process view who wish to help people with self-control difficulties ought to take these barriers seriously.

4. When access is limited: the case of ADHD

In this section, I consider the implications of the access theory of self-control for a case where self-control is often limited. As mentioned above, many neurodevelopmental and mental health conditions modulate which self-control behaviours are feasible. Attention-deficit hyperactivity disorder (ADHD) is only one such condition, but I have chosen it as an example for two reasons: first, it is a fairly common disorder, and second, its symptom presentation, which spans self-control-associated difficulties across various domains from the occupational to the interpersonal, is sufficiently similar to nonclinical self-control problems for it to serve as an illustrative example (by contrast, addictive behaviour, often used as an example in philosophical examinations of self-control, is typically described as involving self-control problems in one domain only). Second, ADHD is associated with deficits on all measures of self-control, from inhibitory control and delayed gratification to trait self-control, and its symptomology corresponds to difficulties in self-control on any theory of self-control.[25]

Our first puzzle is interpersonal variation in self-control in people with ADHD. The symptomology of ADHD is thought to stem from differences in executive functioning. For many people with ADHD, their disorder is reflected in self-control problems across multiple domains,

---

[24] See, e.g., Luque et al. (2017) for the impact of cues and habituation on salience.

[25] For an in-depth examination of the relationship between self-control and ADHD, see Barkley (1997). For an up-to-date overview of the disorder, see Barkley (2015). Note that the role of the environment for self-control is also emphasized in discussions of addiction and self-control, such as in the essays comprising an edited volume on the topic (Levy 2013).

resulting, e.g., in poorer work performance and unstable relationships. There is, however, plenty of anecdotal evidence of high achievers with ADHD. The claim that ADHD involves limitations in self-control may seem at odds with the observation that some people with ADHD lead careers and lives that seem to place significant demands on their self-control capacity, such as entrepreneurial and academic careers.

For the access theory, this interpersonal variation is consistent with the observation that ADHD is associated with problems in self-control. People with ADHD face special considerations when it comes to access to self-control: an environment that is harmless for someone without ADHD may generate barriers to self-control for someone with it. For example, environmental distractions, while annoying, can be relatively harmless for the self-control of someone without ADHD but have a pronounced detrimental effect on the self-control of someone with the disorder. ADHD may furthermore make certain self-control behaviours, such as ones that rely on inhibitory control, attention shifting, and working memory, unfeasible or excessively effortful for the agent.

However, there are other self-control behaviours that it does not impact, or on which its impact is small. These latter behaviours include many environmental strategies and might plausibly also include some intrapsychic strategies such as construal. Some people with ADHD are well placed to discover and enact self-control behaviours that are a good fit for them, and thus may form a broad enough 'toolkit' for self-control to have awareness of means to self-control that are well suited for a variety of situations. Acknowledging the efficacy of these agents' self-control strategies is consistent with emphasizing the robust adverse effects of ADHD on the self-control of most people with the disorder; it furthermore implies that we ought to develop effective ways to help people with ADHD discover such self-control behaviours that are suited for these agents in the circumstances they live in.

Medication can help people with ADHD succeed in self-control tasks. However, medication does not remove ADHD-associated self-control difficulties altogether; and for some people with ADHD, suitable medication is not available. It is therefore important to also manage ADHD with non-medical means. Helping people with ADHD succeed in self-control entails facilitating situation modification strategies such as requesting frequent feedback or increased monitoring, adding visual reminders, and developing suitable situation selection strategies[26].

However, there is considerable variance in to what extent strategies of situation selection and situation modification are available for people with ADHD. For example, not all people with ADHD have supportive social and occupational environments where frequent feedback and monitoring could be arranged, or sufficient control over their living space that visual reminders could be installed. Furthermore, epistemic barriers can prevent people with ADHD from engaging in these practices. Notably, the tendency of laypeople, educators, and scientists alike to emphasize intramental strategies of self-control can prevent people with ADHD from learning about situational strategies of self-control that would be a better fit for their neural makeup.

Of further interest is that socioeconomic disadvantage and ADHD are associated (Miller et al. 2016). The symptoms of ADHD generate difficulty in employment and educational achievement, due to which adults with ADHD tend to earn less than peers (ibid.). The heritability of

[26] See Barkley (1997) for an account on the relationship of self-control and ADHD, and on how self-control can be fostered in people with ADHD.

ADHD compounds the problem: children with ADHD are statistically likely to have at least one parent with ADHD, whose educational and occupational struggles make children with ADHD more likely than peers to be born to lower income families. As SES modulates access to self-control behaviours both directly and indirectly, low-SES people with ADHD are in a double bind: not only is it harder for them to enact intramental self-control behaviours, but they also encounter epistemic barriers and lack means for situational self-control strategies due to their socioeconomic disadvantage[27].

As a result, the lived circumstances of many agents with ADHD, particularly circumstances of socioeconomic disadvantage, place further special challenges on discovering and enacting feasible self-control behaviours.

Many mental and neurological disorders adversely impact the executive functions, resulting in similar barriers to self-control. However, not all such disorders are fully analogous with ADHD with respect to these barriers. For example, while ADHD is consistent with lower measures of cognitive control, Tourette's syndrome, characterized by the presence of chronic motor and vocal tics, may not be associated with problems in cognitive control. Intriguingly, Mueller et al. (2006) found *greater* levels of cognitive control in youth with Tourette's compared to peers in an oculomotor task: the mechanisms underlying the tics, then, are dissimilar to the mechanisms underlying ADHD-associated traits. In some disorders, their associated atypical behaviours are reflexive in character rather than resulting from failure to manage motivational conflicts; theorizing about self-control has limited explanatory force in such cases.

The access theory claims that epistemic and practical barriers to situational strategies of self-control set people with ADHD (and with other disorders or lived circumstances that adversely impact their executive functioning) at a markedly disadvantaged position regarding self-control. These barriers may severely undermine the self-control of people with conditions such as ADHD, who may experience intramental strategies as unfeasible or disproportionately effortful. Barriers like these are also present for people without conditions associated with impaired executive functioning. However, their impact is smaller for people for whom intramental strategies are not such an uphill battle. The latter group of people continues to access some means to self-control even if they are prevented from accessing all such means. Yet the efficacy of environmental practices for self-control, demonstrated repeatedly in non-clinical samples, shows that if discourse on self-control were altered to emphasize environmental practices, this would benefit agents with and without ADHD alike.

Epistemic barriers are particularly hard for agents to identify because of the dimensionality of self-control impairments such as ADHD. That is, their different neurobiology may not make certain self-control practices, such as 'willpower' strategies, impossible – it merely makes them excessively effortful and difficult. As a result, one may hold on to the hope that with time, as one

---

[27] Many risks associated with ADHD also have an association with SES: while the precise mechanisms of interaction are unknown, there is some indication that the joint presence of low SES and ADHD is a larger risk factor than either of the two alone. Social risks associated with ADHD include, for example, substance use, problems in personal relationships, and problems with the law (Barkley 2015).). For example, people with ADHD are overrepresented in the criminal justice system (Koi, Uusitalo & Tuominen 2018). Incarceration is an example of a situation where access to situational self-control strategies is markedly limited.

patiently flexes one's 'mental muscle', these difficult practices would become easier. However, knowledge of a fuller array of effective self-control behaviours would better enable these agents to seek out a 'toolkit' of means to self-control that do not entail disproportionate difficulty.

If the access theory of self-control, as described in the previous sections, is correct, then it follows that by removing barriers to self-control described here, we are not just helping people with ADHD cope with poor self-control. Rather, we are improving their self-control.

Those who remain unconvinced about the broad functionalist view of self-control presented in this paper can, nevertheless, endorse a weaker version of the access view. As mentioned above, those who accept Sripada's (2020) process view should acknowledge the role of environmental supports and constraints in developing skill in cognitive control. However, even those who reject that self-control is skilled could endorse a variation of the access view: they may claim that these environmental strategies are compensation for poor self-control rather than part of self-control yet agree that these strategies are as valuable as de facto self-control for resolving motivational conflicts. Proponents of such stances can agree with proponents of the access theory that we ought to take practical steps to enable access to such strategies.

5. Conclusion

For the access theory of self-control, self-control is a set of practices and behaviours. Individual differences in self-control are explained by differences in access to these practices and behaviours. Differences in access arise not just from individual biology but also from the past and present environment. If the access theory is true, then individual differences in self-control are not merely a matter of inherent character: they are a matter of justice. By emphasizing intramental self-control behaviours over environmental ones, present discourse on self-control obscures the availability and efficacy of environmental self-control strategies, setting people with disorders such as ADHD at a disadvantage. Yet these epistemic barriers on self-control can be alleviated by emphasizing the heterogeneity of self-control behaviours. Practical barriers, by contrast, could be lifted by adjusting practices at schools, workplaces, et cetera to allow for a wider range of environmental practices in those contexts.

Willpower approaches to self-control have contributed to obscuring the need for actively removing barriers to self-control strategies. Further research on the precise policy changes and educational practices by which access to self-control could best be improved is urgently needed.

References

Ainslie, G. (2001). *Breakdown of Will.* Cambridge: Cambridge University Press.

Audi, R. (1991). Intention, cognitive commitment, and planning. *Synthese* 86 (3): 361–378.

Barkley, R. A. (1997*). ADHD and the Nature of Self-Control.* New York: Guilford Press.

Barkley. R. A. (2015). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment (4th ed.).* The Guilford Press.

Barnes, E. (2016). *The Minority Body: A Theory of Disability.* Oxford: Oxford University Press.

Baumeister, R. F. & Exline, J. J. (1999). Virtue, Personality, and Social Relations: Self-Control as the Moral Muscle. *Journal of Personality* 67 (6): 1165–1194.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of Self-Control. *Current Directions in Psychological Science* 16(6): 351–355.

Cohen, J. R., Berkman, E. T., & Lieberman, M. D. (2013). Intentional and incidental self-control in ventrolateral PFC. In Stuss, D. T. & Knight, R. T. (eds*.), Principles of Frontal Lobe Function: Second Edition.* Oxford: Oxford University Press.

Duckworth, A. L. & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality* 45 (3): 259–268.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science* 11 (1): 35–55.

Farah, M. J. (2017). The neuroscience of socioeconomic status: Correlates, causes, and consequences. *Neuron* 96 (1): 56–71.

Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Personality and Social Psychology Review* 15 (4): 352–366.

Fujita, K., Carnevale, J. J. & Trope, Y. (2018). Understanding self-control as a whole vs. part dynamic. *Neuroethics* 11 (3): 283–296.

Gray, R. (2009). A model of motor inhibition for a complex skill: Baseball batting. *Journal of Experimental Psychology: Applied*, 15 (2): 91–105.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science* 11 (4): 546–573.

Henden, E. (2008). What is self-control? *Philosophical Psychology* 21 (1): 69–90.

Herdova, M. (2017). Self-control and mechanisms of behavior: Why self-control is not a natural mental kind. *Philosophical Psychology* 30 (6): 731–762.

Holton, R. (2003). How is strength of will possible? In Stroud, S. & Tappolet, C. (eds.), *Weakness of Will and Practical Irrationality.* Oxford: Oxford University Press.

Horstkötter, D. (2015). Self-control and normativity: Theories in social psychology revisited. *Theory & Psychology* 25 (1): 25–44.

Inzlicht, M. & Friese, M. (2019). The past, present, and future of ego depletion. *Social Psychology* 50: 370–378.

Kalis, A. (2018). Self-control as a normative capacity. *Ratio* 31 (S1): 65–80.

Kennett, J. (2001). *Agency and Responsibility: A Common-Sense Moral Psychology.* Oxford University Press.

Koi, P., Uusitalo, S., & Tuominen, J. (2018). Self-control in responsibility enhancement and criminal rehabilitation. *Criminal Law and Philosophy* 12 (2):227-244.

Koi, P. (2020). Born which way? ADHD, situational self-control, and responsibility. *Neuroethics.* https://doi.org/10.1007/s12152-020-09439-3

Levy, N. (2011). Resisting weakness of will. *Philosophy and Phenomenological Research* 82 (1): 134–155.

Levy, N. (2013) (ed.). *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience.* Oxford: Oxford University Press.

Levy, N. (2017). Of marshmallows and moderation. In Sinnott-Armstrong, W. & Miller, C. B. (eds.), *Moral Psychology Volume 5: Virtue and Character.* Cambridge, MA: MIT Press.

Luque, D., Beesley, T., Morris, R.W., Jack, B.N., Griffiths, O., Whitford, T.J., & Le Pelley, M.E. (2017). Goal-directed and habit-like modulations of stimulus processing during reinforcement learning. *The Journal of Neuroscience* 37 (11): 3009–3017.

Mele, A. R. (1987*). Irrationality: An essay on akrasia, self-deception, and self-control.* Oxford: Oxford University Press.

Mele, A. R. (1995). *Autonomous agents: From self-control to autonomy.* Oxford: Oxford University Press.

Miller, G. E., Yu, T., Chen, E., & Brody, G. H. (2015). Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. *PNAS* 112 (33): 10325-10330.

Miller, L. L., Gustafsson, H. C., Tipsord, J., Song, M., Nousen, E., Dieckmann, N., & Nigg, J. T. (2018). Is the association of ADHD with socio-economic disadvantage explained by child comorbid externalizing or parent ADHD? *Journal of Abnormal Child Psychology* 46(5): 951–963.

Mischel, W. & Ebbesen, E. B. 1970. Attention in delay of gratification. *Journal of Personality and Social Psychology* 16: 329–337.

Mischel, W., Ayduk, O., Berman, M. G., et al. (2011). 'Willpower' over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience* 6 (2): 252–256.

Mueller, S.C., Jackson, G.M., Dhalla, R., Datsopoulos, S., & Hollis, C.P. (2006). Enhanced cognitive control in young people with Tourette's syndrome. *Current Biology* 16 (6): 570–573.

Muraven, M., & Baumeister, R. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin* 126 (2): 247–59.

Oliver, M. (1996). *Understanding Disability: From Theory to Practice.* Basingstoke: Macmillan.

Shakespeare, T. (2006). *Disability Rights and Wrongs.* New York: Routledge.

Sripada, C. S. (2014). How is willpower possible? The puzzle of synchronic self-control and the divided mind. *Noûs* 48 (1): 41–47.

Sripada, C. S. (2020). The atoms of self-control. *Noûs*, early view. https://doi.org/10.1111/nous.12332

Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: noise or part of the signal? *Nature Reviews Neuroscience* 6: 389–397.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality* 72 (2): 271–324.

Tremain, S. L. (2015). *Foucault and Feminist Philosophy of Disability*. Ann Arbor: University of Michigan Press.

Vohs, K., Schmeichel, B., Lohmann, S., Gronau, Q. F., Finley, A. J., Others, M., … Albarracín, D. (forthcoming). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science.*

Polaris Koi

(forthcoming in Cambridge Quarterly of Healthcare Ethics)

Towards informed user decisions about pharmacological cognitive enhancement

Abstract

Pharmacological cognitive enhancement (PCE) refers to the use of pharmaceuticals to improve cognitive function when that use is not intended to prevent or treat disease. Those who favour a liberal approach to PCE trust users to make informed decisions about whether enhancing is in their best interest. The author argues that making informed decisions about PCE requires a nuanced risk-benefit analysis that is not accessible for many users. Presently, the PCE use of prescription medications such as methylphenidate and modafinil is widespread, but most commonly happens without medical supervision. Direct and indirect barriers generate a situation where the risks and benefits of PCE are inequitably distributed; as a result, PCE is sometimes not in the user's best interest. This is likely to also be the case for future pharmaceuticals. As a result, even if PCE pharmaceuticals were equitably distributed, its associated risks and benefits would not be. The paper concludes with a discussion of the prospects of the clinical consultation on one hand, and e-health solutions on the other, in ameliorating the situation, arguing for cautious optimism.

Key words: human enhancement, pharmacological cognitive enhancement, cognitive function, risk-benefit analysis, Methylphenidate, Modafinil, e-health, decision-making

1. Introduction

Cognitive enhancement refers to measures taken to improve cognitive function when the purpose of the intervention is not the prevention or treatment of a disorder. Cognitive function is an umbrella concept that encompasses a range of mental functions including information processing, learning, memory, and executive functions such as inhibitory control, attentional guidance and forward-looking planning processes. In this paper, I discuss informed decision-making, particularly risk-benefit assessment, about current and future pharmacological cognitive enhancement (PCE).

Pharmacological cognitive enhancement refers to the use of pharmaceuticals in an effort to improve cognitive function. Currently, medications such as methylphenidate and modafinil, typically used as prescription drugs for attention deficit hyperactivity disorder

(ADHD) and for sleep disorders, are sometimes used by people without those disorders in an effort to improve their attentional and executive functioning capacities. Methylphenidate inhibits dopamine reuptake and increases the availability of dopamine and norepinephrine in the brain, a mechanism thought to explain its effect on patients with ADHD. Modafinil is used in particular for its wakefulness-promoting effects; however, exact mechanism underlying these effects remains unknown.

Not all cognitive enhancers are pharmaceuticals. For example, improved nutrition, caffeine, sugar, physical exercise, sleep, meditation and training can all improve cognitive function[1]. However, this does not entail that they would be ethically on a par.

Ethical considerations surrounding the individuals who decide to engage in PCE include concerns about the safety and efficacy of PCE, as well as concerns about the autonomy and authenticity of users. The general idea among proponents of PCE is that much like with cosmetic surgery, risking adverse effects is acceptable, so long as there is a reasonable expectation that this is, on balance, in the user's best interest; whether this is so is highly personal. For proponents of PCE, patients' autonomy should be respected by trusting their own assessment of whether they could benefit from PCE. For example, Nick Bostrom and Anders Sandberg suggest that the user should be allowed "to decide whether the benefits outweigh the potential risks, based on advice from medical professionals and her own estimates of how the intervention might affect her personal goals and her way of life."[2] Likewise, Vojin Rakic argues that the main arguments levelled against PCE "are not prima facie sufficient to prohibit interested individuals from using them"[3]. In other words, a liberal approach to PCE is justified by appeal to patient autonomy. It suggests that if medical advice is available, the user is well placed to assess the risks and benefits of PCE and decide accordingly; and to limit her choices would therefore intrude on her autonomy. Critics of PCE, by contrast, have raised autonomy-based considerations such as whether the decision to enhance can be autonomous, and whether the use of PCE could undermine autonomy and authenticity.

PCE also raises questions regarding justice and fairness. These include the concern that those who engage in PCE would stand at an unfair advantage to those who do not, and the concern that opportunities for obtaining the pharmaceuticals would be inequitably distributed, if considerations such as prohibitive pricing make it available only to a (wealthy) minority of people[4].

While the fairness of PCE has been much discussed in terms of whether the pharmaceuticals could be made universally available, whether they pose harms for non-users

(via 'doping' or competitive advantage), and whether they constitute a viable target for the use of scarce healthcare resources, the distribution of *risks and benefits* of PCE among actual and prospective users has been little discussed. In this essay, I discuss the complexities of risk-benefit assessments in deciding to enhance and argue that the risks and benefits of PCE are currently inequitably distributed among users. I further argue that this inequitability stems in part from both direct and indirect barriers to informed user decision-making about whether to engage in PCE.

In other words, even when prospective users are not directly barred from making their own decisions about PCE much like Bostrom and Sandberg suggest, we ought not take for granted that all prospective users would be well placed to make those decisions. Indirect barriers to informed decision-making result in decisions that are not in the user's best interest. In effect, the argument from patient autonomy can only be successful if prospective users have access to information, but this access is currently inequitably distributed. A liberal and equitable approach to PCE cannot simply assume that agents are well placed to assess whether they could benefit, but must actively enable agents to make that assessment.

A couple of caveats are in order about what this essay does not do. First, this essay limits its discussion to adults: the use of PCE in children is subject to further controversy that falls outside the scope of this paper. Second, I take no stance concerning the legality of PCE, although it is worth noting that legislation can work to build or dispel barriers to accurate information. Nor do I claim that equitable access to informed user decisions would be sufficient for PCE to be ethically acceptable. Multiple concerns, including safety, efficacy, authenticity, and whether enhancements can be collectively self-defeating, must be taken into account in the overall assessment of the ethics of PCE[5]. Rather, the aim of the paper is to highlight informed user decision-making as an important, neglected aspect in assessing whether the individual and collective benefits of PCE outweigh its harms, and an aspect that may, if unaddressed, be the demise of the libertarian approach to PCE.

2. The argument for PCE from patient autonomy

Pro-enhancement arguments often proceed from the standpoint of liberalism, within which patient autonomy plays a key role in medical decision-making. While there are multiple philosophical accounts of autonomy, within this framework, autonomy in the context of medical decision-making is generally taken as comprised of competence, authenticity, and liberty. Competence entails that the agent is sufficiently informed and possesses adequate reasoning capabilities in order to be able to assess what their options are, reliably evaluate

them and identify which option best matches their wishes and needs. Note that in the medical context, particularly, local competence suffices: for example, a person suffering from social anxiety may have inaccurate beliefs about how others regard them, but be perfectly capable of deciding whether to undergo major surgery. Authenticity entails, at a minimum, that the person decides on the basis of their own values and preferences. Liberty, here, entails that the agent is not being coerced, manipulated, or pressured to choose in a specific way. In the context of medical decision-making, it is assumed that adults, if armed with information, are sufficiently autonomous to make decisions. Paternalism is rejected on grounds that people vary in what values and preferences matter most to them, and are therefore uniquely well placed to make these choices.

Against this backdrop of highlighting procedural autonomy, pro-enhancement arguments defend the claim that adults should be free to make their own decisions about whether or not to enhance. For example, take Henry Greely et al.:

> From assembly line workers to surgeons, many different kinds of worker may benefit from enhancement and want access to it, yet they may also need protection from pressure to enhance.[6]

Here, Greely and colleagues emphasize the desiderata of equitable access to the pharmaceutical; the core threat to autonomy is coercion either to enhance or not to enhance. The same themes continue to be highlighted throughout the literature, such as in Thomas Metzinger and Elisabeth Hildt:

> In open, democratic societies [policies concerning cognitive enhancement] […] should be guided by a general principle of liberalism: In principle, the individual citizen's freedom and autonomy in dealing with their own brain and in choosing their own desired states of mind (including all of their phenomenal and cognitive properties) should be maximized.[7]

These authors, and others, have expressed their confidence that adults should freely choose whether they wish to engage in PCE. The theme of autonomy concerning PCE has been explored with an eye on whether legalizing PCE would limit autonomy by decreasing liberty, if there is a risk that peer pressure or other forms of coercion to enhance would emerge; and some have worried that the pharmaceutical effects of PCE would decrease autonomy by

decreasing authenticity[8]; or else suggested these effects may instead increase autonomy by increasing competence[9].

The role of competence in decisions to enhance has been a sidenote, however, plausibly because it is assumed that decisions to use PCE happen in similar epistemic conditions to other medical decision-making. Obviously, competent decision-making requires accurate information, and those pleading for liberal legislation around PCE also include calls to "thoroughly and *objectively* inform the public on the effects, side effects, chances, and risks of cognition-enhancing substances"[10] or to "increase public understanding of cognitive enhancement"[11]. Informing the people is treated as both a necessity and a sidenote.

As I will next demonstrate, however, engaging in PCE under limited or unreliable information is currently closer to a norm than an exception. Making accurate assessments of the risks and benefits of PCE requires nuanced and personalized information that is not straightforward to procure. Inequitable access to accurate information makes users poorly placed to make competent decisions about PCE. This need for nuanced and personalized information is unlikely to be a contingency specific to the current medical and social landscape, but rather is a factor that generalizes to future PCE. As a result, deregulation alone does not suffice for enabling the sort of procedural patient autonomy that liberals about enhancement endorse.

3. Assessing safety and efficacy in users requires nuanced and personalized information

In assessing the risks and benefits of PCE for the individuals who engage in it, the safety and efficacy of the pharmaceuticals used is a major concern. Currently, pharmaceuticals such as methylphenidate, piracetam, mixed amphetamine salts and modafinil are used by adults to improve their cognitive performance. Due to the varied means of procurement of pharmaceuticals for PCE, prevalence of use is hard to measure; previous studies within specific sample populations have indicated prevalence of use as high as 25 per cent on some college campuses[12].

In a recent nationally representative sample (N=102,000), by contrast, 1,9% of U.S. adults use stimulant medications for purposes other than the management of a disorder without use disorders. Out of these users, 56,3% reported using the medication to help them stay alert or concentrate; the second largest purpose of use was to help study (21,9%)[13]. This suggests that while not as common as studies of specific populations have indicated, the PCE use of stimulant medication is a widespread phenomenon. The data gleaned from this study also indicates that the PCE use of prescription stimulants may not be the tool of high

achievers that some suggest, being positively correlated with a household income of less than $20,000 and with prior and current substance use disorders; it may, instead, sometimes indicate an attempt to improve a life of hardship. While there likely are pronounced geographic, and cultural differences in the prevalence of use between the United States and other populations, studies of prevalence of PCE on college campuses and in other specific populations outside the U.S. establish this is not limited to U.S. alone[14].

Drug regulation organs within the US Food and Drug Administration (FDA) and European Medical Agencies (EMA) have not approved PCE as an indication for pharmaceuticals such as methylphenidate or modafinil. PCE is nevertheless promoted by influential personalities: for example, in a recent self-help book, entrepreneur and media personality Dave Asprey claims that modafinil "gives you superhuman mental processing powers with few to no downsides"[15]. While these claims are false, they are received by a significant fan following and contribute to PCE as a growing phenomenon found in various walks of life, rather than merely a form of 'college doping'.

From a public health perspective, one pertinent concern is whether those who engage in PCE are at risk of harming their health. Medications such as methylphenidate and modafinil are often touted as safe; however, all medications have side effects, and the safety of each medical intervention should therefore be established by assessing the ratio of risks to benefits, which is subject to some interpersonal variance based on counterindications and what the individual can reasonably hope to gain from the intervention.

While early studies indicated some promise for cognitive enhancers, subsequent replications and meta-analyses have showed the cognitive enhancing effects of current pharmaceuticals to be modest at best. Both non-users and users of PCE tend to overestimate their effects, which is one possible source of both exaggerated claims, like Asprey's, and exaggerated concerns surrounding autonomy and authenticity for these pharmaceuticals. Anecdotal experiences of 'superhuman mental processing powers' may thus result from the improvements in mood, motivation and confidence that these disproportionate expectations cause.[16] Furthermore, there is some debate about the overall efficacy of major PCEs such as methylphenidate and modafinil, including in clinical use in the treatment of disorders such as ADHD; their effect in the treatment of ADHD is modest[17], which urges further caution about making strong claims about the benefits of these pharmaceuticals. The hope is that future pharmacology may yet produce interventions that deliver more robust effects than current pharmaceuticals do.

Alterations in cognitive function resulting from current medications impact a broad swath of neural functioning. Due to the interconnected nature of neural mechanisms, it is likely that this will also be the case for future pharmaceuticals. The approach found in some earlier literature[18] on cognitive enhancement that operates on an idealized notion of a perfectly safe future cognitive enhancer should therefore be considered useful as a thought experiment only. Actual PCE interventions, current and future, are not likely to be limited to cognitive functioning in their effects.

Despite being generally well tolerated, PCE substances each have significant side effects that severely decrease their applicability for some individuals. For example, modafinil and methylphenidate each may cause weight loss[19], which is a counterindication when the prospective user is underweight or suffers from an eating disorder. In a review of studies concerning PCE medication efficacy in healthy individuals, observed adverse effects for modafinil included dizziness, headache, nausea, dry mouth, abdominal pain, sleep disturbances including insomnia, palpitations, tachycardia, nervousness and restlessness[20]. These adverse effects, while not dramatic, have cumulative effects that may undermine the goals of users in engaging in PCE. A college student, stressed and tired from working a part-time job while getting her degree, may engage in PCE in the hopes that her improved cognitive performance would make assignments quicker to complete and thus make her life less stressful. But if insomnia and nervousness result from the medication, her life may turn out to feel more, not less, exhausting.

The risks for insomnia, nervousness and restlessness are particularly important to take into consideration for prospective users with mental disorders, such as Major Depressive Disorder (MDD), or with a history of the same. This is because lack of sleep and anxiety are risk factors for many mental disorders[21]. In these cases, both the prospective risks and the prospective benefits of PCE may be pronounced: on one hand, the user risks triggering a new episode of their relapsed disorder (or intensifying a current one). On the other, disorders such as MDD adversely impact the executive functions[22], and persons with such disorders may therefore have a more to gain from PCE. Assessing risks and benefits in users therefore requires nuanced reflection that considers the promises and risks of PCE *for that user*.

Informed and reflective decision-making about PCE stands to alleviate these concerns greatly. Informed users may assess to what extent the possible adverse effects of pharmaceuticals would be harmful for them, and to reflect on the possible results of PCE and whether their expectations are realistic. Based on such an individualized assessment, risks and benefits could be more accurately gauged. Such a multifactorial assessment is not an easy

task, however, and it would be made greatly more feasible if accessible, verified information and guidance were available, such as via consultation with a physician.

However, such a consultation is not within easy reach of all prospective users. In the next section, I will argue that the current heterogeneity of PCE procurement generates barriers to informed decision-making about PCE. As a result, some users risk considerably larger harms than others.

4. Barriers to informed decision-making raise problems for the autonomy view

One source of the variable safety of PCE is the considerable variance in how users obtain these pharmaceuticals. While there are geographic and demographic differences in obtainment, stimulant medications are most often procured from friends or relatives, either for free, by buying, or by stealing[23]. Pharmaceuticals for PCE may also be obtained from online sources[24] or from the pharmacy with a physician's prescription. Legality greatly influences user strategies in obtainment.

When pharmaceuticals are obtained from friends and family, while users may conduct online or library research on the medications, they will also be subject to culturally transmitted impressions of the safety and efficacy of the medication. Uncertainties about the pharmaceutical are addressed, in part, through learning about peer experiences and opinions. These experiences and opinions may amount to the transmission of disproportionate optimism, resulting in the user inaccurately assessing the risks and benefits of PCE and choosing to enhance based on inaccurate information. The reverse may also occur, as adverse peer experiences may cause disproportionate pessimism, depriving someone who would benefit from the possible positive effects of PCE.

Uncertainties surrounding risks and benefits do not evaporate by trying out the pharmaceutical. As mentioned above, improved productivity after consuming a pharmaceutical may simply result from increased optimism and motivation. Furthermore, users may also, even after repeated use, be uncertain about whether the pharmaceutical has been effective. Here, a lengthy citation from a study by Margit Anne Petersen et al. is pertinent. 'Harrison', a MA student and one of their interviewees, describes his struggles in assessing the efficacy of stimulants as cognitive enhancers.

> There are times where I wonder if stimulants do anything positive at all. Certainly superficially it would seem like they are, you have increased energy, this sort of intense connection, focus on the material, but at the higher doses, it starts to become

> slightly frantic. . .and if you're writing something very quickly, you have this sensation that you're being very productive and that you are enhanced as a worker, but just because you're writing quickly doesn't necessarily mean that you are writing more or of higher quality. […] It's no question that I would, if possible like to stop taking stimulants. I would like to stop, but really the deciding factor is, do they make me better at working? And if the answer is yes, then it's something I'm willing to risk and willing to take on the burden of. . . in order to do more things. But if they don't then it's really stupid to be taking them.[25]

Harrison's thoughts about his PCE highlight two matters. First, that conducting an accurate risk-benefit analysis is difficult for users, even after repeated use. And second, that the risk-benefit analysis may be highly personalized. Harrison comes across as saying that if he can "do more things", then he is willing to endure a fair bit of adverse effects. Some other persons would doubtless rank productivity lower in importance than, say, adverse impacts on sleep quality. But Harrison also has a hard time judging whether the stimulants in fact do improve his productivity. He is acutely aware that his current decision-making about PCE is not as well-informed as he would like it to be.

Users like Harrison can rely on a variety of resources for information: they can consult online information, friends, relatives, medical textbooks, and physicians. However, information is not universally or equitably accessible. Not all users have the requisite educational background (regardless of whether that is the result of formal education or self-education) to be able to discriminate which sources of information are trustworthy; for this reason, self-directed online information searches will only support informed decision making for those users with suitable background knowledge. While some friends and relatives may have accurate information about pharmaceuticals, peer networks are also a source of culturally transmitted myths about PCE. Petersen et al. call this set of beliefs a 'folk pharmacology' and note that the folk pharmacology influences not just deliberation about whether or not to engage in PCE, but also interpretations of the positive and adverse effects of PCE[26].

Returning to the criteria for autonomous patient decision making – those of competence, authenticity, and liberty – people like Harrison clearly are not being directly coerced to PCE. However, even well-educated adults may lack local competence for making decisions about PCE. It can be hard for them to assess whether PCE in fact helps or hinders their own values and preferences.

As the case of Harrison illustrates, there is good reason to believe that it is in fact rare for an agent to be well placed to assess whether PCE stands to benefit them. This is a serious problem for the liberal approach given its assumption that the user knows best.

5. Do real or ideal clinical practices enable informed decisions?

To the above, the proponent of the liberal view will likely respond that medical advice can resolve our concerns about decision-making competence. Consulting a physician may be the most reliable aid for informed decision-making. Of course, not all physician recommendations are sound. However, physicians are extraordinarily well placed to give informed counsel on the topic, and therefore consulting with a physician with the requisite background information can be considered a benchmark for informed decision-making, placing decisions about PCE on a par with other difficult medical choices.

Users are however not equally well placed to access such a consultation. First, not all users have access to the sort of robust healthcare where consultation on PCE would be available. Second, even when that is available, due to legal parameters surrounding PCE, not all users would feel comfortable discussing the prospective or ongoing use of PCE with a physician, particularly if the pharmaceuticals are not obtained in a manner sanctioned by the medical establishment.

It may then appear that obtaining the pharmaceuticals via a prescription by a physician would better enable users to access accurate information about those pharmaceuticals, within the same clinical relationship where the prescription is written. However, obtainment from a physician, while rare in comparison to obtainment from friends and family, is likewise diverse in its methods. It bears repeating that the consumption of stimulants for the treatment of a disorder is not PCE. Users of PCE may obtain a prescription by finding a physician willing to prescribe the pharmaceutical for off-label use; by finding a physician willing to falsely diagnose a disorder in order to prescribe the pharmaceutical, seemingly for an approved indication, in full knowledge that the intended use is PCE; or by leading their physician to believe that they do in fact have an approved indication for the medication, such as, by convincing a physician that they have ADHD. Setting aside ethical considerations surrounding wilful misdiagnosis as the focus here is on user decisions, only in the first two patient-physician interactions is the physician well placed to counsel the patient in making informed decisions about PCE. When the physician has false information about the health of the patient, their advice concerning the risk-benefit ratio will be coloured by that false information. Correspondingly, the EMA's conclusion is that for modafinil, only in the

case of narcolepsy do the benefits outweigh the risks[27]. Even if the EMA's assessment were to be taken as a generalization, and the risk-benefit ratio of modafinil were adequate for *some* PCE users, this highlights that physicians' impressions about the health of their patient will strongly impact how they counsel patients wishing to gain from PCE. The procurement of a pharmaceutical by falsely portraying a disorder thus prevents the patient from discussing the actual projected risks and benefits with that physician.

In sum, users use a variety of methods to procure pharmaceuticals for PCE. It continues to be uncommon for a physician to prescribe the pharmaceutical in full knowledge that it will be used for PCE: more often, the methods of procurement reflect a variety of ways to avoid, rather than make use of, consulting medical professionals concerning PCE. The variable means of procurement are a result of widespread prohibitive guidelines and laws surrounding PCE, as well as a result of variable access to healthcare services.

The status quo of PCE distribution is therefore doubly inequitable. First, some people, but not all people, have access to pharmaceuticals for PCE. Access to pharmaceuticals seems to track access either to a peer group where these are circulated, or to a medical practitioner who is either willing to prescribe the pharmaceutical in full knowledge of that PCE is its intended use, or else can be misled.

But there is a further structure of inequitable distribution relating to PCE. Namely, access to informed decision-making about PCE is also inequitably distributed. Out of those who engage in PCE, only some are able to make their assessments about the risks and benefits of PCE based on accurate information: others will either make them under uncertainty, like Harrison, or under false expectations, like readers of Asprey.

Perhaps for future pharmaceuticals, their enhancing effects would be pronounced enough that it would be easier to assess whether one benefits from it. However, as discussed above, cognitive effects without side effects are a theoretical possibility only. In practice, future pharmaceuticals, too, will be riskier for some than they are for others. The need for nuanced risk-benefit analyses and informed decision-making therefore persists even as pharmacology advances.

The barriers for informed decision-making, however, are a contingency that could be ameliorated. If, for example, physician consultations surrounding PCE would be made universally accessible as well as socially and legally acceptable, this can be expected to have a formidable impact on the barriers surrounding informed decision making about PCE. This is a theme for the next section.

5. The clinical solution may not suffice

In this paper, my aim is not to defend or oppose the universal availability of PCE; that topic has been discussed elsewhere in the literature[28]. If, however, PCE should become universally available, should everyone be able to access it after consulting with their doctor? In light of the above discussion of the role of medical professionals in informed decision-making, this would seem like it would help guide users to accurate information about PCE. Clinical consultation can be treated as an adequate baseline for informed decision-making: after a clinical consultation, or its equivalent in terms of personalized information, we can have reasonable confidence that the user is able to make an informed decision about PCE.

However, even if ideal, a scheme like this runs into the problem of scarcity. Proponents of PCE sometimes argue that current pharmaceuticals are not, or need not be, prohibitively costly and that there is no principled obstacle for wider-spread production; likewise, for future PCE, if the cost of the future pharmaceutical is reasonable, there would be no principled reason not to produce enough to provide adequate coverage on a par with other publicly accessible health services in welfare states. If the pharmaceutical can be produced with a modest cost, it need not be the privilege of the few. However, medical professionals are not so easily mass produced. Even if pharmaceuticals were to cost pennies, consultations with doctors would not. In other words, in a world where PCE pharmaceuticals would be equitably distributed, access to informed decision-making about PCE – and, as a result, the risks and benefits of PCE – would continue to be inequitably distributed.

Could society ensure equitable access to physician consultation regarding PCE, such as via training more doctors? This solution, while seemingly attractive, runs into the problems of scarcity and prioritization. Given that there are copious global health problems, a pressing point about prioritization of healthcare resources remains: one can with good reason ask whether clinicians should put their time in consulting people on PCE, when they could conceivably spend that time in other pursuits, such as the treatment of disease. While many optimistic futures and drastic policy changes can be envisioned, reasoning in medical ethics that hinges on first resolving the problem of the scarcity of expert labour may not be the most fruitful approach.

Additionally, access to clinical consultation, even on matters of treating disease, is inequitably distributed. In some countries, the lack of a substantial public healthcare system prevents some from accessing adequate health services; but even in countries with such a system, some socioeconomic disparity in access to clinical consultation persists. One source for these disparities is that public healthcare systems are typically complex to navigate, and

citizens are not equally well informed about what services they are entitled to and how these services can be accessed. As it stands, equitable access to clinicians is a work in progress, including in countries with robust public health programmes. As a result, even if universal clinical consultation on PCE were offered, which users would in fact benefit from this resource would likely track other measures of socioeconomic disadvantage.

Recalling that the role of the consultation is to provide personalized, accurate information, in other words, even in an abundance of any given substance used for PCE, there may not be an abundance of accurate information. If PCE were cheap and easily available, but accurate and nuanced information about it were as hard to come by as it is now, then there would likely be many more finding themselves in Harrison's shoes: using PCE but wondering about whether it is in fact helping, or whether they are instead doing something unwise. Yet others would have come to conclusions about enhancing or not enhancing that would be founded on misinformation. This comes at a cost: some who gain little or no benefit may be unduly bearing the burden of risks and adverse effects, whereas some who stand to gain much may decide not to enhance based on false or exaggerated beliefs about the risks of enhancing. While the decision to enhance should always be based on the needs and wants of the subject, improved availability of accurate information could help subjects better assess whether PCE responds to those needs.

A further concern in providing adequate consultation for prospective users of PCE is the complex role of past and current psychiatric symptoms. There are multiple reasons to treat this as a central, rather than peripheral, concern for the ethics of PCE. First, many of the adverse side effects of stimulants currently used for PCE, such as insomnia, anxiety, and loss of appetite[29], are particularly harmful for people with past or current substance use, eating, or psychiatric disorders, as for many such disorders these effects present a clinical risk.

Second, the prevalence of the use of PCE is documented as higher in people with Major Depressive Disorder and with suicidal ideation than in persons without these mental health problems, suggesting that PCE may be used to self-medicate. Many of the adverse effects associated with stimulant medication carry a risk of new or intensified depressive episodes.

Third, these users may stand to gain more from PCE. Many psychiatric illnesses adversely impact the executive functions, whose functioning stimulant medication has a modest positive effect on. For some psychiatric patients, this modest effect may have a large impact for their overall well-being even if it is not intended to cure the disease. For persons

with existing psychiatric symptoms, stimulants therefore pose a higher risk, but may also pose larger benefits for some within this group.

Fourth, the demographic data surrounding current stimulant use for PCE suggests that many PCE users may come from vulnerable populations, as the phenomenon is correlated not just with psychiatric history but also with many other measures of health and socioeconomic status. In a recent study of stimulant use prevalence in the United States, the use of stimulants seven times or more in a month was correlated with having less than a high school education, with past and current substance use disorders, with depressive episodes, and with being disabled for work[30]. In other words, the popular imagination of the cognitive enhancer as a hotshot law student may not accurately represent the persons that health policy surrounding PCE will have the largest impact on.

From a distributive justice perspective, then, even if equitable access to the pharmaceuticals for PCE could be ensured, the risks and benefits of these pharmaceuticals would not be equitably distributed. This is because the extent to which users would in fact be well placed to make decisions about PCE would not be equitably distributed. Liberals about PCE should respond to this concern by urging practical steps towards equitable access to accurate and personalized consultation to improve access to informed decision-making. Yet the scarcity of expert labour makes this desideratum difficult to put to practice.

However, perhaps there is yet hope for the liberal stance: there may be ways to provide accurate and personalized information without the requirement of prohibitive amounts of expert labour. Particularly, the recent increase in electronic health services (e-health, for short) carries both promises and burdens. In the next, final section of this paper, I discuss whether e-health could help improve equitable access to accurate information about PCE.

6. Accessing personalized information via electronic health services

The above discussion, which highlighted clinical consultation as a sufficiently reliable support for informed decision-making, has likely struck some readers as quaintly old-fashioned. After all, for many users of health services, clinical consultation is not their first stop in acquiring medical information: the internet is.

However, while online medical information abounds, some of it is not trustworthy and very little of it is personalized. The role of the clinical consultation, after all, is not simply to convey general facts about PCE, such as can be found in medical textbooks, but to address

how they relate to the specific user's general health, whether there are contraindications for use, and how likely that user is to benefit.

While information supplied by general medical information websites is too general to fulfil this role, in principle, interactive health systems may satisfy these demands. Indeed, patients increasingly employ various interactive online information tools, such as symptom checkers, as a means to help assess their health needs. These interactive tools range from simple programmed questionnaires, such as symptom checkers, to complex artificial intelligence algorithms. In the United Kingdom, the National Health Service's (NHS) use of various mobile health tools has been described as empowering the patient: for this narrative, apps, health sensor devices, personal health software and the like are described as tools that patients may use to shift control of their health increasingly from the doctor to the patient. While this narrative is worth questioning, digital solutions are fairly well suited for the role of personalized information seeking and may take the role of a consulting companion in personalized healthcare, helping patients gain the local competence required for them to make informed decisions about their health[31].

Depending on the outcome alternatives for the digital consultation, e-health can take on an important triaging role. One example of this has been seen during the Covid-19 pandemic, during which many public health systems utilized advisory symptom checkers that advised patients on whether their symptoms satisfy the grounds for getting tested for the virus according to local public health policy. All of this gives grounds for some cautious optimism about the use of digital solutions for the provision of personalized guidance about PCE.

This optimism, however, is met by three lines of criticism. First, while digital solutions are less resource intensive than clinical consultation, they do use resources, and any use of healthcare resources must be justified against other possible targets for those resources. The development of digital solutions, then, is only justifiable if it can be done at an acceptable cost. Another line of criticism is that digital solutions would not provide accurate enough personalized guidance. For this line of criticism, however, it can be responded that well-designed Artificial Intelligence (AI) diagnostics and counselling already sometimes outperform consultation by human clinicians. While poorly designed AI can yield poor recommendations, such as the well-known problem of recommendations unduly based on race or gender, these are not an inevitable feature of digital solutions, but can instead be avoided by skilful design of the digital solution[32].

However, there is a third line of criticism that is much harder for digital solutions to shake off. Namely, the sort of information that is pertinent for an accurate risk-benefit

consultation about PCE is rather more sensitive than is the information required for assessing whether the patient is a candidate for a virus test. This raises dire concerns surrounding data privacy and data ownership.

For example, for assessing the possible risks and benefits of current PCE pharmaceuticals, awareness of factors such as prior psychiatric symptoms is important, as it is crucial that the user is aware that some forms of PCE may cause symptoms such as anxiety and that their existing symptoms may be compounded by the use of PCE. On the other hand, some persons with a prior history of psychiatric illness may benefit more than others from PCE, for which reason it would not be an unequivocally good solution to simply advise persons with past or current mental health symptoms not to enhance. As many psychiatric conditions adversely impact the executive functions, on which stimulant has a modest positive effect, for persons with these conditions this modest effect may have a large impact for their overall well-being even if it is not intended to cure or manage the disease. Furthermore, for some persons who have symptoms but have not received a diagnosis or mental health services, they may seek out PCE rather than those services in an effort to manage those symptoms. A system on a par with clinical consultation would notice this and direct such individuals to a mental health professional.

Accurate risk-benefit analyses therefore require detailed information about the prospective user's mental health. Given the particularly sensitive character of such information, stringent regulation of data privacy and data ownership[33] are prerequisites for the consideration of processing sensitive data within an e-health interface. In other words, in addition to ethical considerations concerning patient safety and distributive justice, as discussed above, the longstanding ethical quandaries of information ethics become pertinent when looking into e-health for solutions to the scarcity problem. These information ethical concerns are not specific to PCE, but rather, apply to e-health more generally. However, they ought to be emphasized in this context, particularly given the sensitivity of the data required.

The prevalence of persons with diagnosed or undiagnosed psychiatric symptoms in users of PCE also suggests that any system for accessing personalized information about PCE must be designed to properly respond to vulnerable patient groups. One possibility is for the role of the e-health application to be a triaging one, directing persons with strong counterindications or from potentially vulnerable groups to a consultation with a human clinician. However, such a solution is only as equitable as factual access to such a consultation is. Vulnerable patient populations currently face a range of barriers to accessing the full range of health services, resulting in poor public health outcomes[34]. It is likely that e-

health triaging and direction to clinical consultation would fail to reach all patient groups equally, and that targeted improvements in access to health services would be a prerequisite for truly equitable access to PCE. It should be highlighted, again, that this is no marginal issue for PCE, since the population level research on current PCE indicates that persons with past or present psychiatric symptomology are strongly represented in PCE users.

7. Conclusion

Inequalities related to enhancement have typically been discussed in terms of whether enhancing is an unfair advantage. However, as I have above demonstrated, all is not fair among users of cognitive enhancers. In this essay, I have highlighted that the risks and benefits surrounding PCE are not equitably distributed among users. In my analysis, I have argued that differences in access to informed user decision-making are a source of these disparities.

While liberals about PCE believe that the patient knows best whether or not PCE is right for them, as the above discussion highlights, few users are well placed to assess the costs and benefits of PCE. Instead, decisions about whether or not to engage in PCE are often made under uncertainty or false impressions.

While the popular imagination portrays the persons using PCE as hotshot law students and ambitious professionals, demographic data indicates that many who enhance come from vulnerable and disadvantaged groups. If access to informed decision-making about PCE is not ensured, there is a danger that users from these groups end up engaging in PCE with adverse rather than beneficial results to their performance and well-being.

Access to informed user decision-making is paramount for an accurate risk-benefit distribution. Presently, for PCE, such access is spotty at best. From a public health perspective, improved access to improved user decisions yields improved health and well-being for current and prospective PCE users. The more accessible PCE pharmaceuticals become, whether through affordable pricing, steps towards deregulation, or a combination of both, the more pressing the issue of informed user decision-making becomes.

Finally, I have discussed e-health services as one possible means to help improve access to informed user decision-making. E-health services are booming, and it is possible that the e-health scenario for PCE consultation, described above, is closer to a default near future tendency than it is to a reasoned solution. However, the emergence of clinical software solutions, in particular, has taken medical ethicists and regulatory organs somewhat by surprise: for example, in the U.S., the booming industry of mobile health applications has

been minimally regulated as a result of many regulatory guidelines pertaining to hardware rather than software. Raising these public health concerns should occur before, not after, a practice is widespread, so as to prevent harms that could have been identified in advance. To do so, ethicists need to keep an eye not just on the philosophical and conceptual aspects of PCE, but also on the contingent, empirical, public health factors that shape the real consequences of the policy at hand.

Policy concerning PCE should endeavour to enable equitable access to informed user decisions about PCE. Even if better, more effective PCE substances are devised, the need for informed decision-making persists. If e-health solutions are applied for this purpose, however, medical and information ethics must join forces to meet the challenge of generating an accessible, safe, and ethically sustainable digital solution.

---

[1] Dresler M, Sandberg A, Ohla K, Bublitz C, Trenado C, Mroczko-Wasowicz A, et al. Non-pharmacological cognitive enhancement. *Neuropharmacology* 2013:64:529–43.
[2] Bostrom N, Sandberg A. Cognitive enhancement: Methods, ethics, regulatory challenges. *Science & Engineering Ethics* 2009:15:311–41, at 323.
[3] Rakic V. Commentary: Cognitive Enhancement: Are the claims of critics "good enough"? *Cambridge Quarterly of Healthcare Ethics* 2017:26:4:693–8.
[4] Faber N, Savulescu J, Douglas T. Why is cognitive enhancement deemed unacceptable? The role of fairness, deservingness, and hollow achievements. *Frontiers in Psychology* 2016:7:232.
[5] Maslen H, Faudmüller N, Savulescu J. Pharmacological cognitive enhancement – how neuroscience could advance ethical debate. *Frontiers in Systems Neuroscience* 2014:8. DOI: 10.3389/fnsys.2014.00107 ; Bognar G. Enhancement and equality. *Ethical Perspectives* 2012:19(1):11–32.
[6] Greely H, Sahakian B, Harris J, Kessler RC, Gazzaniga M, Campbell P, et al. Towards responsible use of cognitive enhancing drugs by the healthy. *Nature* 2008:456(7224):702–5, at 705.
[7] Metzinger T, Hildt E. Cognitive enhancement. In: Illes J, Sahakian B, eds. *Oxford Handbook of Neuroethics*. Oxford: Oxford University Press; 2012;245–64, at 259.
[8] Habermas J. *The Future of Human Nature*. Cambridge: Polity Press: 2003.
[9] Schaefer GO, Kahane G, Savulescu J. Autonomy and enhancement. *Neuroethics* 2014:7:123–36.
[10] See note 7, Metzinger & Hildt, at 359. Italics in the original.
[11] See note 6, Greely et al., at 705.
[12] See note 6, Greely et al.
[13] Compton WM, Han B, Blanco C, Johnson K, Jones CM. Prevalence and correlates of prescription stimulant use, misuse, use disorders, and motivations for misuse among adults in the United States. *The American Journal of Psychiatry* 2018;175:741–55.
[14] Ricci G. Pharmacological human enhancement: An overview of the looming bioethical and regulatory challenges. *Frontiers in Psychiatry* 2020:11:53. doi: 10.3389/fpsyt.2020.00053
[15] Asprey D. *Game Changers: What Leaders, Innovators and Mavericks Do to Win at Life*. London: Thorsons: 2018, p. 54.
[16] Petersen MA, Enghoff O, Demant J. The uncertainties of enhancement: A mixed-methods study on the use of substances for cognitive enhancement and it's unintended consequences. *Performance Enhancement and Health* 2019:6:111–20; see also Ilieva IP, Boland J, Farah MJ. Objective and subjective cognitive enhancing effects among college students: Prescription status, motives, theory of planned behavior, knowledge and self-diagnostic. *Neuropharmacology* 2013:64:496–505.
[17] See, e.g. Cortese S, Adamo N, Del Giovane C, Mohr-Jensen C, Hayes AJ, Carucci S, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *The Lancet Psychiatry* 2018:5;9:727–38.
[18] See note 2, Bostrom & Sandberg.
[19] See note 17, Cortese et al.
[20] Repantis D, Schlattmann P, Laisney O, Heuser I. Modafinil and methylphenidate for neuroenhancement in healthy individuals: A systematic review. *Pharmacological Research* 2010:62:187–206.

[21] Herstentein E. et al. Insomnia as a predictor of mental disorders: A systematic review and meta-analysis. *Sleep Medicine Reviews* 43: 96–105.
[22] Malloy-Diniz LF, Miranda DM, Grassi-Oliveira R. Editorial: Executive functions in psychiatric disorders. *Frontiers in Psycholo*gy 2017:8:1461. doi: 10.3389/fpsyg.2017.01461
[23] See note 13, Compton et al.; see also Singh I, Bard I, Jackson J. Robust resilience and substantial interest: A survey of pharmacological cognitive enhancement among university students in the UK and Ireland. *Plos One* 2014:9(1):e105969.
[24] Sahakian BJ, LaBuzetta JL. *Bad Moves: How Decision Making Goes Wrong, And the Ethics of Smart Drugs.* Oxford: Oxford University Press, 2013: p. 120.
[25] See note 16, Petersen et al.
[26] See note 16, Petersen et al.
[27] European Medicines Agency. *Assessment report for modafinil containing medicinal products*. EMA, London. EMA/4038/2011.
[28] See, e.g., Dubljevic V. Prohibition or coffee shops: Regulation of amphetamine and methylphenidate for enhancement use by healthy adults. *American Journal of Bioethics* 2013:13(7):23–33.
[29] See note 13, Compton et al.
[30] See note 13, Compton et al. at 743.
[31] Morley J, Floridi L. The limits of empowerment: How to reframe the role of mHealth tools in the healthcare ecosystem. *Science and Engineering Ethics* 2020:26(3):1159–83.
[32] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* 2019:25:44–56.
[33] Wilkowska W, Ziefle M. Privacy and data security in e-health: Requirements from the user's perspective. *Health Informatics Journal* 2012:18(1):191–201.
[34] Johannes B, Graaf D, Blatt B, George D, Gonzalo JD. A multi-site exploration of barriers faced by Vulnerable patient populations: A qualitative analysis exploring the needs of patients for targeted interventions in new models of care delivery. *Primary Health Care Research & Development* 2018. Available from: https://www.cambridge.org/core/journals/primary-health-care-research-and-development/article/multisite-exploration-of-barriers-faced-by-vulnerable-patient-populations-a-qualitative-analysis-exploring-the-needs-of-patients-for-targeted-interventions-in-new-models-of-care-delivery/ (accessed April 29, 2021).