Towards a Theory of Control

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2013

CHRISTOPHER DAVID OVEN DEN

School of Social Sciences
# CONTENTS

**ABSTRACT** ......................................................................................................................... 4

**DECLARATION** .................................................................................................................... 5

**COPYRIGHT STATEMENT** ..................................................................................................... 6

**ACKNOWLEDGEMENTS** ......................................................................................................... 7

**INTRODUCTION** .................................................................................................................. 9

I - What is control? ..................................................................................................................... 18

II – Three Species of Control .................................................................................................... 24

III – Sketch for a Theory of Agential Control ......................................................................... 38

IV – The Collection .................................................................................................................. 66

**THE ANTI-AKRASIA CHIP** .................................................................................................. 79

1. Guidance Control .................................................................................................................. 83

2. The Anti-Akrasia Chip .......................................................................................................... 90

3. Objections and Replies. ........................................................................................................ 95

4. Summary and Prospects ........................................................................................................ 118

References................................................................................................................................. 121

**GETTING SPECIFIC WITH MANLEY AND WASSERMAN** ................................................ 122

1. Rejecting the purely restrictive ........................................................................................... 124

2. Steel, Schmeel ....................................................................................................................... 128

3. A dual approach to analysing dispositions ......................................................................... 138

4. Counting cases ..................................................................................................................... 141

5. Measuring C-case-space ...................................................................................................... 149

Conclusions ............................................................................................................................... 161

References................................................................................................................................. 161

**RELEVANT ABILITIES** ........................................................................................................ 163

1. The New Dispositionalism ..................................................................................................... 166

2. Alternate possibilities for action ......................................................................................... 178
**ABSTRACT**

The University of Manchester  
Christopher David Ovenden  
Doctor of Philosophy  
Towards a Theory of Control  
September 2013

Control is a concept that has received surprisingly little attention in the philosophy of action and ethics, given its *prima facie* ties to freedom, responsibility, intentionality and agency more generally. In this collection I take the first step towards an account of agential control: the kind of control that agents commonly exercise over actions, events, and even other agents.

In the introduction I give a sketch of the complete thesis on control: characterising agential control as consisting primarily in the restriction or guidance of some process, and secondarily in the continuous monitoring of that same process. I go on to suggest that the primary aspect of control involves an agent’s having the ability to effectively intervene in the process that they are controlling.

The collection itself consists of three journal style papers that, whilst not being focussed explicitly on control themselves, begin to fill out the sketch in my introduction: roughly, I think that control requires an ability to intervene (effectively, an ability to do otherwise), I think that ability should be understood as a kind of disposition to effectively intervene in a process should an agent try, and I think that to build a satisfactory conditional account of dispositions we need to appeal to the recently proposed contextualist account of dispositions from David Manley and Ryan Wasserman.

The three papers aim to support each of these thoughts: The first paper, ‘The Anti-Akrasia Chip’, presents a counterexample to the well-known Fischer-Ravizza account of guidance control and suggests that what that account lacks is an emphasis on an agent’s being able to effectively intervene in their own behaviour; the second paper, ‘Getting Specific with Manley and Wasserman’, uses a novel counterexample to motivate a particular reading on Manley and Wasserman’s contextualist account of dispositions; and the final paper, ‘Relevant Abilities’, involves a defence of dispositional compatibilism by introducing the notion of a *relevant* ability: one grounded by the contextualist account of dispositions developed in the previous paper.
DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
COPYRIGHT STATEMENT

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
ACKNOWLEDGEMENTS

There are so many people who have had a hand in getting this PhD thesis to where it is now; I suspect I could fill every page in this collection with names and still not have listed them all.

I should like especially to extend my sincerest thanks to Tom Smith, for his unwavering support and guidance through the many ups and downs of the past few years, and for always finding time for one ‘last’ supervision; to Joel Smith, for countless invaluable supervisions, and for always managing to kick me back into gear right when I needed it; to Michael Clark, for friendship greater than which I couldn’t have asked for, and practical jokes I genuinely didn’t ask for; to the postgrads and staff of the University of Manchester Philosophy department, for their many helpful comments and suggestions on near every word in this collection, and for sitting through so many talks on Fischer and Ravizza’s Responsibility and Control they might well be able to recite the whole book by heart; to my family, for supporting me every step of the way and suffering through many a philosophy talk over dinner; and of course to Mary, for sticking with me through thick and thin, helping me smile through every rough patch, keeping me sane, and putting up with me at my very grumpiest.

I couldn’t have done it without you.
For my parents,

David & Francesca
INTRODUCTION

CHAPTER SUMMARY

The papers contained in this collection constitute a small part of a much larger thesis on the nature of control. In this introduction I try to give a sketch of that larger thesis and show where the papers fit in to its overall scheme. This, necessarily, results in a rather long introduction involving my quite rapidly running through a number of control-relevant definitions, several of which cut across one another, and moving quickly over areas of the overall thesis that deserve a dissertation all of their own. That being the case, it will be helpful to have a map of this chapter to follow.

I first do some work to motivate an enquiry into the nature of control and situate the thesis in the current philosophical literature on freedom, responsibility and agency. In particular I highlight the lack of a fully worked out account of control in current writing.

In §I – What is control? I aim to provide a rough characterisation of control based upon its uses in ordinary language; I consider the definitions provided for control in its verb and noun forms by the Oxford English dictionary and settle on a gloss of control as ‘monitoring and restricting or guiding some x.’

In §II – Three Species of Control I identify three different kinds of control that agents can enjoy: mechanical, authoritative and agential. I point out agential control as the most relevant to freedom and responsibility and then go on to briefly discuss some ways this kind of control can be exercised:

- Self-control
- Moral control
- Conscious and automatic control
- Direct and indirect control

In §III – Sketch for a Theory of Agential Control I lay out the rough outlines for the theory of control that unites the three papers in the collection. I begin by distinguishing between two modes of agential control: controlling and being in control.
of whether. I then move on to discuss the criterion for being in control of something: I first identify two Frankfurtian intuitions that I suggest ought to inform our theory of control: i) that being in control does not require an ability to do otherwise, and ii) that being in control requires an ability to effectively intervene; the conflict between these two intuitions I suggest can be resolved by the introduction of a distinction between two kinds of ability: those involving what I call metaphysical access and what I call counterfactual access.

Next, I look briefly at the awareness constraints on being in control before considering the context sensitivity of control attributions.

Finally I present my sketch analysis of being in control and identify processes as being the proper objects of agential control (no matter its superficial object: action, agent or event).

In §IV – The Collection I give an overview of each paper of this collection and identify where they fit in to the overall thesis sketched in the previous section.

Over recent decades, philosophers have spilt a great deal of ink over the likes of free will, alternate possibilities, moral responsibility, agency, reasons, beliefs, desires, intentions and intentionality. Relatively little, however, has been written on the nature of control; specifically, the kind of control that agents regularly exercise over their own behaviour, events in their environment, and even other agents.

This ought to be surprising given the role that control potentially plays in accounts of many of these other concepts: Acting freely, morally responsibly and intentionally all prima facie look like they involve a certain level of control: If an agent was not in control of some action of theirs $A$, then how could $A$ have been an expression of their agency? How could they have performed $A$ freely or intentionally? How can that agent be responsible for $A$ if its performance was not under their control?
In fact, a great many classic philosophical problems can be rephrased in such a way that they appear to be problems of control. Take, for instance, Davidson’s (1973) case of a wayward, or ‘deviant,’ causal chain:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. (1973, p. 79)

This kind of case is particularly problematic for causal accounts of freedom or intention, where an action is free or intentional if it is caused by mental states that rationalise that action; Davidson, for instance, comments:

I tried analysing, ‘A is free to do x (or can do x)’ in terms of the conditional, ‘He would do x intentionally if he had attitudes that rationalized his doing x’. Even if we read this subjunctive conditional as implying a casual relation, we can see now that it is not adequate. If the agent does x intentionally, then his doing x is caused by his attitudes that rationalize x. But since there may be wayward causal chains, we cannot say that if attitudes that would rationalize x cause an agent to do x, then he does x intentionally. (1973, p. 79)

Essentially, if an intentional or free action is one that is caused by an agents beliefs and desires (specifically those that make performing such an action rational), then the climber’s loosening his hold ought to be free and intentional. Clearly though, the
climber does not act intentionally, and the possibility of such wayward causal chains from reasons to action illustrates a startling flaw with the simple causal theory of action.¹

Interestingly, we might think that the failure of Davidson’s climber to act intentionally is actually a failure of control: the climber has all the relevant mental states in order to act freely and intentionally when they loosen their grip, their belief that they can rid themselves of weight and danger by loosening their grip on the rope and desire to do so certainly rationalise loosening their grip, it is just that their action is not controlled by them in the right kind of way. In this example, and others like it, the mental states of the agent cause their actions to occur in such a way that they are cut out of the picture and act ‘beside themself’. What is lacking from the climber’s situation, so that we could describe his loosening his grip as a free and intentional action of his, is that his doing so was controlled by him or suitably under his control. An analysis of control, or in this case controlling some action, might therefore be instrumental in exploring cases of deviance further and in developing accounts of non-deviant action or intentional and free action more generally.

Similarly, control presents a potentially untapped angle on so called ‘Frankfurt cases’ introduced by Harry Frankfurt in his (1969) ‘alternate possibilities and moral responsibility’ (though which really herald back to Locke’s (1961) sleeping man example²) and frequently discussed in literature on abilities, freedom and responsibility. Take Frankfurt’s (1969) classic case:

---
¹ For further discussion of responses to wayward, or deviant, causal chains see: (Peacocke, 1979); (Mitchell, 1982); (Mele, 1987); and (Stout, 2010).
² See Locke 1961, bk. 2, chap. 21, sec.10
Suppose someone-Black, let us say-wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way.... Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. (1969, pp. 6-7)

The intuition we are supposed to draw from this is that whilst Jones is morally responsible for whatever action he performs, he is unable to perform any other action, and thus that alternate possibilities are unnecessary for moral action (and perhaps also freedom).

The case becomes less clear, however, when we start to question what exactly Jones is morally responsible for. Suppose that Black wants Jones to shoot and kill the mayor and that this is the very action that Jones performs (with the result that the mayor dies). Jones seems obviously responsible for his action, ‘shooting the mayor’, but what about the subsequent outcome, ‘the mayor’s death’? It is far less obvious what we ought to say about Jones’s responsibility for this event (given that no alternative could have been brought about).
This case and its many variations have become something of a meme in the literature on action, freedom and responsibility, spawning a huge and ever growing literature. However, thinking about these kinds of cases in terms of control, rather than just the responsibility of the agent involved, presents a way of bringing out further intuitions on such cases that might otherwise have been masked by considerations of freedom and responsibility and potentially help us to make some progress in the long standing debates on them: Intuitively, Jones controls his own action, and is in control of his shooting the mayor, given the inactivity of Black’s device, but is not in control of whether he shoots the mayor or whether the mayor dies. On the other hand, Black appears to be in control of the situation without directly controlling Jones: he is in control of whether Jones shoots the mayor, and in turn whether the mayor is shot; he does not, however, control Jones’s behaviour in an active sense.

Reflections such as these can help us to develop our account of control; for instance, from the brief assessment above it appears that there is a distinction to be made between being in control of some action, or controlling some action, and being in control of whether some action or event comes about (I go on to discuss this distinction in §III below). However, more than that, if control turns out in some way to be a constituent of moral and free action, then these intuitions about control ought to ground more robust analyses of freedom and responsibility.

In spite of these ties to the literature on freedom and responsibility, comparatively little has been done to give an analysis of control. The term is and

---

3 See, for instance, (Fischer, 1994; 1995; 2002a; 2002b); (Widerker D., 1995a; 1995b); (Widerker & Katzoff, 1996); (Kane, 1996); (Ginet, 1996); (Wyma, 1997); (O'Connor, 1993); (Warfield, 1996); (Hunt, 1996); (Speak, 1999); (Mele & David, 1998); (Stump, 1990; 1996; 1999); (McKenna, 1997); (Otsuka, 1998); (Copp, 1997); (Heinaman, 1986); and (Vihvelin, 2000; 2008).
utilised in current literature, but this is usually in service to some other concept and
does not equate to a fully worked out account of control itself.

Pettit (2001), for instance, discusses a theory of freedom as ‘rational control’
according to which an action is free “so far as it is an exercise of rational control or
power on the part of the agent” (2001, pp. 34-5) and where “[agents] will enjoy rational
control or power just so far as they operate properly as intentional subjects: that is, as
subjects of intentional states like beliefs and desires” (2001, p. 35).\(^4\) This turns out,
roughly, to mean that one’s actions be ‘rationally directed’, or perhaps determined in
some way, by one’s beliefs and desires and for those beliefs and desires to be rationally
amended in light of new information that becomes available (so that, for instance, we
will expect an agent enjoying rational control who is exposed to evidence that not-p
to revise their belief that p) (2001, p. 35).

Control, in this case, might almost be thought of as a simple relation of
determination: one has rational control when one’s actions are rationally determined
by one’s beliefs and desires, with those beliefs and desires themselves being rationally
determined by available evidence. Whilst this might capture part of what controlling
some action involves, it certainly falls short of a full analysis of the concept.

Similarly, Widerker (2005), uses control to characterise the Libertarian thesis
about freedom: “[some] of the actions we perform are free, that is within our control
in the sense that it was within our power not to perform them.” Here, control is
characterised as a kind of power or ability to do otherwise; specifically, to not perform,
or to prevent the performance of, some action or other. This, however, is as far as the

\(^4\) Pettit also goes on to discuss two further theories of freedom: as what he calls ‘volitional control’
(Pettit 2001: Ch.3) and what he calls ‘discursive control’ (2001: Ch.4), eventually settling on freedom
as discursive control. In each of these cases, as with rational control, the concept of control itself is not
fully analysed, but is rather a subservient concept to freedom and responsibility: discursive control
almost being a stand-in name for freedom. I focus on ‘rational control’ here for example’s sake only.
characterisation of control goes and, again, whilst having some kind of ability or power to do otherwise might be an aspect or element of having control, we can imagine kinds of control in which no ability to do otherwise (at least of the incompatibilist variety) is present at all: for instance, John Martin Fischer and Mark Ravizza (1998) distinguish between two kinds of control: regulative control and guidance control, the latter of which they claim does not require an ability to do otherwise.

Fischer and Ravizza characterise ‘guidance control’ as the process of actually guiding an action through its performance (with or without an ability to do anything else) and they do go on to offer an analysis of this concept: roughly, that an agent has this type of control whenever their behaviour is issued by a process that is suitably responsive to reasons and is genuinely a process of their own. However, this account is largely driven by their focus upon moral responsibility, and once again control ends up being called on not for its own analysis, but in service to some other concept: according to their overarching scheme, an agent is morally responsible just so long as they have guidance control (or exercises guidance control at relevant points in the history leading up to their present situation) and so the term ‘guidance control’ becomes little more than a placeholder for the kind of state one is in when one is morally responsible for one’s actions.

Finally, a great deal of philosophical literature exists on akrasia—an agent’s acting contrary to what they judge they have most reason to do—and so by extension on ‘self-control’, which is generally treated as the contrary of akrasia: conforming one’s actions to what one judges one has most reason to do. Whilst this is a commonplace usage of ‘control’ in philosophical literature, and it might capture some notion of control, it is no fully worked out analysis of control.

---

5 See, for instance, (Mele, 1995) and (Watson, 1977).
Holton (2009) notes that there are two distinct uses of self-control: one, as above, being the opposite of akrasia, where this involves acting in accordance with what one judges best; and two being roughly equivalent to what he calls strength of will, which involves having the willpower not to reconsider resolutions\(^6\) one has made to act (Holton, 2009, p. 128 fn.21).\(^7\) As above, it seems to me that both of these senses might represent an aspect of control, and so should have a part in a wider account of the concept, but do not capture what control, or controlling, essentially entails.

The aim of this collection is to take a step towards a fully worked out account of what I shall call agential control: the kind of control that agents exert over themselves, their environments and, at times, other agents around them; it is this kind of control that I think lies at the heart of many accounts of freedom and responsibility, and holds strong ties to agency and intentionality. This collection of three papers makes a start at filling in some of the details of the sketch for a theory of agential control that I lay out in this introduction.

It is worth noting that I approach control in this collection from a compatibilist angle. In the last few decades some philosophers, notably John Martin Fischer, have turned away from straightforward compatibilism concerning the relationship between freedom and determinism and have adopted instead what Fischer refers to as semi-compatibilism: the thesis that moral responsibility is compatible with determinism, whether or not free will turns out to be.\(^8\)

If that is semi-compatibilism, then my project in this collection might well be described as semi-demi-compatibilism: I think that, irrespective of whether free will

---

\(^6\) Holton characterises resolutions as a special kind of intention that is “designed to stay firm in the face of future contrary inclinations or beliefs” (2009, p. 10).

\(^7\) See also (Holton 2009: Ch.6) for the account of strength of will.

\(^8\) See, for instance, (Fischer & Ravizza, 1998) and (Fischer, 2002b; 2012).
and moral responsibility turn out to be compatible with determinism, if our universe is causally deterministic (or indeed causally-indeterministic) agents will enjoy agential control over certain aspects of their behaviour. In actual fact, I think that freedom and responsibility are also compatible with a causally deterministic universe, but as compatibilists (more generally) we need to start at the bottom of the agential ladder with control over action and build out accounts of free and responsible action from there.

I begin, in this introduction, by considering control’s usage in ordinary language along with some motivating cases that aim to bring out our intuitions on different kinds of control before going on to give a brief sketch of a theory of agential control.

I - WHAT IS CONTROL?

Control can mean a variety of different things, not all of which are relevant to agency, so it will be instructive to first compile a list of some of the things that control can mean in its everyday usage and then consider how those meanings relate to the more technical usage of ‘control’ relevant to agency and action.

To start us off, let’s consider some of the definitions under which ‘control’ is used in ordinary language. The Oxford English Dictionary provides the following definitions for the verb ‘control’: 9

1. trans. To check or verify, and hence to regulate (payments, receipts, or accounts generally): orig. by comparison with a ‘counter-roll’ or duplicate register; now in the wider sense of 4.

2. transf. To check by comparison, and test the accuracy of (statements, stories, or their authors). arch.

3. Hence:

a. To take to task, call to account, rebuke, reprove (a person). Const. of, for. Obs.

b. To challenge, find fault with, censure, reprehend, object to (a thing). Obs.

4.

a. To exercise restraint or direction upon the free action of; to hold sway over, exercise power or authority over; to dominate, command.

b. To hold in check, curb, restrain from action; to hinder, prevent (? obs.).

c. refl. To hold in check or repress one’s passions or emotions; so to control one’s feelings, tears, etc.

5.

a. To overpower, overmaster. Obs.

b. Law. To overrule (a judgement or sentence).

c. Fencing. to control the point: ‘to bear or beat it down’ (Gifford in loco).

Note that the abbreviation ‘Obs.’ means ‘obsolete,’ so definitions (3a), (3b), (5a) and possibly (4b) are taken by the OED to no longer be in general use. I include them, and refer to them, because I think they help to characterise the overall nature of control, even if this meaning is not generally appealed to in everyday English any longer.

Of all of these, the definitions under (4) look to be the most relevant to agency. Notice that (4a) characterises control as ‘exercising restraint or direction upon the free action of [some thing]’ (my emphasis), illustrating a dual nature to what we might call the primary, and what turns out to be active, aspect of control: first, restraining and
curbing some process or thing through an exercise of power or authority, hindering and preventing (from (4b)), and holding in check and repressing (from (4c)); second, directing and commanding the free action of some thing, exercising power to guide that thing as one sees fit. Call these, respectively, the *restrictive* and *directive* elements of control’s primary aspect.

The definitions under (3) also suggest the *restrictive* element of control’s nature, though with more emphasis on actual confrontation of a subject that needs to be restrained: calling to account and rebuking (from (3a)), challenging and censuring (from (3b)). It is in this sense that we might say that alcohol or violent media is controlled in certain areas (through preventing access to it), or that a teacher might control an unruly class by challenging, reprehending and rebuking them; in each case the emphasis is on taking some action to restrict and prevent some thing or group in accordance with what one sees fit.

Finally, the definitions under (5) give a characterisation for the verb as the taking of and expression of control (as restricting and directing): one controls the point in fencing with an act that puts one in a position to further direct the duel; one, at times, controls an unruly person by overpowering and overmastering them, again putting oneself in a position to further direct them; and one can control a judgement by overruling it, stepping in to overturn it and prevent it from passing.

Definitions (1) and (2) illustrate what we might call the *secondary* aspect of control: to verify and check, to regulate some system, person or process in accordance with some set standard. This, we might think of as the silent partner to the Primary aspect of control: For the purposes of agency wherever there is control by restriction or direction, there must be awareness of the thing controlled; there must be a
monitoring and checking of the thing controlled so that useful direction can be given or effective restriction put in place.

As a verb, and in its relation to agency, ‘to control’ equates primarily to dominating and commanding something by directing and guiding its behaviour as one sees fit, but also by restraining and hindering its free action when, and secondarily to monitoring something, maintaining an awareness over it and verifying it against some standard.\(^\text{10}\)

The OED also provides the following definitions for the noun ‘control:\(^\text{11}\)

1. 
   a. The fact of controlling, or of checking and directing action; the function or power of directing and regulating; domination, command, sway.
   
   b. Colloq. phr. *everything(‘s) under control*: all is as it should be; everything is in order.
   
   c. *Radio. Control*, artificial regulation of the dynamic range of a programme output, by means of a control potentiometer, to bring it within the limits of an electrical medium of communication, these limits being determined by overloading on the one hand and under-modulation on the other.

2. Restraint, check. without control: unrestrainedly, freely.

\(^{10}\) Why not call restricting, directing and monitoring the primary, secondary and tertiary aspects of control, respectively, rather than restricting and directing the primary and monitoring the secondary? Quite simply, because control would appear always to appear to involve monitoring *alongside* restricting or directing: either restricting and monitoring, directing and monitoring, or both restricting and directing whilst monitoring. Restriction and direction can come in the absence of one another, and so one aspect of control can be filled by either or both of them. To suggest that all three of these elements were independent aspects would be to suggest that they all had to be present in any given case of control. Of course, it might be argued that restricting really is a form of directing: for instance, I might direct, or guide, the movements of a rat in a maze by restricting its access to all but one route. Whilst this is certainly true, and there is of course some overlap to these elements, I want to illustrate the impression from the definitions that these are two different ways of controlling, and the directive element I have in mind is of a more active sort than the purely restrictive: actually getting involved in the process issuing in a given event (taking over the driving of a car, rather than simply blocking off roads to make you drive where I want).

3.

a. A method or means of restraint; a check. Also, a means adopted, esp. by the government, for the regulation of prices, the consumption of goods, etc.; a restriction (usu. in pl.).

b. In mod. scientific use: A standard of comparison used to check the inferences deduced from an experiment, by application of the ‘Method of Difference’. Often attrib. as in control-experiment, a test experiment devised with this end in view.

c. In automobile racing, a section of the road, usually through a town or village, over which speed is controlled; also, a point on the road or track where officials are stationed and contesting cars are halted for examination and repairs; similarly, a station in aircraft racing. Also attrib.

d. The apparatus by means of which a machine, as an aeroplane or motor vehicle, is controlled during operation; also, any of the mechanisms of a control apparatus, or (in pl.) collectively for the complete apparatus. So dual control: see DUAL adj. and n. Compounds.

e. Bridge. A card which will enable its holder to win a trick in a given suit at a desired point in the play.

f. Computing. In full control unit. That part of a computer which controls the operation of the other units and in recent computers interprets the coded instructions.

4.

a. A person who acts as a check; a controller.

b. Spiritualism. A spirit who controls the words and actions of a medium in a trance

The most relevant definitions of control in its noun form to agency appear to be (1a),(1b),(2) and (3a), casting control, especially (1a), as a power to restrict, direct
and dominate a particular object,\textsuperscript{12} representing the primary aspect of control. (1c), (3b) and (3c), whilst they are not directly applicable to agency, also echo the two aspects present in the verb form definitions: (1c) refers to the \textit{regulation} of the dynamic range of a programming output, \textit{restricting} that output so that it falls within the desired range (thereby relating to the primary aspect of control identified above); (3b) refers to something by which an experiment can be checked and compared (thereby relating to the secondary aspect of control identified above); and (3c) refers both to a section of road over which speed is restricted (thereby relating to the primary aspect) and also to a point at which examination and monitoring can be carried out (thereby relating to the secondary aspect).

The other definitions, (3d), (3e), (4a) and (4b), appear to be largely derivative from the other notions of control discussed above: controls for machines are presumably so named because they can be \textit{used} by some agent to control (in the sense of directing) the operation of the machines to which they are linked;\textsuperscript{13} a person who acts as a control or check is referred to as a \textit{control} because their purpose is to exercise control over some area; in bridge a \textit{control} card plausibly derives its name from the control, in the sense of (1a), it gives its holder over a given hand of cards; likewise, a spirit that controls the words and actions of a medium is a \textit{control} due to its nature of exercising control, in the sense of (1a), over the medium: directing and dominating them.

Control in terms of agency, then, involves:

\textbf{Primarily:} the restricting or directing of actions, events or agents; and

\begin{flushleft}
\textsuperscript{12} I use object in a loose sense here to refer to the \textit{object of control} in any given instance of controlling, which could be an action, agent or event.  \\
\textsuperscript{13} Though, in fact, I think there is also a unique meaning of control relating to this notion; see §II, \textit{Mechanical Control} below.
\end{flushleft}
Secondarily: in large part, the monitoring and checking of those same actions, events and agents (or perhaps more properly, of the processes leading to those actions, events or agent’s behaviour) to inform one’s restricting and directing and ensure that they conform to what one wills or desires.

II – THREE SPECIES OF CONTROL

I think we can distinguish between roughly three kinds of control from the more general definition given above of a restricting, directing and monitoring relationship.

To see this three way distinction consider the following example:

Building Site: A crane operator on a building site sits in the cab of his crane and controls the crane arm with a series of levers on the crane control panel. Throughout the day the operator moves a series of crates from one area of the site to another. The site is run by a foreman who supervises operations from a temporary office on the site.

There are at least three kinds of control present in this example: first, and at the most basic level, there is the relationship that holds between the levers on the crane control panel and the crane arm. The levers control the crane arm in a mechanistic sense: there is a mechanical connection between the two such that moving a lever in such-and-such a manner will move the crane arm in some corresponding manner (though perhaps not vice-versa) allowing the crane arm to be manipulated and directed through the use of the levers on the control panel; call this mechanical control.

Second, the crane operator actually monitors the movement of the crane arm and utilises the levers on the control panel to direct its ongoing movement as he sees fit and towards some goal (in this case the moving of crates around the site). The driver
is the agent of the crane’s movements in this case, they correspond to his will and actions; call this *agential control*.

Finally the foreman supervises the whole operation: he is ‘in charge’ of the site and accountable for all of the goings on located on the premises. He does not directly control the crane arm, nor the crane operator.

Let’s consider these three kinds of control in a little more detail. I start with mechanical and authoritative control before moving on to agential, this last kind being my main focus in this collection.

**Mechanical Control**

I noted in §I above that control can be used, in the sense of noun definition (3d) to refer to “The apparatus by means of which a machine, as an aeroplane or motor vehicle, is controlled during operation; also, any of the mechanisms of a control apparatus.” I reasoned that this terminology is derived from the possibility of *using* that apparatus to restrict and direct the relevant machine. However, I think there is something more to mechanical control than just the possibility of using it to direct some other object.

Consider, for instance, that when we say ‘the moon controls the tides,’ we do not mean that it is restricting or directing the tides; the moon cannot command the tides to raise and lower. Nor do we mean that we could *use* the moon to take control of the tides (as a pilot might utilise the yoke in the cockpit to direct an aircraft). Instead, we mean something like ‘the tides are caused to behave in the way that they do by the position of the moon,’ or simply ‘the moon’s position causes the tides to raise and lower as they do.’
In this way, I think there is a very simple causal way in which we use ‘control’: The moon’s controlling the tides, for instance, amounts simply to the moon’s causing the tides to rise and fall in some regular and reliably pattern based upon its own movements around the Earth. A light-switch controls a light by virtue of the causal connections that hold between the two: if the switch is flipped the light will come on. There is undoubtedly still some link with utility in some cases, though with an emphasis on the causal connection that underpins that utility.

So, for instance, when we ask what controls a light in a room, we are asking `what mechanism can be used to cause the light to turn on or off?’ When we ask what controls an automatic door's opening when we walk near it, we are asking something like: `what mechanism causes that door to open when we walk towards it?'; and when we ask, `How do I control the wing flaps' we mean something like: `what mechanism can I use to cause the wing flaps to open or close?’

What about a rock, though: suppose that one picks up a rock which can be reliably used to turn off mobile phones (by breaking them), is that rock thereby a mechanical control for mobile phones? Presumably, we would want to say ‘no.’ What, then, aside from the propensity to reliably cause some change of state is necessary for mechanical control?

Perhaps we need to tighten our conditions for something’s being a mechanical control: we might try adding the condition that the implement be specially designed for controlling the relevant object. So for instance, the levers on the crane’s control box in crane 1 above have been designed specifically to move the crane’s arm in various ways. The rock, on the other hand, was not designed at all, let alone designed with the function of turning phones off.
Imagine now, though, that someone invents a phone-mallet, a large rubber mallet, specifically for turning a phone off by hitting it (and breaking it). Would this phone mallet count as a control for all phones because it has been designed specifically to turn them off (by breaking them)? Again, I am inclined to think that we will say ‘no.’

The problem with these cases, is that the stone and the phone-mallet do not interface with the devices that we are imagining they might control in the right sort of way: perhaps built into the notion of a control apparatus is that it achieves this control through a certain relevant medium, or that it brings about these changes of state in a way relevant to the controlled object itself: a phone is not typically designed to be turned off by being broken, so something that can only cause a phone to turn off by breaking it will not count as a control for a phone. On the other hand, an atomic warhead is designed to explode, so something which can reliably be used to cause it to explode may well count as a control for that warhead.

That is not to say that there is not some sense of control according to which we can use sense in the way described here: Jane might control Joe (in the sense of intentionally directing his movements) by hitting him with a stick, hitting his left side to make him go right and on his right side to make him go left; but Jane’s stick is not a mechanical control for Joe: the stick is not connected in a mechanical fashion to Joe; swinging the stick to the left or right may or may not cause Joe to move right or left, depending upon where it is swung. Contrast this to the controls on a Boeing747, the button controlling the wing-flaps will reliably cause the relevant change in the wing-flaps if pushed (so long as the mechanism is operating correctly).

It is not entirely clear how we should finalise the conditions for being a mechanical control of some x, I think that this will be determined in some way by how
we define a mechanical connection between two objects, and more work may need to be done here. This falls outside of the scope of this thesis on *agential control* however, so this sketch will suffice for now.

**Authoritative Control**

Let’s turn our attention now to the foreman; we might say: the foreman is in control of the goings on of the site, and hence in control of the crane and its operation. He is not, however, in control of the crane in the same way as the operator: the operator actually guides the crane arm using the controls in the crane’s cab, he monitors the crane arm as it moves and adjusts his own manipulation of the controls accordingly to direct it in the way that he wants. The foreman, on the other hand, is sat in the site office and may not even be watching the crane at all.

Effectively, the foreman is in control of the crane (and the rest of the site) in the sense that he is *in charge* of it and its operations and has the authority to dictate what operations the crane should and should not perform, whether or not he actually exercises that authority. What’s more he is *responsible* for the goings on of the site, even if he does not exercise control: if there is an accident on the site or company property is damaged he will be held accountable, perhaps more so if he was not monitoring the site as he ought to have been.

Interestingly, this suggests that an agent, or group, with authoritative control over some $x$ can fail to control $x$ and still maintain their authoritative control: that is, agential control (as exercised by the crane operator) implies success—if you have agential control then you are actually regulating or directing something—whereas authoritative control does not. For instance a department of government, say a tax regulating body, might be said to be *in control* of regulating the tax laws in the country
and yet fail to appropriately regulate those laws, or even fail to regulate them entirely; the body would remain, officially, the department in control of regulating tax, in that it still has the authority to exercise control, even if it fails to do so appropriately. One only loses authoritative control over something when one no longer has the authority to ‘step in’ and control it, should the need arise.

This kind of control plausibly has links to moral responsibility (perhaps more so than to freedom): we can imagine that parents who have authoritative control over their children, for instance, will be held morally responsible for how their children behave whether or not they actually succeed in controlling them. That being said, I consider this a kind of control less relevant to agency and it shall not therefore be my primarily concern in this collection.

Agential Control

Now let’s turn our attention to the crane operator: the operator exercises control over the crane by monitoring it and directing it, thus meeting both the primary and secondary aspects of my gloss above. He pays close attention to where the crane arm is, where it is going and at what speed and he utilises the controls along with the knowledge gained from that attention to guide the crane arm in the way that he wants. If the crane arm starts to move in a way that he does not intend, he is poised and ready to alter his manipulation of the controls (or perhaps just stop manipulating them altogether) so as to bring the crane arm’s movements in line with his intended guidance.

Notice that unlike the foreman, who has authoritative control, if the operator stops paying attention to the crane arm and loses track of its position he will effectively lose agential control over it. He might retain some sort of mechanical control over it.
in that his moving the levers on the control panel will cause some corresponding movement in the crane arm, but he will no longer be guiding it: how could he, when he is unsure of its position?

To see this point, imagine that you need to guide someone by phone to a certain location in a city centre, say the town hall. You have a map of the city in front of you and can quickly plot a route for them from any street in the city centre. However, the person on the phone does not know where they are. What is the first thing you will tell them to do? Look for some landmark or street-name so that you can work out where they are currently (and possibly which way they are currently travelling); it would be impossible to generate a route along which to direct them unless you first know where they are. Similarly, when exercising agential control over something, one needs to maintain awareness over that thing so as to have a hope of restricting and guiding it in the way that one sees fit.

Before moving on to the sketch analysis of agential control it will be worth briefly running through some of the ways in which we think of agents as exercising agential control in the philosophical literature; for instance, agents can exercise self control, moral control, direct control, indirect control, conscious control and automatic control. A lot of these distinctions cut across one another, being generated in some cases by the particular object of control (in the case of self-control, for instance) and in others by the standards by which an agent directs their actions (in the case of moral control) or the manner in which control is exercised (in the cases of direct, indirect, conscious and automatic control). Whilst I think these subsidiary notions of control are important and do have a role to play in philosophical discourse,

---

14 This list is not intended to be exhaustive.
I think that our account of control ought first and foremost aim to tackle the more fundamental notion of agential control on which they are built.

**Self-control**

We have already seen above that a common usage of ‘control’, in the philosophical literature at least, is in the term ‘self-control’, which is largely used as a contrary to akrasia: one has self-control, according to this reading, just so long as one’s behaviour conforms to what one judges one ought to do.

This strikes me as a relatively uninteresting usage of control and not an instance of agential control proper: one’s behaviour could conform to one’s judgements and yet be out of one’s control in the true agential sense. There is room, that is, in one’s behaviour conforming to one’s judgements for cases of causal deviance, which we have already seen appear to constitute instances of lacking agential control. For instance, a man hanging from the ledge of a burning building might judge that he had best let go of the ledge and fall to the firemen’s safety blanket below; this judgement might so unnerve him (with the realisation of what he must do) that his grip loosens causing him to falls down into the safety blanket. The man’s behaviour conforms to his judgement, but he hardly controls his letting go of the ledge.

In order for self-control (of this contra-akratic kind) to be a form of agential control, the agent must actively direct their behaviour *in accordance* with their judgements.

For a more interesting notion of self-control, we might consider Holton’s (2009, Ch.6) conception of *strength of will*, which he characterises as not reconsidering or abandoning resolutions that we have made to act (a resolution, according to Holton, is a special kind of intention resistant to reconsideration that we make in order to ensure that we carry out a certain plan of action, even though we
accept that we might not want to carry out that plan, or judge doing so best, when the time comes (Holton, 2009, p. 10)). If self-control is conceived of as simply ‘sticking to one’s intentions in the face of contrary inclinations’, however, then as with the first gloss of the contra-akratic self-control above, I think this would fall short of agential control and allow for wayward causal chains (I might stick to my intention to not-A-at-t in the face of a desire to A-at-t, but only because my reflecting on my intention distracted me and caused me to miss my chance, for instance).

In order for strength of Will to be a case of agential control, as with contra-akratic action, the agent must be the one to direct their sticking to their resolutions.

**Moral control**

Moral control might be characterised as ‘the kind of control required for being morally responsible.’ This conception of control is evident, though certainly not exclusively found, in the Fischer-Ravizza account of guidance control and is, in effect, a strict form of agential control: one in which moral reasons must play a role in directing and or restricting an agent’s behaviour.

Briefly, Fischer and Ravizza think that an agent is morally responsible for performing some action if and only if they had guidance control over that action’s performance, or exercised guidance control at some relevant point along the way to performing that action (1998, p.241, fn.2). Having guidance control consists in two elements: the mechanism issuing in the action must be suitably responsive to reasons, and the mechanism must be properly the agent’s own (the agent must have appropriately taken ownership of that mechanism). The clues that guidance control is really this stricter moral form of agential control can be found in the conditions for a mechanism’s being what they call moderately reasons-responsive (the appropriate level of responsiveness for control according to Fischer and Ravizza):
…holding fixed the operation of a K-type mechanism, the agent would recognize reasons (some of which are moral) in such a way as to give rise to an understandable pattern (from the viewpoint of a third party who understands the agent’s values and beliefs) and would react to at least one sufficient reason to do otherwise (in some possible scenario).

(1998, pp. 243-4)

An agent needs to be able to recognize and then respond to moral reasons, according to the above criterion, in order to be exercising guidance control over their behaviour. This is largely due to Fischer and Ravizza’s aim in introducing the notion of guidance control: the Fischer-Ravizza view is a semi-compatible one that purports to demonstrate that moral responsibility is compatible with determinism. As guidance control is necessary and sufficient for moral responsibility, according to the view, of course it must involve an agent’s recognizing moral reasons. I think, though, that this would be far too heavy a constraint to place upon agential control more generally, which is supposed to capture the connection between agent and action when they are genuinely guiding their behaviour. Receptiveness and responsiveness to moral reasons very plausibly are necessary for being a moral agent and morally responsible for one’s actions. How, after all, could we expect an agent that is incapable of recognizing moral reasons of behaving morally? However, such an agent could still be expected to exercise control over their behaviour using the reasons that they are able to recognize.

Imagine, for instance, a psychopath or amoralist, call him Patrick, who is simply incapable of recognizing moral reasons;\(^\text{15}\) when Patrick performs a morally

---

\(^{15}\) This view of psychopaths, as incapable of recognizing moral reasons has recently been challenged; according to (Cima, Tonnaer, & Hauser, 2010) psychopaths appear to understand the difference between right and wrong but do not care about such knowledge and are not motivated by such considerations. It is not clear whether we ought to classify knowing that something is right or wrong in the absence of being appropriately motivated by the knowledge as ‘recognising a moral reason;’ that is,
reprehensible act he is not guided by moral reasons of any kind: they neither motivate nor restrict him. However, that should not mean that he is incapable of directing his behaviour entirely: Suppose that Patrick decides that the quickest route to getting a promotion at work is to kill his boss; he meticulously plans and executes (no pun intended) the murder with the highest degree of precision. Patrick cannot recognise moral reasons, and this goes some way to explaining why he considered killing his boss a viable option for raising the chances of his promotion, but this does not prevent him from directing his behaviour in a highly organised and intellectual manner in accordance with the reasons that he does recognise.

Even agents who are able to recognise moral reasons can sometimes guide their behaviour without appealing to moral reasons: suppose you have to cross the street; you press the crossing-button, wait for the walk-signal to go green and then stroll across the road. Was your crossing the street guided by you? It seems obvious that it was. Was your crossing the street guided also by moral considerations? Presumably not: moral concerns rarely arise in mundane tasks such as crossing the road.

Further, why should moral considerations have such an important role in determining the character of agential control? Consider an aesthetically-challenged, though highly moral agent, who is able to recognise and respond to moral reasons but

recognising a moral reason might inherently involve being appropriately motivated by that reason, where if one is not morally motivated one has not really recognised a moral reason.

Smith (1995), for instance, has argued that a requirement on having a mastery of moral terms is that “if an agent judges that it is right for her to φ in circumstances C, then she is either motivated to φ in C or she is practically irrational” (1995, pp. 61 & 71-76). An amoralist might have a facility for moral language, and be able to reliably point out right and wrong actions based upon their observations of other agents’ judgements, but as they are neither motivated by their moral assessments, nor practically irrational, they do not have a mastery of moral terms and therefore do not really make moral judgements (1995, p. 76). Similarly, we might think that without a mastery of moral concepts, an amoralist would be unable to genuinely recognise moral reasons; just as a colour-blind person who can reliably point out red objects as ‘red’ is unable to genuinely recognise ‘red.’
is incapable of recognising aesthetic reasons. Should we say that they lack agential control? Presumably not, they are perfectly capable of guiding their behaviour, just not in any aesthetically pleasing way.

Moral control, then, is a particular kind of agential control, one requiring moral reasons to play a central role in an agent’s directing their behaviour; agential control remains the more fundamental kind of control, however, and the one more interesting in terms of agency and freedom.

**Conscious and automatic control**

A common distinction that we might draw between exercising control is being in control of something consciously and having control over something in an automatic sense.

Consider Wittgenstein’s famous question: what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?\(^{16}\) I think we can ask a more telling question: what is the difference between my raising my arm due to an electric shock, say, my idly raising my arm to scratch my head, and my consciously raising my arm to draw my name in the air with my finger.

Working backwards: in the third case I consciously guide my arm’s movement, monitoring it as I do, aware of its relative position directing it as required; in the second case I do not consciously guide the motion of my hand, but if necessary I can intervene in its motion and ‘take over’ its guidance from my sub-conscious (or whatever sub-personal mechanism might be guiding it); finally, in the first case I do not consciously guide my arm’s movements and nor could I guide it: I cannot intervene in the motion and curtail the motion caused by the electric shock.

---

\(^{16}\) Wittgenstein 1958, §621
In the first case I clearly have no control over my arm: I do not direct its movements myself and am unable to restrict or abort its raising even if I so wanted.

In the second case I have automatic control: my behaviour is being directed by some sub-conscious, or perhaps sub-personal, psychological mechanism that is a part of me in a broader sense of self: that is, I can be identified with just my conscious will (narrow sense of self) or I can be identified with a whole host of psychological, and perhaps also organic, mechanisms that operate within my body (broad sense of self). So I might control my behaviour automatically in the broad sense whilst not controlling it in the narrow sense.

In the third case I have conscious control: my arm’s movement is directly guided by me in a process of continual monitoring of its position and movements so as to inform further active guidance.

The difference between the second and third case, then, is in the mechanism of mine that is directing my arm’s raising: in the second case it is a sub-conscious mechanism of mine that directs my behaviour, whereas in the third it is my conscious deliberative self that does the directing.

**Direct vs. Indirect Control**

Finally, we sometimes refer to agents exercising control directly and indirectly; that is, an agent might be exercising, or have, control over something here and now, or they may be able to exercise control now so as to maintain control later when they otherwise might not.

To elaborate, consider a man with anarchic hand syndrome, a condition in which one’s hand will often act without ones conscious direction,\(^{17}\) who straps his arm

---

\(^{17}\) See, for instance, (Della Sala, Marchetti, & Spinnler, 1991) and (Krikitos, Breen, & Mattingley, 2005).
down now so that later at dinner his anarchic hand will not steal food from the plates of other dinner guests (a frequent complaint of anarchic and alien hand patients). In doing so, the man exercises indirect control over his anarchic hand so as to curtail its movements later when he will be unable to directly control it.

As with conscious and unconscious control above, these are not a different kind of control, but rather represent two different manners of exercising agential control over something: presently, for direct control, or in view of maintaining control at a later time, for indirect control.

Indirect control especially potentially has a large role to play in a theory of morally responsible action based upon control of action: agents are frequently held accountable for actions that they performed when they were not capable of properly controlling themselves precisely because they did not indirectly control themselves by taking various measures at a previous time.

A perfect example of this is drink driving: an agent who hits a pedestrian whilst drink driving is almost universally considered to be morally responsible for doing so; but they may have been so drunk at the time that they were unable to genuinely direct their car in a manner sufficient for having agential control over it or whether they hit the pedestrian. Now generally a lack of control over whether something occurred entails a lack of responsibility for that event’s occurring; however, in this kind of case what seems important is that the agent should have indirectly controlled themself and taken measures to prevent their drink driving: leaving their car at home, organising a lift, or simply not drinking knowing that they might decide to drive later whilst drunk.
It is potentially due to their indirect control over their situation that leads us to think of them as morally responsible.\textsuperscript{18}

\section*{III – SKETCH FOR A THEORY OF AGENTIAL CONTROL}

Having identified \textit{agential control} with, roughly, restricting, directing and monitoring something in this section I shall try to lay out the sketch for an account of this kind of control. I begin by noting that there exist two \textit{modes} of agential control, an active and a passive, before moving on to look at the conditions we might place on being in control of some action, event or agent before and then, finally, presenting my sketch for an analysis of agential control.

\textbf{Controlling vs. being in control of \textit{whether}}

The first thing to notice about agential control as it applies to actions, agents and events, is that it can come in two modes: an agent can be \textit{controlling} some action, agent or event, or they can be \textit{in control of whether} they act in such-and-such a way, some other agent behaves in such-and-such a way, or some event comes about. I concentrate here solely on control over actions, but what I say can be extended to agents and events (and I suggest how to do this in §III – \textit{A Sketch}).

\textit{Controlling} some $x$ involves actively restricting or directing that $x$ as one sees fit. So, for instance, an agent behind the wheel of a car might come to a turn she needs to make, slow the car down, indicate, turn the wheel and guide the car through the turn. As she guides the car she is constantly monitoring its speed and position relative

\footnote{Fischer and Ravizza make this point about guidance control also, though not in terms of indirect control: an agent can be responsible for some action that they perform (drink driving, say) whilst lacking guidance control just so long as they exercised guidance control in getting themselves into that state. So an agent might exercise guidance control in getting themselves drunk, well aware that this will diminish their ability to control themselves later on (1998, p. 195).}
to the curb and the road so as to continuously update her directing of the car and make
minor adjustments to its heading as and when they are needed.

An agent is in control of whether some $x$ behaves in such-and-such a way, on
the other hand, if they are able to bring it about that $x$ either does or does not behave
in such-and-such a way. So, for instance, our driver guides her car through the turn,
but she was able at any time to abort this process and either stop the car or guide it
somewhere else instead. She is able, in that case, to determine whether or not the car
is guided around the corner.

Interestingly, having control over some $x$ in one mode, does not entail having
control over that $x$ in the other; to illustrate this, consider the following pair of cases:

**Crane 1:** A crane operator is sat in the operation box of a crane and
uses the levers and buttons that make up the controls to guide the
movements of the crane and load a crate onto a cargo ship, all the while
he surveys and keeps track of the crane's movements.

The crane operator is *controlling* the crane in virtue of his engaging in a process of
continually monitoring the crane's position and movements whilst directing those
movements using the controls.

As well as being in a process of controlling the crane he is also in a state of
being in control of whether the crane arm moves in this way or that: Using the controls,
the operator can bring it about that the crane moves left, say, or does not move left.

A simple view of the relationship between controlling $x$ and being in control
of $x$ might go as follows: $S$ is in control of $x$ iff $S$ is controlling $x$. The following
example shows this simple view to be incorrect, however:
Crane 2: A crane operator is sat in the operation box of a crane. The crane arm is being guided by a computer program which loads crates onto a cargo ship nearby. All the while the crane operator surveys and keeps track of the crane's movements. If at any time he thinks that the crane's movements need to be adjusted or stopped entirely he can engage the controls in front of him to override the computer program and manually guide the crane.

In this example the crane operator is not *controlling* the crane arm as the operator did in crane 1: the computer program is (consider if we asked "who (or what) is controlling that crane?" the appropriate response would seem to be "a computer program" rather than "the crane operator" given that he is not doing anything to actively guide its movements). However, the crane operator is still 'in control of the crane' in that he is in control of whether the crane operates in the way that it does under the computer program’s guidance: the operator is appropriately placed to ‘step in’ and take over should the need arise.

Being in control of whether some action or event comes about, then, involves being placed appropriately to bring that action or event, or prevent it from being brought about, through taking control of the process issuing it should the need arise. This too will involve a process of monitoring x, but in the case of being in control, the emphasis is on being able to effectively control (restrict or direct) x, and having the knowledge that one can do so along with the corresponding know-how.
The computer program is also ‘in control’ of the crane arm, in so far as a computer can have agential-type control, though perhaps to some lesser degree given that its executive power over the arm can be overridden by the operator.

Before moving on, note that this distinction between the two modes of agential control—controlling and being in control of whether—raises the following point about the phrase ‘being in control:’

Describing an agent being in control of some x turns out to be a general phrase that could point either to their controlling that x or their being in control of whether that x operates in some way or issues in certain behaviours. We can see this ambiguity by considering the following set of cases:

Mastermind 1: An evil scientist, known as ‘the mastermind’, has planted a device in the brain of Adam, a high ranking US official, in order to control his behaviour as a part of a larger plot to overthrow the government. The device allows the mastermind to take over the physical movements of Adam’s body whenever he so wishes and drive Adam’s body in the same way that you or I might drive a car using specially designed equipment in his laboratory. The mastermind sets about using Adam’s body to plant incriminating documents about the President in the CIA’s databank. Whilst the mastermind’s device is active Adam loses all abilities to consciously direct his bodily movements.

---

19 We might think here that the computer program does not have agential control over the crane arm, but rather plays the role of an agent in what would standardly be a relationship of agential control. We might want to say, if that is the case, that the computer program simulates or mimics agential control over the crane arm.

20 These cases are roughly inspired by Stump’s (2002, pp. 47-8) ‘alien master’ example, which is itself inspired by Robert Robert Heinlein’s The Puppetmasters.
In this case, the mastermind is in control of Adam in the sense that he is controlling his body: he takes over Adam’s motor functions and guides Adam’s body almost as if it were his own.

Mastermind 2: An evil scientist, known as ‘the mastermind’, has planted a device in the brain of Adam, a high ranking US official, in order to control his behaviour as a part of a larger plot to overthrow the government. Adam has previously been convinced by the mastermind’s co-conspirators to plant incriminating documents about the President in the CIA’s databank. The mastermind uses his device to monitor Adam’s mental states and is able, if Adam wavers in his intention to plant the information, to force Adam to carry out his original plan of planting the incriminating documents. Alternatively, the device can be used to kill Adam with a high voltage shock to the brain, in the event that the conspirators need to abort the plan. As it turns out, Adam does not waver in his intention to plant the documents, and the conspirators do not opt to abort the plan. Whilst the mastermind is monitoring Adam though the device, Adam is able to consciously direct his body as normal.

In this second case, it seems right to say that mastermind-2 is in control of Adam, indeed we can imagine that this is what that mastermind-2 tells his co-conspirators (“don’t worry about Adam, he’s under my control now”). However, he is not controlling Adam in the way that mastermind-1 was; rather, mastermind-2 is in control of whether Adam performs the particular actions that the co-conspirators wish him to
perform: mastermind-2 can ensure, in fact, that Adam plants the incriminating documents as per the conspirators’ plans, or that he does nothing (by killing him).\textsuperscript{21}

\textit{Mastermind 3:} An evil scientist, known as ‘the mastermind’, has planted a device in the brain of Adam, a high ranking US official, in order to control his behaviour as a part of a larger plot to overthrow the government. The mastermind uses the device to systematically implant intentions into Adam’s mind and guide his planting incriminating documents about the President in the CIA’s databank. Once an intention has been implanted in his mind, the mastermind leaves the altered-Adam to carry out the intended actions of his own accord.

In this final case, we might say that mastermind-3 is \textit{controlling} Adam’s behaviour by implanting various intentions into his mind and also that he is \textit{in control of whether} Adam performs various actions, in that he can implant an intention into Adam that will be effective in bringing about whatever action he wants. However, Mastermind-3 is not \textit{controlling} Adam’s behaviour in the same way as Mastermind-1: Mastermind-1 \textit{drives} Adam’s body as you or I might drive a car, operating his limbs with external controls and actively guiding them through their movements. Mastermind-3 does not actually guide the precise movements of Adam’s limbs, he allows Adam to control that; rather he guides Adam as a whole by implanting intentions into his mind.

\textsuperscript{21} We might imagine a case more like \textit{Black and Jones} in which the mastermind can only use the device to ensure that Adam brings about the planned actions, but not prevent him from doing so. In such a case, the mastermind \textit{would not} be in control of whether Adam performs the planned actions, and hence not in control of Adam. This might seem counter-intuitive at first, but imagine that in this case the co-conspirators \textit{do} change their mind and want to abort: “you said you were in control of him!” they might say to the mastermind, but of course whether Adam plants the incriminating documents or not is somewhat up to Adam alone at this point, the mastermind can only hope that Adam changes his mind and then \textit{not} activate his device; that hardly sounds like control.

Of course, if the mastermind \textit{does} activate his device when Adam wavers in his intention he will be \textit{controlling} Adam by forcing him to perform the planned actions, so in this imagined case the mastermind is still able to control Adam, but only to one end.
When we say that Mastermind-3 is in control of Adam then, we might mean that he is both: i) controlling what Adam does (in a broad sense: he controls what actions Adam performs but not the performance of them), and ii) in control of whether Adam performs various actions, in that he can prevent him from performing some action A by implanting in him the intention to do not-A. Adam, or at least altered-Adam remains in control of his behaviour in a somewhat narrow sense: he actually guides the movement of his body in accordance with the intentions the mastermind implants in him.

Two Frankfurrian Intuitions

What does control, in either of the above modes, consist in?

I have already settled on the rough characterisation of control as primarily restricting and directing and secondarily monitoring, but what are the criterion for meeting these two aspects? What, that is, does restricting, directing and monitoring something involve? I concentrate on the primary aspect of control in the next few sections before moving on to the secondary in §III – Awareness.

I think that Fischer and Ravizza are right to suggest that control is tied up somehow with guidance: there is something right about the Fischer-Ravizza intuition that controlling something involves actually guiding that action through its performance, with or without the opportunity to do anything else. I do not think, however, that this means that control does not require some kind of ability to do otherwise (in fact, as I’ll note in §V, Guidance Control and the Anti-Akrasia Chip below, I think that once we are careful to distinguish between what I shall call metaphysical and counterfactual abilities, we can see that on Fischer and Ravizza’s analysis guidance control requires a certain kind of ability to do otherwise).
Fischer and Ravizza, of course, base their intuition about control and alternate possibilities on a Frankfurtian intuition from (Frankfurt, 1969); I think there is a second Frankfurtian intuition on guidance that we ought to adopt in our theory of control. I take these two intuitions in turn:

**No alternatives needed**

The first Frankfurtian intuition is a familiar one, and one championed by Fischer and Ravizza in their (1998) ‘Responsibility and control:’ specifically, that in order to be morally responsible, one need not be able to do otherwise.

This intuition is easily demonstrated by Frankfurt cases, take this shortened version of the classic Jones and Black case discussed above:

*Black and Jones:* Black, an evil scientist, implants a device in Jones’s brain. If Jones wavers in his intention to kill the Mayor, the device will be activated, forcing Jones to remain faithful to his original intention. As it turns out, however, Jones murders the Mayor and the device remains inactive.\(^{22}\)

Now the original intuition aimed at by these cases is that an agent can be morally responsible for some action even though they are unable to do otherwise. However, we can easily apply this case to control also: it seems quite clear that Jones, and not Black, controls his shooting the mayor; Jones is not coerced or compelled, he acts upon his own practical deliberation mechanism without any tampering from Black, and yet he is unable to do otherwise. Frankfurt cases also appear to show, therefore, that control does not require an ability to do otherwise.

---

\(^{22}\) This version of Black and Jones is taken from (Whittle, 2010, p. 8).
**Effective Intervention**

The second Frankfurtian intuition that I think ought to inform our account of control comes from his (1978) ‘The problem of action:’

A driver whose automobile is coasting downhill in virtue of gravitational forces alone may be entirely satisfied with its speed and direction, so he may never intervene to adjust its movement in any way. This does not show that the movement of his automobile was not under his guidance. What counts was that he was prepared to intervene if necessary and that he was in a position to do so more or less effectively.

(1978, p. 160)

The intuition here, then, is that in order for something to be under one’s guidance, one need not be actively involved in the issuing of that thing, be it an action or event; rather, one need only be able to effectively intervene in the process bringing that something about.

Now, the rough definition of control that I settled upon in §I above was: primarily, restricting and directing some action in accordance with some standard, and secondarily the monitoring of the putatively controlled object to inform ongoing restriction and direction. The primary aspect, I submit, involves placing an object under one’s guidance. That being the case, if we accept the second Frankfurtian intuition, control involves the ability to effectively intervene in the putatively controlled process.

These two intuitions conflict, though: on the one hand, we have seen that control does not require an ability to do otherwise, and on the other we have seen that
it involves having an ability to effectively intervene (which essentially comes down to an ability to do otherwise).

Metaphysical and Counterfactual Access

I think the answer to the tension alluded to above between the two Frankfurtian intuitions is that they refer to two different modes of control, and hence two different kinds of ability: In the case of the *no alternatives needed* intuition, when we say that an agent does not need to be able to do otherwise to have control we mean that in order to be controlling their behaviour, they do not need the opportunity to perform any other action; that is, given their current situation, they need not be able to actually do anything differently. Think here of Jones in Frankfurt’s classic case: Jones *controls* his shooting of the mayor in the sense of guiding and directing it, whilst not being in control of whether he shoots the mayor, in that he has no opportunity to do otherwise. However, he still maintains his capacity, or capability,\(^{23}\) to do otherwise as the mechanism on which he acts (his deliberation about whether to shoot the mayor) is one that is perfectly capable of issuing in alternative actions given the presence of the right reasons.

In the case of the *effective intervention* intuition, when we say that an agent needs to be able to effectively intervene in some process in order to control it, we mean that in order to be *in control of whether* that process issues in this or that action or event, they need to have the opportunity to bring about two or more actions or events (through refraining from acting or intervening). Think here of the crane operator from *Crane 2*: the operator is not directly controlling the crane arm, rather they allow the computer to direct its movements, but at any time they are able to effectively intervene

\(^{23}\) I use these two terms interchangeably.
in the motion of the crane arm and move it in the way that they see fit. They are thereby in control of *whether* the crane arm goes left or right, back or forwards, and so forth, in that they have the opportunity (as well as the capacity) to effectively intervene in the crane’s operation (using the control panel) and bring about the movements they see fit.

Now, causal determinism *would* remove an agent’s ability to do otherwise, without necessarily removing their capacity to do otherwise: it might still be true, for instance, that if they were to try to do otherwise that they would succeed in doing so. That being the case an agent might remain able to do otherwise, even if they have no opportunity to do so; indeed, this has been the strategy taken by many compatibilists in the face of worries about the compatibility of freedom and determinism (Moore, 1912; Ayer, 1954) and, more recently, against the threat to the principle of alternate possibilities from Frankfurt cases (Fara, 2005; Vihvelin, 2004; Smith, 1997 & 2003). This same strategy, I think, needs to be employed in our theory of control: when we are talking about *controlling* some action, say, we are interested in an agent’s having the capacity to effectively intervene in their behaviour; when we are talking being in control of *whether* they perform some action we are interested in an agent’s having the opportunity to effectively intervene in their behaviour.

We might try to bring the difference between opportunities and capacities out by distinguishing between two types of access that agents can have to possible actions: metaphysical and counterfactual.24

---

24 In fact, I think that for the purposes of control, both opportunity and capacity will require counterfactual but not metaphysical control, as I go on to discuss in §III, *Contextualism in Control* below.
Say that an agent has metaphysical access to some action A, if from their current circumstances they can actually end up in a world in which they A; that is, given their own causal history, they have the power to bring it about that they A. Now, every agent has metaphysical access to the actions that they do, or will, in fact perform, in as much as those actions will have been accessed from their own causal history.25

In order to have metaphysical access to alternate possibilities, however, one requires access to more than one possible action from just one causal history. Metaphysical access to alternate possibilities, then, is effectively the same as having what Fischer and Ravizza (1998) call ‘regulative control’: the dual power to perform some action A or to perform some other action B instead.

25 In this sense, one has metaphysical access to the actions one actually performs in just the same way that, according to Mele (2003), one has the simple ability to perform all and every action that one does and will perform.
So if at \( t_0 \) an agent is presented with a choice between doing either A or B, makes a decision between A and B at \( t_1 \), and completes their chosen action at \( t_2 \), the possible world in which the agent does A at \( t_2 \) and the possible world in which the agent does B at \( t_2 \) will have a shared segment from \( t_0 \) to \( t_1 \), or if you prefer, they will be two worlds that have ‘branched off’ from the one world running from \( t_0 \) to \( t_1 \). In figure 1, point x at \( t_1 \) forms a node at which one possible world diverges into two.

The agent at \( t_1 \), then, has genuine metaphysical access to two possible worlds; and which world they end up in will be determined by the choice they make at point x. This strikes me as a very strong reading of an ability to do otherwise: it is the agent’s choice alone that determines which world they end up in, and that choice is itself not determined by the preceding causal history (else we would not end up with the branching world view). Clearly, an ability involving this kind of metaphysical access \textit{would} be ruled out by determinism (and also by the counterfactual interveners such as Black in Frankfurt cases) in that each causal history would have at most one possible future dictated by the laws of physics and any given state of that world at a time.

Consider now though a weaker notion of ability that is apparent in conditional analyses of abilities: according to a conditional account, an agent has the ability to A just in case they would A if they tried to A. In a possible world framework I suggest we read this as: an agent S is able to A at \( t_1 \) just in case there is a possible world in which they try to A at \( t_1 \) and consequently A at \( t_2 \). Of course, that possible world could be the actual world, or it could be a non-actual possible world with a different causal history.

This weaker reading involves having what I shall call \textit{counterfactual access}: in a world in which I try to A at \( t_1 \), and then A at \( t_2 \) I have counterfactual access to B-
ing just in case there is some other possible world in which I try to B at t₁ and B at t₂ where those events may be entirely determined by the preceding causal history from t₀ to t₁ (see fig.2). It is not, then, because I could access a different possible world from the same causal history that I am able to do otherwise in this weaker sense, but because I would have done otherwise given some relevant changes to the history of my own world (evaluated by considering a distinct world in which those changes are realised).

The truth of causal determinism does indeed rule out metaphysical access to alternate possibilities—in a deterministic world the past and laws of nature will jointly determine whether an agent chooses to A or chooses to B, and whether they are subsequently successful in their A-ing or B-ing, leaving no room for diverging possible

---

26 Some might prefer this claim to involve a counterpart of mine acting differently in different possible worlds. I shall not make any assumptions about the correct way to understand identity across possible worlds. Instead, I invite the reader to interpret these claims in the manner they prefer.
worlds. However, that an agent has counterfactual access to alternate possibilities looks to be compatible with determinism: in figure 2, for example, the two worlds depicted could both be deterministic in nature: that the agent performs A in one and B in another is down to a difference in the segments of both worlds from \( t_0 \) to \( t_1 \) that precede the decision made at \( y \) and \( y_1 \).

If agents in Frankfurt cases do have abilities to do otherwise, then, it is with respect to their having counterfactual, rather than metaphysical, access to other possible actions.

Now, given the above definitions, an agent with metaphysical access will also have counterfactual access to alternate possibilities. Consider the agent who has just A’d at \( t_2 \) in figure 1: there is a distinct possible world in which they tried to B at \( t_1 \) and subsequently B’d at \( t_2 \), it just happens that that world and their own shared a segment in their causal history. Figure 2, on the other hand, represents an agent with only counterfactual access: In each world there is just one action that they will end up performing (as dictated by the causal history of that world and the laws of nature, say) and there is no shared segment between the two possible worlds that could grant metaphysical access to either world from a shared past—the segments from \( x,y \) and \( x_1,y_1 \) are distinct non-identical histories—so there is no chance of metaphysical access to alternate possibilities. Nonetheless, the agent A’s in one world and B’s in the other, due to a relevant change in the causal history, and so they have counterfactual access to those two possibilities.

As noted above, we might think of this difference between metaphysical and counterfactual access as one of opportunity and capability: in saying that an agent has metaphysical access to actions A and B (and is therefore able to A or B in the strong sense) we are attributing to them the opportunity to actually ‘end up’ at one of those
two options, so to speak, from their current situation given the laws of nature in their world; call this *absolute-opportunity* (in that one can actually end up in more than one causal history). In saying that an agent has counterfactual access to actions $A$ and $B$ (and is therefore able to $A$ or $B$ in the weaker sense) we are attributing to that agent the *capacity* or *capability* to issue in those actions given certain inputs by way of their particular causal history.

To illustrate this further consider two coffee machines, *Meta*-machine and *Counter*-machine: *Meta* only takes £1 coins and will dispense either a black or a white coffee determined by the user. When we say that it is able to produce a black or white coffee (in so far as coffee machines are able to do anything) we mean something like: I can put in a £1 coin and receive a black coffee, or I can put in a £1 coin and receive a white coffee. The same input, excluding some button pushing by the user, can lead to two different outcomes. This is roughly the picture for metaphysical access: one input can actually lead to two or more different outcomes.

*Counter*, on the other hand, is a good deal older than meta-machine and takes either a ten pence piece or a shilling. If you put in a ten pence piece, *counter* will produce a black coffee whereas if you put in a shilling it will produce a white coffee. Now, assuming for the sake of this example that due to decimalisation of UK current shillings no longer exist in any form, there is a clear sense in which *counter* is unable to issue white coffees: You don’t have the required input and feeding it a ten pence piece will only over lead to its producing a black coffee. However there is also a sense in which it clearly is able to issue in both white and black coffee: nothing about its internal workings have changed since shillings were demonetarised, it is still stocked with the requisite ingredients, if you had a shilling to put in then it would produce a white coffee; so it is capable of producing white coffees, there is just no opportunity
for doing so. This is roughly the picture for counterfactual access; an agent that A’s may have been capable of issuing in B instead, given some relevant change to their causal history, but might still have lacked the opportunity to express that capability (due to determinism or Frankfurtian-style counterfactual interveners, say).

One final point to consider here: I have said that having counterfactual access to an alternate possibility A involves having the capacity to perform A given a relevant change in one’s causal history; it is not clear just what might count as a relevant change, though.

We don’t want to say that just any change that leads to an agent performing some different action will grant them counterfactual access to that alternate possibility, and to see this consider the following set of cases:

*Holiday 1:* Smith is planning a holiday to either Rome or Athens. He takes into account the price of each trip, the activities he can enjoy in both cities and decided upon Rome. Suppose, though, that no matter what Smiths reasoning had been, no matter what reasons he considered for or against going to either Rome or Athens, he would always decide on Rome.

*Holiday 2:* Again Smith is planning a holiday to either Rome or Athens and is weighing up the two options. He decides on Rome, but this time suppose that had he recognised more reason to go to Athens instead, he would have decided on Athens rather than Rome. That is, whilst Smith goes to Rome, there is a possible world in which he thinks there is more reason to go to Athens and subsequently decides to go to Athens.
Holiday 3: This time Smith is planning a holiday and as in Holiday 1, he will only ever choose Rome, regardless of what reasons he identifies for going to Athens instead. However, Black—our cunning neuroscientist, now bored with toying with Jones and working for the Greek tourist board—has implanted a device in Smith’s brain which, if activated will cause Jones to choose to go to Athens. As it turns out, Black’s device is not activated and Smith goes to Rome, however there is a possible world in which the device is activated and Smith chooses to go to Athens instead.

In the first case, Smith has no counterfactual access to alternate possibilities (where his holiday is concerned at any rate) in that there are no changes to his causal history that would cause him to choose to go to Athens over Rome. That being the case, let us say that he further lacks counterfactual possibilities as well as access.

In the second case, Smith chooses to go to Rome, but a change in his mental states—in this case his beliefs about where he has most reason to go on holiday—could cause him to choose to go to Athens instead. That being the case, I want to describe him as having counterfactual access to the possibility of going to Athens and this entails that he also has the counterfactual possibility of going to Athens.

In the final case, I want to say that whilst it is counterfactually possible for Smith to go to Athens, he does not have counterfactual access to that possibility. In attributing counterfactual access to an alternate possibility to an agent, remember, we are intending to give an assessment of the agent’s capacities to act given particular features of their causal history and Smith, alone, is incapable of issuing in a choice to go to Athens. Black’s intervention is entirely irrelevant to our assessment of Smith’s capabilities to choose where to go on holiday and so cannot grant him counterfactual
access to any alternate possibilities. In terms of our coffee machine example, this would be like Black opening up the back of counter and pouring hot chocolate out of its dispenser. Clearly this represents a possible dispensation from counter, but it in no way effects its capacity to dispense hot chocolate.

Black’s intervention in Holiday 3, then, is not a change of the relevant kind to grant counterfactual access to alternate possibilities, else we would end up attributing counterfactual access to every possible action to all agents at all times—we will always be able to construct some possible world scenario in which an agent is forced to perform some action by an interfering neurosurgeon like Black. The change in Smith’s mental states in Holiday 2, however, do constitute a change of the relevant kind to ground counterfactual access to alternate possibilities.

This description is obviously incomplete: Just what it is about a change in Smith’s mental states, but not Black’s intervention, that makes such an event a relevant change to counterfactual access remains a mystery, and there is room for cases of causal deviance within the framework outlined above: Smith might have believed that he had more reason to go to Athens than Rome and subsequently chosen to go to Athens but only because a chip in his brain jolted his brain into so choosing; supposing that in worlds in which he lacks such a chip a change in mental states would have no effect on his decision, I want to say that Smith does not have counterfactual access to a decision to go to Athens even though it appears that a relevant kind of change in his causal history would lead to just that decision being made.

To return to the apparent conflict between our two Frankfurtian intuitions: in the case of no alternatives needed, we can interpret the claim that agents need not be able to do otherwise as the claim that they do not need metaphysical access to doing otherwise in order to be in control; in the case of effective intervention, we can interpret
the claim that an agent requires an ability to effectively intervene in some process in order to be in control of it as the claim that agents need counterfactual access to successful intervention in order to be in control of something.

I shall return to this point on the kind of ability that being in control requires in §III – A Sketch; presently; let us move on to looking at the requirements on the secondary aspect of control: monitoring.

Awareness

We saw above, both in the initial gloss of control in §I and also in my brief discussion at the start of §II – Agential Control, that being in control of some \( x \) requires that an agent be monitoring that \( x \).

If an agent is not monitoring the object\(^{27}\) that they are trying to control, then they cannot set about either controlling it, or properly be in control of whether it behaves in this way or that: doing so would require them to know how it was in fact behaving so that it could either be actively restricted or directed or ready to intervene in the processes leading to its behaviour should they see fit.

It turns out, though, that simply monitoring something is not enough to grant control over that thing (given a relevant ability to intervene), as demonstrated by the following example:

Novice Pilot: Leslie, a novice pilot, is trying to land a passenger airline after the cockpit crew have been knocked unconscious. It comes time to open the wing flaps but, unfortunately, Leslie does not know which button on the control panel controls them.

\(^{27}\) I use object in the lose sense of ‘object of control’, so that it could be an action, event, agent or object proper.
Now, we might say loosely that Leslie is controlling the plane, and this might be true in some diminished sense of controlling (given that he does not know how to control all of the plane’s workings). More precisely I think we should say that insofar as Leslie utilises the plane’s yoke and throttle to guide the plane this way and that, he does actually guide the plane’s movements and is in control of the plane’s flight-path; however, Leslie is not controlling the wing-flaps themselves (at least not in an agential sense: if he hits the button to open them randomly or by accident he might control them in a mechanical sense) and nor is he in control of whether the wing-flaps open or not: he does not know which button controls the flaps and so he cannot possibly be aware of how to direct such an operation.

Leslie is able to effectively intervene in the plane’s operation and open the wing-flaps, though: the button is right there on the control panel for him to push, if only he knew which one. That being the case, we might say that opening the wing-flaps is within Leslie’s control; however, even if he is aware of his ability to open the wing-flaps (Presumably he knows that some button or other opens the flaps), his lack of know-how—the knowledge of how to exercise his ability to open the wing-flaps—prevents him from being in control of them.

Being in control of whether some action or event comes about, then, requires knowledge of both i) the possession of an ability to effectively intervene, an ii) the knowledge of how to express that ability to intervene: knowledge of the ability without the know-how of how to exercise it is insufficient for having agential control.

What’s more, it is not enough that one merely think, albeit correctly, that one is able to effectively intervene in some process; something more like knowledge is required for being in control. Consider the following example: Suppose that Paul gets a new smartphone with face recognition software built into it. When reading a
document on the device, if you perform a long blink, the phone is set up to recognise that gesture and perform a ‘page down’ operation (so that the text scrolls down one screens worth of text). However, the software is new, and somewhat flawed, and will pick up 30% of all normal blinks as a long blink and thus perform an unwanted page down procedure. Unaware of this functionality, Paul at first becomes frustrated that the document he is reading keeps skipping forward. After a time he notices that it seems always to coincide with his blinking, though not every time. He begins to wonder if he is actually controlling the scrolling of his document with his blinks (albeit intermittently). Because of the 30% recognition rate, Paul ends up unsure as to whether it is mere coincidence that the document’s scrolling always coincides with his blinking. After contacting the smartphone manufacturer and learning about the feature Paul returns to you to triumphantly confirm that he was controlling the scrolling.

Now, I think it is clearly true that Paul was controlling the scrolling of his document with his blinks; but, I think he was doing so in a mechanical or causal sense (the phone’s camera and software being a mechanical control for the device which he was unknowingly utilising), and indeed I think this is what he means when he tells you that he really was controlling the device with his blinking. He was not, however, directing the scrolling, or using his blinks to command the device to perform that function with authority. How could he? He did not know whether the device really was responding to his blinks.

So, whilst utilising mechanical control does not require that one know that one can effectively intervene in a particular process, we may well have mechanical control over a great many things without knowing it. Controlling, however, in the sense of intentional guidance requires that one not only be able to effectively intervene in a process, but also have the awareness and know-how to effectively exercise that ability.
Contextualism in Control

I said above that an agent’s controlling some action involves their having the capacity to do otherwise and hence relevant counterfactual access to doing otherwise, whereas being in control of whether they perform some action involves having the opportunity to do otherwise, but where this need not be absolute-opportunity as I defined it in §III, Metaphysical and Counterfactual Access above.

To expand on this point, Manley and Wasserman (2007 & 2008) have recently suggested that contextualism might help to assuage some of the long standing worries about dispositional analyses and, more specifically, help us in linking conditional and dispositional statements. I think that much the same is true of abilities and, by extension, agential control (and in fact I suggest that we build a contextualist account of abilities onto a Manley-Wasserman style account of dispositions).

In some contexts, asking whether someone is in control of whether $P$ or not-$P$ obtains might invoke a notion of ability involving metaphysical access, and hence an absolute-opportunity, to doing otherwise, whilst in other contexts that same question might invoke a much weaker notion of ability involving only counterfactual access. For instance, a hard determinist might say: “for any $A$-ing, no one is in control of whether they $A$”, and by this they mean that, given the truth of causal determination (which they have assumed) everyone lacks the absolute-opportunity to do otherwise; that is, the possibility of genuinely having access to more than one outcome from just one causal history.

However, having the opportunity (rather than absolute-opportunity) to $A$ can mean simply that one’s circumstances are conducive to the expression of one’s capacity to $A$; that is, if one were to try to express one’s capacity to $A$, one would be successful in $A$-ing, and this, of course, would involve counterfactual rather than
metaphysical access. So, for instance, a compatibilist might say “John is in control of whether he *As*”, and mean by this that nothing about John’s actual circumstances would prevent him from *A*-ing or not-*A*-ing were he to try. Given that we are talking about what John is able to do in his circumstances, however, this would have to be read as counterfactual access to those worlds that most closely resemble the actual world.28

The difference between controlling some action and being in control of whether that action is performed, then, is not a difference in the kind of ability involved (in terms of metaphysical or counterfactual access); rather, it is a difference in the worlds at which we assess an agent’s ability to effectively intervene and bring about some other action.

In *Black and Jones*, for instance, the relevant worlds at which to assess whether Jones controlled his shooting the mayor (in the active sense of *controlling*) are plausibly those in which he deliberates uninhibited about what he will do and then acts upon the outcome of this deliberation: that is, we will not limit our assessment of Jones’s ability to do otherwise than shoot the mayor to those worlds in which Black is present, because Black plays no role in the actual directing of Jones’s deliberating about and subsequently shooting the mayor. Here, Jones’s controlling the shooting of the mayor is grounded in a capacity of his: to deliberate and issue in different kinds of action; he is at least capable of choosing not to shoot the mayor.29

---

28 I move to talk of worlds from that of ‘histories’ in the exposition of counterfactual and metaphysical access for ease of expression. Not much is lost in this, and the reader is free to translate back into talk of histories if they so wish.
29 For reasons why it would be a problem if Jones’s deliberation could not issue in any action other than ‘shooting the mayor’ see Fischer and Ravizza’s discussion of guidance control (1998: Ch.2, esp. §. III & IV; & Ch.3). Their main interest is in in the capabilities of *mechanisms* to respond appropriately to reasons, specifically that the actual sequence mechanism (the mechanism that issued in an agents behaviour in the actual sequence of events) is capable of responding differently to reasons. Without such responsiveness, or responsiveness at the right level, agents would not appear to be genuinely guiding (and hence in control) of their actions (1998: Ch.3).
Conversely, the relevant worlds at which to assess if Jones was in control of whether the mayor was shot appear to be confined to those in which Black is present: here we are assessing Jones’s opportunity to do otherwise than shoot the mayor, and clearly his actual circumstances are not conducive to his doing so.

Consider a second example, though, in which an agent both controls their actions and has the opportunity (in terms of counterfactual access) to do otherwise:

**White and Smith:** White tells Smith he approves of his plan to shoot the mayor, but there is no device and no way for White to force Smith to shoot the mayor. After deliberating about what to do, Smith shoots the mayor.

In this case, Smith clearly controls (in the active sense of controlling) his shooting the mayor, but he also has the relevant counterfactual access required for being in control of whether he shoots the mayor or not, because if he were to try not to shoot the mayor he would not. He has the opportunity to not shoot the mayor, and this is the case even if Smith’s world is causally deterministic: his circumstances do not involve any factors that would prevent him from not shooting the mayor were he to try (or not try to shoot him as the case may be).

Just what determines the relevance of certain possible worlds to an ability or control assessment is a topic for a different dissertation, though we might make a start by adopting a Fischer-Ravizza-esque ‘alternate sequence’ approach (for controlling, at least). According to such an approach we need to look at the actual sequence of events and the mechanisms that brought about the relevant action or event and then make our ability assessment only at those worlds in which that same kind of mechanism operates. This is effectively what we did with Jones’s shooting the mayor.
above: we assessed his capability to not shoot the mayor by assessing his deliberative mechanism’s capacity to issue in his choosing not to shoot the mayor.

For being in control of whether some action or event comes about an actual sequence approach will not work, though we might instead try an actual circumstance approach, where we look only at those worlds in which the agent’s actual circumstances obtain. This is not a complete answer, though, as we have already seen that in different contexts ‘being in control of whether’ can point to different kinds of ability.

A Sketch

With the above sections in mind, we are ready to provide our sketch for our theory of agential control and also to analyse some control terminology that we might frequently employ to describe an agent’s relationship to various actions or events.

First, though, I have noted above that agents can have agential control over their own behaviour, events unfurling around them and even other agents. It is worth briefly thinking about whether there is a proper object of control: that is, a common kind of thing over which an agent has control in each of these three cases.

Intuitively, and from the above discussions, being in control of one’s own actions involves determining in some way (by restricting or directing; or being able to effectively restrict and direct) what one does; being in control of some event occurring involves determining that events coming about; and being in control of some other agent involves determining what actions that other agent performs.

The common object to all of these kinds of control, I think, is most naturally conceived of as a process: in being in control over one’s actions, one is in control over the process leading to those actions; in being control of an event’s occurring, one is in
control of the process leading to that event; and in being in control of some other agent, one is in control of the processes leading to that agent’s actions.

With that in mind we can state our conditions for controlling and being in control of whether in terms of processes:

An agent \( S \) controls some process \( P \) iff

i) \( S \) plays an active role in restricting or directing \( P \)

ii) \( S \) is able to effectively intervene in \( P \)

iii) \( S \) is aware of their ability to intervene in \( P \) and how to exercise that ability

An agent \( S \) is in control of whether some process \( P \) issues in some event \( E \) iff

i) \( S \) is monitoring \( P \) accurately

ii) \( S \) is able to effectively intervene in the \( P \)

iii) \( S \) is aware of their ability to intervene in \( P \) and of how to exercise that ability.

Note that there are several ways that an agent could intervene in a process leading to some event; they could either restrict or direct that process so that it issues in some other event, or they could restrict and direct the process so that it is effectively aborted and does not issue in any event at all.

We can also note here that the way that processes are individuated will be contextually determined by the way that we frame our questions about control: suppose that a missile is headed for London in the UK and you have access to its navigation controls; using the controls you are able to affect the missile’s flight-path, but only so far as to direct the missile to hit Manchester rather than London.\(^{30}\) It seems

\(^{30}\) This example roughly mirrors Fischer and Ravizza’s ‘missile 3’ case (1998, p. 95).
in this case that you are in control of whether the missile hits London and of whether the missile hits Manchester (and therefore, plausibly, in control of whether the missile hits London or Manchester); however, you are not in control of whether the missile hits the UK. I think this can be explained by the individuation of the mechanisms that we are assessing your ability to intervene in: on the one hand you are able to intervene in the process guiding the missile to London (or Manchester), and on the other you are not able to intervene in the process guiding the missile to the UK.

It might be suggested that there also needs to be sort of constraint on what an agent is willing to do in bringing about certain events. That is, whilst an agent might be able to intervene in some process and bring about a certain event, be aware of that ability and have the know-how to express it, it might be contended that they can still fail to be fully in control of some events occurring.

For instance, suppose that you pass a homeless man on the street begging for change. Are you in control of whether they sleep on the street tonight? According to my analysis, we might think that you are: so long as you stay aware of the man’s condition and know that he is out there, you are able to get him off the street and may be fully aware of this ability: You could offer to let him stay in your home, or even just give him your home and take his place on the street. Still, I think one would be inclined to say that they are not in control of whether a homeless person lives the life that they do on the street: after all, you had nothing to do with his ending up on the streets, perhaps thousands of people could also help him out (and none do), you aren’t doing anything wrong by walking past and leaving him; the man is not your responsibility, we might say.
I think, in contrast to the above, that you are in control of whether that man sleeps on the street tonight, and that this position is most likely to arise out of following Fischer and Ravizza in treating control and responsibility as interchangeable.

It is a matter of contention (or at least should be) whether we are individually or collectively responsible for the homeless in our society, but these considerations ought to not to inform our intuitions about whether we are in control of the homeless remaining so: moral responsibility might require control, but that is not to say that control is sufficient for moral responsibility.

Moral considerations aside, it perhaps remains uncomfortable to think that we are in control of the suffering of another human being, especially when society at large has adopted a stance of generally ignoring the plight of people living on the streets; none-the-less, I maintain that if you can effectively intervene in some ongoing situation or process and are aware of that ability and how to exercise it, that process and the events it issues are under your control.

This is, by admission, only a sketch of a theory of agential control, and it would not be possible to complete it in this introduction, for a start we need a full account of what processes are and how to individuate them; it is hopefully sufficiently clear, though, for us to see how such an account might work, and begin thinking about how we could begin to build accounts of freedom and responsibility upon it.

IV – THE COLLECTION
The main body of this submission is divided into three papers that collectively constitute the groundwork for the ability element of the account of control sketched in the previous sections. Each paper, whilst not specifically focussed on control, does some work to build up the account of ‘ability to effectively intervene’ or support the requirement of such an ability for being in control: Briefly, I take agential control to
involve, in part, an agent’s being able to effectively intervene in a process leading to action; I take this ability to involve context sensitivity and best understood as a kind of disposition to intervene if one were to try. Accordingly, the first paper in the collection attacks the notion of control as not involving an ability to do otherwise, and motivates the need for an ability to intervene; the second paper is focussed upon modifying Manley and Wasserman’s (2008) contextualist account of dispositions; and the final paper defends counterfactual access style compatibilist approaches and suggests the contextualist account of abilities that I suggest we build upon Manley and Wasserman’s dispositional analysis.

Below I give a brief overview of each paper along with some points on the material not explicitly raised in the papers themselves.

Guidance Control and the Anti-Akrasia Chip

In the first paper, ‘The Anti-Akrasia Chip’, I tackle the well-known guidance control account defended by Fischer and Ravizza in their (1998) Responsibility and Control. My main target in this paper is the ownership constraint that Fischer and Ravizza place upon having guidance control. The main point of this paper is to turn us away from accounts of control that are heavily directed by our intuitions about moral responsibility and to motivate the idea that control requires some kind of ability to do otherwise.

Briefly, Fischer and Ravizza distinguish between two kinds of control: regulative control, which involves having the dual power to perform some action A or some other action B instead, and guidance control, which involves actually guiding some action A through its performance. They identify guidance control as “the

31 It might at first appear that having guidance control equates to what I have termed controlling some x, whilst having regulative control equates to being in control of whether some x comes about; however,
freedom-relevant condition necessary and sufficient for moral responsibility” (1998, p. 241, fn.2) and hence, since they are interested in establishing the compatibility of responsibility and determinism, concentrate their account on this kind of control.

An agent has guidance control, according to the Fischer Ravizza account, whenever their behaviour is issued from one of their own moderately reasons responsive mechanisms (where a mechanism is, roughly, a process leading to action); a mechanism is moderately reasons-responsive to the extent that, holding fixed the operation of that mechanism, an agent would recognise reasons in an understandable pattern\(^{32}\) and to react to at least one sufficient reason to do otherwise across alternate scenarios in which there exist reasons for doing otherwise. An agent can *make* such a mechanism their own, on the Fischer-Ravizza account, by *taking responsibility* for it, which involves an agent forming an attitude towards themselves so that they see themselves as the agent of the behaviour issued by that mechanism and an apt target for the reactive attitudes it might generate.

I present a case in which an agent meets the conditions for guidance control, and yet intuitively is not actually guiding their behaviour; on the basis of this case I argue that either taking responsibility for a suitably responsive mechanism is not sufficient for having guidance control over some action, or guidance control does not equate to actually guiding (having agential control over some action).

\(^{32}\) An agent’s reason recognition must demonstrate that they understand how reasons fit together (Fischer & Ravizza 1998, ch3, §IV, esp. pp. 70-1); so, for instance, if Jennifer would buy a ticket to watch her favourite basketball team play if the ticket cost any amount *except* $1000 (which she judges to be too much money), her reasons-recognition would reveal a confusing pattern: after all, if $1000 is *too much*, why would she buy it for $1001, or $1002? Etc. (1998, pp. 66-7)
A large part of the Fischer-Ravizza analysis is driven by their focus upon moral responsibility and whilst I concede in the paper that taking responsibility for some action or the mechanism that brought it about very plausibly is a requirement on being morally responsible, I argue that there is no principled reason for thinking that taking responsibility for some action will grant one any control over that action. In fact, taking responsibility, I suggest, is something an agent ought to do once they realise that they are, or were, in control of something.

There is something right about there being an ownership element to control: when one has control over an action, that action is yours in a sense tied very closely to agency; it is precisely this kind of ownership that we saw was lacking in cases of causal deviance: an agent might act, but not truly own the action, having acted beside themselves. However, I think that, rather than being determined by some attitude of the agent, the ownership element of control is granted by an agent’s monitoring some process alongside their having the ability to intervene and corresponding awareness that they could do so. An agent could, according to my account, know that a child is drowning, know also that they could intervene and prevent this from happening (with the corresponding-know how), refuse to take responsibility for whether the child drowns or not (or see it as their doing, even in part), and yet still be in control of whether the child drowns.

Interestingly, whilst Fischer and Ravizza see their distinction between regulative and guidance control as one that represents a division between having abilities to do otherwise and lacking them, I think that guidance control, specifically a mechanism’s being moderately reasons-responsive does require an agent to be able to

---

33 this is evident from the Strawsonian notion of taking responsibility requiring that one see oneself as an apt target for the reactive attitudes that the behaviour issued by a mechanism might generate.
do otherwise in the sense of having counterfactual access to alternate possibilities. I
do not go into this point in the paper, and so I shall expand upon it briefly here:

Fischer and Ravizza make the following claim about mechanisms.

…in a Frankfurt-type case the agent could not have done otherwise, and thus the agent is not reasons-responsive. But it is crucial to see that in these cases the kind of mechanism that actually operates is reasons-responsive, even though the kind of mechanism that would operate – that is, that does operate in the alternative scenario – is not reasons-responsive. (1998, p. 38)

This passage nicely captures the Fischer-Ravizza treatment of Frankfurt cases: In Black and Jones the actual sequence of events involves Jones’s acting, of his own accord, in accordance with his deliberation and subsequent decision to shoot the mayor. The actual sequence mechanism, therefore, is Jones’s acting on his deliberative faculties (without the interference of Black). The actual sequence mechanism is moderately reasons-responsive, according to Fischer and Ravizza, because Jones’s normal deliberative faculties are able to recognise and respond to reasons; however, if Jones were to try to do otherwise, he would be prevented from doing so by Black, so Jones is not able to respond to reasons and is not reasons-responsive. Conversely, the alternate sequence mechanism, that would operate were Jones to show some inclination for trying to do otherwise, would be Black’s using his device to force Jones to shoot the mayor; this alternate sequence mechanism is not reasons-responsive, as no matter what reasons were present ‘manipulation by Black to force Jones to shoot the mayor’ will always result in Jones’s shooting the mayor.
Now, there are some interesting points to raise here about mechanism individuation, but I shall concentrate instead on treatment of Jones as non reasons-responsive. I think that Fischer and Ravizza have confused capability with opportunity: Jones clearly lacks the opportunity to do otherwise, or, to put it in terms of reasons-responsiveness, Jones lacks the opportunity to respond to reasons in his actual circumstances; but why should that make us think that Jones lacks the capacity to respond to reasons?

Agents’ capabilities are typically grounded in the capabilities of the various mechanisms that make them up: I have the capacity to walk because some part of me (my body) is capable of walking; I have the capacity to deliberate because some part of my mental faculties are capable of deliberating; and I submit that it is in virtue of Jones’s deliberative faculty’s capacity to respond to reasons that he too is capable of responding to reasons; how else, after all, could an agent be capable of doing otherwise than by some mechanism of theirs being able to issue in different actions?

If the above assessment is correct then it turns out that, on a counterfactual access reading, moderate reasons-responsiveness actually requires an ability to do otherwise. I think our own theory of control ought to bring this counterfactual ability centre stage.

---

34 Why for instance should the alternate sequence mechanism be individuated as ‘Black’s manipulating Jones to force him to shoot the mayor’ rather than simple ‘Black’s manipulating Jones’, the latter of which could be responsive to reasons; I discuss this point at the end of ‘The Anti-Akrasia Chip’, see pp. 108.
Dispositional Contextualism

In the second paper in the collection, ‘Getting Specific with Manley and Wasserman’, I critique and propose a modification to Manley and Wasserman’s (2008) recently proposed analysis of dispositions.

As indicated in my sketch of a theory of control above, I think that Fischer and Ravizza are wrong to discard abilities to do otherwise from their account of guidance control; whilst having *metaphysical access* to alternate possibilities may turn out to be unnecessary for actually guiding one’s behaviour, I submit that having *counterfactual access* to alternate possibilities is a necessary constraint on having agential control.

Now, as counterfactual access is tied to conditional type statements such as ‘S would A if S tried to A (in certain relevant circumstances)’, our account of control needs to be built upon an account of abilities that appropriately reflects this connection. Specifically, I think that the ability to effectively intervene in some process should be understood as the disposition to intervene in that process if one were to try.

Conditional accounts of dispositions have traditionally been faced with various damning counterexamples, making the prospect of building a convincing account of abilities upon such an account seem rather bleak. However, recently Manley and Wasserman have suggested that contextualism might help us to solve these longstanding worries:

Briefly, simple conditional accounts of dispositions of the form ‘N is disposed to M when C iff N would M if C’ are susceptible to cases of finking, masking and mimicking, all of which seem to demonstrate that such accounts are unworkable. One strategy for resolving these problems is to ‘get specific’; that is, to make the antecedent conditions for the manifestation of a disposition more specific so as to rule out the
presence of dispositional finks, masks or mimics that might otherwise interfere with out assessment of an object’s dispositions; I call this the restrictive approach to analysing dispositions, in that it involves the restricting of the set of cases at which we assess a dispositional ascription to a narrow subset as determined by the specific manifestation conditions.

Manley and Wasserman have convincingly argued that this kind of restrictive approach will not ultimately get conditional accounts out of trouble: dispositions can be masked even in the most specific manifestation conditions. They provide the following kind of example: A fragile glass might be disposed to break when dropped, and this disposition might even correspond to the highly specific disposition to ‘break when dropped on Earth from one metre up onto a solid surface with a Shore durometer measurement of 90A, through a substance with a density of 1.2 kg/m$^3$’; however, this glass has a strong spot, or reverse Achilles’ heel, such that if it lands in just the right way, at just the right angle it will not break, no matter how great the impact force of its landing. One could drop this glass in the specific manifestation conditions identified and have it land on its strong spot, and therefore fail to break; still the glass is fragile, and still it is disposed to break when dropped (it would most of the time).

On the basis of this counterexample, Manley and Wasserman’s argue that we ought to abandon the restrictive type approaches typified by the getting specific strategy and adopt instead what I call a proportional approach to analysing dispositions. To this end they present the following analysis of dispositions:

\[(PROP) \quad \text{N is disposed to M when C iff N would M in some suitable proportion of C-cases}\]
A C-case can be read roughly here as a case in which circumstances of kind C obtain (so if a glass is disposed to break when dropped, then it will break in a suitable proportion of all dropping cases). Just what counts as a ‘suitable proportion’ of C-cases is determined by the context of the dispositional ascription; so for instance, on a building site where heavy blows are common the standards for fragility might be quite low, such that an object only has to break in a small proportion of cases when struck in order to count as fragile; conversely, in a home where hard blows are infrequent the standards for fragility might be much higher such that an object has to break when struck in a much greater proportion of cases in order to count as fragile. The same object, then, might truthfully be fragile on a building site and non-fragile in the home.

(PROP) deals easily with the finking, masking, mimicking and Achilles’ heel type counterexamples mentioned above: the cases in which an object fails to manifest a disposition in that dispositions characteristic stimulus conditions represent a small proportion of the total set of C-cases; so long as the object does manifest the relevant disposition in a suitable proportion of that total set, the finking, masking, mimicking and Achilles’ heel cases will have no real effect. What’s more, (PROP) makes good on our intuitions about the contextual sensitivity of dispositional predicates across a range of contexts (what is fragile for a giant might obviously be non-fragile for a human) and accounts for the gradability of dispositional predicates: we often make statements of the form ‘x is more fragile than y’, which according to (PROP) would mean that x breaks in more fragility relevant C-cases than y.

My concern with (PROP) is that by taking a wholly proportional approach to analysing dispositions, and allowing for no restriction over the C-cases at which we assess dispositional ascriptions at all, we risk allowing what I call dispositional swamping cases:
Suppose that we attribute to some object N a disposition to M when C in a
certain context K; suppose further that in all K-relevant cases N will M when C, but
in all non-K-relevant cases N will not M when C. Intuitively, I think we ought to say
that this object is disposed to M when C in K, given that it Ms when C in all K-relevant
cases; however, in a swamping case the non-K-relevant cases far outweigh the K-
relevant cases, thereby skewing the proportion of cases in which N Ms when C to the
extent that it does not come out as being so disposed according to (PROP); and this is
possible no matter how low we set the standard for fragility in a given context.

I therefore suggest that (PROP) is best interpreted as a dual approach analysis
of dispositions: one that is both proportional and restrictive. This is achieved by
allowing context to determine both the proportion of C-cases that must result in
manifestation in order that an object have a particular disposition and the set of C-
cases at which the disposition will be assessed. In this way, (PROP) avoids all of the
traditional problems for conditional accounts as well as my swamping
counterexamples.

It is this dual reading of (PROP) that I think could form the basis of an account
of the ability to intervene required by agential control.

Abilities in Context

In the final paper, ‘Relevant Abilities’, I defend dispositional compatibilism from an
objection by Whittle (2010). From this defence I suggest the contextualist theory of
abilities, R-ABLE, that I think can ground the ability to effectively intervene required
for agential control.

Briefly, dispositional compatibilism is the view that we can provide a
dispositional analysis of abilities that will show an ability to do otherwise to be
compatible with determinism and hence establish the compatibility of freedom, or
responsibility, and determinism. This was the position taken by early compatibilists
such as Moore (1912) and Ayer (1954) in defence of the principle of alternate
possibilities (PAP): ‘an agent is free in A-ing only if they could have done otherwise
than A.’

Unfortunately, this original dispositional compatibilism was unworkable:
abilities were treated to a simple conditional analysis such as ‘an S is able to A iff S
would A if he tried to A’, and whilst this appeared to show that abilities were
compatible with determinism\textsuperscript{35} was open to a devastating counter example: a patient
in a coma might walk if they were to try, however being in a coma they are unable to
try and so, intuitively, unable to walk and the simple conditional analysis of abilities
is incorrect. Call this an impossible stimulus condition case.

Aside from this problem, the introduction of Frankfurt cases seemed to show
that abilities, so analysed, were irrelevant to moral and free action: in Black and Jones,
Jones would not do otherwise if he tried because Black is waiting in the wings to force
him to shoot the mayor should he do so. Yet, Jones is free and responsible for shooting
the mayor; that being the case the abilities picked out by the simple conditional
analysis are not necessary for free and moral action, PAP is false, and we should set
about demonstrating the compatibility of freedom and determinism by some other
means.

Recently, compatibilists such as Fara, Vihvelin and Smith, the so called ‘new
dispositionalists’, have revived this position with updated accounts of dispositions.

\textsuperscript{35} It might be causally determined that S will try to A and subsequently A, however it might also be the
case that were S to try to not-A then S would not-A; given the truth of that counterfactual conditional
S is able to do otherwise according to the simple conditional account of abilities (even if it is causally
determined that she will not).
The problem with the old dispositional compatibilism of Moore and Ayer, according to the new dispositionalists, is that it was based on far too simple an account of dispositions; once we update our dispositional analysis to a more sophisticated conditional account, according to which dispositions can be masked whilst not being removed entirely, we will see that agent’s really do retain abilities to do otherwise in Frankfurt cases, and therefore that such cases fail to show that PAP is false.

More recently, Whittle has argued that the new dispositionalists’ defence of PAP is unsuccessful: according to Whittle, the abilities picked out by the new dispositionalists’ updated dispositional accounts of abilities are what she calls global, being general capacities to act, rather than local, opportunities to act. She then goes on to argue, via a range of Frankfurt-type cases, that global abilities do not appear to be relevant to free and moral action: If any kind of ability were relevant it would be local, and these are clearly ruled out by Frankfurt cases.

In this final paper I provide two responses to Whittle’s attack on dispositional compatibilism; I first demonstrating that, because PAP is concerned only with agents’ responsibility for their actions and Whittle’s argument rests upon comparing agents’ responsibility for the outcomes or consequences of their actions across a range of Frankfurt cases, her argument misfires against the dispositional compatibilist defence of PAP.

I then go on to consider her argument against a modified version of PAP, according to which an agent is responsible for some outcome only if they could have brought about some other outcome. Initially Whittle’s argument seems to hit the mark against this version of PAP, however I suggest that dispositional compatibilists need not accept that their accounts pick out global rather than local abilities. By adopting a contextualist account of dispositions, such as the dual approach reading of Manley and
Wasserman’s (PROP), we can generate a dispositional account of abilities that picks out relevant local abilities which are assessed at contextually relevant cases. I go on to demonstrate that these relevant abilities can be used to ground a defence of PAP.

It is this contextualist ability account that I suggest we incorporate into our account of agential control to underpin an agent’s ability to effectively intervene.
THE ANTI-AKRASIA CHIP

Abstract: According to Fischer and Ravizza, an agent has guidance control over their actions, which they gloss as the agent’s actually guiding that action, whenever those actions issue from one of their own moderately reasons-responsive mechanisms. This involves two elements: i) the process leading to their actions’ being suitably responsive to reasons (moderately reasons-responsive); and ii) their taking a certain attitude towards processes of that kind such that they see themselves as the agents of the behaviour those processes issue (what they call ‘taking responsibility’ for a mechanism). I present, and defend, an example in which an agent intentionally acts via a suitably reasons-responsive process which they have taken responsibility for and yet, intuitively, does not actually guide their action. On this basis I argue that either i) taking responsibility for a moderately reasons-responsive mechanism is not sufficient for having guidance control, or ii) guidance control does not equate to actually guiding some action.

Plausibly, a link exists between control and moral responsibility: when an agent is in control of what they are doing, we generally hold them morally responsible for doing it; likewise, when an agent is morally responsible for some action of theirs, we tend to think of them as having had control over that action’s performance. What does it take, then, for an agent to be in control of their behaviour?

Before we answer that question we ought to answer some subsidiary ones: First, what kinds of control can agents have over their behaviour; and second, which type of control is the relevant one for being morally responsible?

According to Fischer and Ravizza, we can distinguish between two different kinds of control that agents can have over their actions: guidance control and
regulative control. Guidance control involves freely performing some action $A$, whereas regulative control involves having the dual power to perform one action $A$ or some other action $B$ instead. (1998, p. 31). They provide the following example to illustrate this distinction:

*Driving Instructor:* Sally is driving with her driving instructor in a dual control car. She comes to a right turn that she needs to make and, at the appropriate time, slows the car, signals, turns the steering wheel to the right, and carefully guides the car around the corner. We can imagine that her driving instructor is quite happy to allow Sally to guide the car to the right on her own, but that if she had shown any inclination to cause the car to go in some other direction he would have intervened and caused the car to go to the right himself (just as it actually does).

They go on: “insofar as Sally actually guides the car in a certain way… she has “guidance control.” [over the car]” (1998, p. 31); given that she cannot make the car do anything other than it actually does (due to the presence of her driving instructor), Sally lacks regulative control over the car (1998, p. 31).

Taking these initial glosses at face value, having guidance control over an action amounts to having actually guided that action through its performance and having done so ‘under one’s own steam,’ as it were; whereas having regulative control

---

36 The use of ‘freely’ in Fischer and Ravizza’s gloss of guidance control should, I think, be interpreted as meaning ‘without coercion or compulsion,’ rather than ‘whilst being able to do otherwise’ (which would amount to having regulative control).

37 This example is taken from (Fischer & Ravizza, 1998, pp. 30-2). They provide two examples, one in which Sally is free to guide the car in any direction she wants and then another in which her driving instructor acts as a counterfactual intervener, ensuring that the car goes just as it actually does. I condense these two examples into one here for the sake of simplicity.
over an action amounts to having the power to determine whether or not that action is performed.

For Fischer and Ravizza, who hold that agents need not be able to do otherwise in order to be morally responsible for their actions, it is guidance control that is the relevant kind for moral responsibility, and they base this intuition on the following familiar kind of case from Harry Frankfurt: 38

*Jones and Black:* Jones intends to kill the mayor and Black, an evil scientist, approves of his doing so. To ensure that Jones carries out his plan, Black implants a device in Jones’s brain which, if he should show even the slightest inclination of not following through on his intention, will force Jones to act on his original intention to kill the mayor. As it happens, Jones kills the mayor of his own accord (without the intervention of Black or the device). 39

Intuitively, Jones is morally responsible for his killing the mayor. It was Jones, and not Black or the device, who killed the mayor; however, he was unable to bring about any other course of action. The parallel to Sally in *Driving Instructor* is clear: Jones lacks regulative control over his actions (just as Sally lacked regulative control over the car) in that he could not do otherwise than he in fact did; however, insofar as he, and no one else, actually guided his act of killing the mayor, Jones has guidance control over his actions. That being the case, it is guidance control (claim Fischer and Ravizza) that is “the freedom-relevant condition necessary and sufficient for moral responsibility” (1998, p. 241, fn.2).

38 See (Frankfurt, 1969)
Under what circumstances, then, does an agent have guidance control? According to Fischer and Ravizza, an agent has guidance control over an action whenever that action issues from one of their own moderately reasons-responsive mechanisms (1998, p. 241). This analysis is quite technical and I dedicate §1 to explaining the concepts that make it up and how they fit together. For the time being: a mechanism can be understood simply as “the process that leads to [an] action,” or the “way [an] action comes about” (1998, p. 38); and is moderately reasons-responsive to the extent that it can issue in different actions given the presence of, and in accordance with, different reasons for acting; what’s more it must respond to the presence of reasons in a regular and intelligible pattern.

As for a mechanism’s being an agent’s own: one can make a mechanism one’s own by taking responsibility for it, which involves forming an attitude towards the behaviour issuing from that mechanism such that one see oneself as the source of that behaviour and as an apt target for the reactive attitudes that it might generate (1998, p. 241). My concern in this paper will largely arise from this second element of guidance control.

Fischer and Ravizza’s ownership condition has already received some discussion: (Mele, 2000), for instance, has argued that taking responsibility is not necessary for moral responsibility, whilst (Stump E., 2002) and (Long, 2004) have challenged the sufficiency of the guidance control conditions, more generally, by

---

40 It is important to note, however, that the mechanism cannot be the entire process, since the entire process would include both the mechanism itself and the inputs to that mechanism. We need to differentiate these two things so that we can make sense of the same mechanism responding to different inputs in alternate sequences. Thanks to an anonymous referee for this point.

41 This amounts to the mechanism’s being weakly reasons-reactive and regularly reasons receptive; I explain and analyse these concepts in greater detail in §1 below.

42 See (Fischer & Ravizza, 2000) for their reply and (Mele, 2006) for a response.
proposing cases in which agents take responsibility for a suitably reasons-responsive mechanisms and yet are not in control of, or morally responsible for, their behaviour.

My point in this paper follows a similar line to Stump and Long, with some important differences in the example motivating my objection, although I concentrate more closely on the ownership element of the guidance control analysis. I contend that, if we accept the initial gloss of guidance control as ‘actually guiding some action,’ taking responsibility for a moderately reasons-responsive mechanism cannot be sufficient for having guidance control—specifically, an attitude on the part of an agent cannot make the relevant difference between their actually guiding their behaviour and its being out of their control—and I present a counterexample in §2 to demonstrate this claim.

If taking responsibility for a moderately reasons-responsive mechanism really is sufficient for having guidance control then I argue that guidance control cannot amount to actually guiding some action and I’ll further suggest that, if this is the case, it is the notion of control corresponding to actually guiding some action that we ought to be interested in.

In §3 I’ll defend my counterexample against some potential objections, then finally in §4 I’ll conclude by briefly looking at what my example suggests about how agents come to have ownership of their behaviour along with some wider implications for the link between control and moral responsibility.

1. GUIDANCE CONTROL
This section is dedicated to presenting Fischer and Ravizza’s somewhat technical analysis of guidance control. Those familiar with their account can skip ahead to §2 (though some of the points highlighted here will be important in the defence of my example later on).
As already noted above, an agent has guidance control over an action whenever that action issues from one of their own moderately reasons-responsive mechanisms. Guidance control therefore consists in two elements:

i) The reasons-responsiveness of the mechanism that issues in action; and

ii) The agent’s ownership of that mechanism.

I shall take these two elements in turn.

Moderate Reasons-responsiveness

According to Fischer and Ravizza’s analysis, a mechanism of kind \( K \) is moderately reasons-responsive (henceforth MRR) to the extent that:

…holding fixed the operation of a \( K \)-type mechanism, the agent would recognize reasons (some of which are moral) in such a way as to give rise to an understandable pattern (from the viewpoint of a third party who understands the agent’s values and beliefs) and would react to at least one sufficient reason to do otherwise (in some possible scenario).

(1998, pp. 243-4)

A mechanism is MRR, then, to the extent that it is what Fischer and Ravizza term ‘regularly reasons-receptive’ and ‘weakly reasons-reactive’ (henceforth RRR and WRR).

A mechanism is WRR if it displays the ‘executive power’ to respond to a sufficient reason to do otherwise (1998, p. 75). That is, it should be possible for the

---

43 My presentation of these two notions (WRR and RRR) follows the order of Fischer and Ravizza’s own exegesis of these concepts, rather than the order in which they appear in the above quote from their conclusion.
mechanism to issue in a different action given the presence of a sufficient reason to do so:

**WRR**: An actual sequence mechanism \( K \) is WRR if (holding fixed the operation of a \( K \)-type mechanism) there exists *some* possible world with the same physical laws as the actual world in which there is a sufficient reason to do otherwise, the agent recognises this reason, and the agent does otherwise for that reason.\(^{44}\)

WRR alone, however, is not sufficiently responsive to allow for guidance control: the agent also needs to display a suitable level of reasons-receptivity—a level that demonstrates that they have an understanding of how reasons fit together (Fischer and Ravizza 1998, p. 71).

A mechanism is RRR, then, to the extent that it displays the ‘cognitive power’ to recognise reasons (where some of those reasons are moral), and for that reasons-recognition to form a regular and intelligible pattern.\(^{45}\) Fischer and Ravizza cash this notion out in terms of an “imaginary interview” with the agent:

On our approach, it is as if a “third party” (the one assessing the moral responsibility of the relevant agent) conducts an “imaginary interview” with the agent. In this interview he asks about various actual and

\[^{44}\text{This formulation of weak reasons-reactivity is roughly the same as Fischer and Ravizza's definition for weak reasons-responsiveness given on pp. 63–4. I take it that there is no issue with this: WRR entails at least weak reasons-reception (such that an agent recognises a reason in at least one possible scenario—otherwise they would be unable to react to any reasons at all) and this is precisely what weak reasons-responsiveness appears to involve (See fn.10 below).}\]

\[^{45}\text{A mechanism that is WRR but displays weaker receptivity to reasons (such that the agent’s reasons-recognition would not give rise to an understandable pattern) would be only weakly reasons-responsive, which turns out, according to Fischer and Ravizza, to be insufficient for guidance control. See (Fischer \& Ravizza, 1998, pp. 65-8).}\]
hypothetical scenarios, and elicits views from the agent as to what would constitute sufficient reasons. (1998, p. 71)

I take it that this “imaginary interview” is intended to be a representation of what reasons the agent actually would recognise in a series of possible scenarios. We can therefore drop the ‘imaginary third party’ and define RRR in terms of an agent’s pattern of reason-recognition across a range of possible worlds, taking into account the reasons present and the agent’s reasons-recognition at each possible world and looking for a discernable pattern across the whole range of possible worlds considered. So, for reasons of clarity (and continuity with the definition for WRR), let’s formalise RRR as follows:

**RRR:** An actual sequence mechanism $K$ is RRR if (holding fixed the operation of a $K$-type mechanism) for a set of possible worlds $W$ – with the same physical laws as the actual world – in which there exists sufficient reason to do otherwise, the agent’s reason-recognition across the members of $W$ gives rise to an understandable pattern (where some of the reasons recognised are moral).

A mechanism is MRR when it displays both of these traits in combination. So, from the two definitions above we arrive at:

**MRR:** An actual sequence mechanism $K$ is MRR if (holding fixed the operation of a $K$-type mechanism) for a set of possible worlds $W$ with the same physical laws as the actual world and in which there exists sufficient reason to do otherwise, (i) the agent’s reason-recognition across the members of $W$ gives rise to an understandable pattern (where
some of the reasons recognised are moral); and (ii) in at least one of the possible worlds in $W$ in which the agent recognises a sufficient reason to do otherwise, the agent does otherwise for that reason.\(^{46}\)

There are two things to highlight in this formulation of MRR:

First, mechanisms involve the doings of particular agents: the agent’s reason-recognition gives rise to an understandable pattern; the agent does otherwise.\(^{47}\) I take this to mean that if a process leading to action involves an agent’s acting, then any alternate sequence involving the same type of process must involve that same agent (or a counterpart of theirs) acting also. That is, no other agent, or entity, can act on the agent’s behalf in the alternate sequence cases.

Second, the analysis includes, and Fischer and Ravizza stress, the claim that an agent’s doing otherwise in alternate sequences must be appropriately connected to their reasons for so acting; that is, when we consider an alternate scenario in which there exists a sufficient reason to do otherwise, the agent recognises that reason, and the agent does otherwise, it is expected that the agent does otherwise for that reason in the relevant sense. Similarly, they note, the actual sequence “must also exhibit the appropriate sort of connection between reasons and actions. So, when an agent has guidance control, we [can] assume that he performs the relevant action intentionally.” (1998, p. 64; my italics).\(^{48}\) Importantly, though, this cannot go so far as to require that

\(^{46}\) I rely on this formulation of MRR rather than the one given by Fisher and Ravizza above in order to give a clearer statement of the two counterfactual conditions that a mechanism must meet in order to be MRR. I assume that there is no difference in meaning between my formulation and theirs.

\(^{47}\) This is a feature of Fischer and Ravizza’s own formulation too—“the agent would recognise reasons… and would react to at least one sufficient reason to do otherwise…” (1998, p. 244)(Bold emphasis is mine).

\(^{48}\) If acting intentionally in alternate sequences were not a requirement on a mechanism’s being MRR then we might end up with the following kind of problematic example: Suppose that an agent, call her Sneezy, is such that she only sneezes when there is sufficient reason to do so. What’s more, if there is greater reason to not sneeze (for instance, if she is hiding from pirates and wants to remain undetected) she will refrain from sneezing. However, in every case in which she sneezes or refrains from sneezing, her behaviour is involuntary (and also, therefore, unintentional). Suppose that she sneezes at $t$ in
an agent be actually guiding their behaviour in alternate sequences (without forcing the analysis into circularity, at least).

**Taking responsibility**

Having one’s actions issue from a MRR mechanism is not quite enough for guidance control however: one also needs to have ownership of that mechanism. More specifically, one needs to *make* that mechanism one’s own by *taking responsibility* for it, which is “a matter of having certain (dispositional) beliefs about oneself (and having acquired these beliefs in appropriate ways)” (1998, p. 243).

This concept of *taking responsibility*, Fischer and Ravizza explain, can be taken as shorthand for ‘taking responsibility for the behaviour issued by those mechanisms’ (1998, p. 241) and generally occurs as the result of an agent’s upbringing:

As a child grows up, he is subject to moral education (imperfect as it may be). The child’s parents – and others – react to the child in ways designed (in part) to get the child to take certain attitudes toward himself… [As a result] the child typically acquires the view of himself as an *agent*… he sees that upshots in the world depend on his choices and bodily movements. Further, the child comes to believe that he is a

response to smelling some pepper; it looks as though the mechanism issuing in her sneezing will be MRR: there is at least one alternate sequence in which she recognises a reason not to sneeze at *t* and consequently refrains from doing so, and her response to sneeze-reasons will form an understandable pattern across a range of possible worlds. Still, if Sneezy were to take ownership in of her sneeze-mechanism (in the appropriate way) she would not be actually guiding her sneezing; after all, it is an involuntary action. The appropriate response on Fischer and Ravizza’s part here seems to be that because her sneeze-mechanism does not issue in intentional actions, it is not really the case that Sneezy *does* anything in the actual or alternate sequences, so that mechanism fails to count as MRR and she cannot have guidance control through its operation.
fair target for certain responses… as a result of the way in which he exercises his agency. (1998, p. 241)

The agent takes responsibility for a particular kind of behaviour (and thereby makes the mechanism that issued it their own), in a way, just by coming to see that behaviour as their own. Call the bundle of beliefs this involves a *responsibility attitude*:

**Responsibility attitude:** $S$ has a responsibility attitude towards a kind of behaviour $B$ if (i) $S$ sees $B$ as an upshot of his agency in the world; and (ii) $S$ sees himself as a fair target for any reactive attitudes that $B$ might generate in others.

I take it that a responsibility attitude need not be formed as part of an agent’s upbringing, this is just the normal course for its formation, and that it can be formed at any point in an agent’s lifetime. It is sufficient for an agent’s having guidance control, then, that they take responsibility for a MRR mechanism.

Before moving on to my counterexample to this sufficiency claim, there is an important proviso to taking responsibility worth mentioning: the agent’s responsibility attitude must be based, in an appropriate manner, on their evidence (1998, p. 213).

This condition is intended, in part, to prevent problems that could potentially arise from cases of manipulation: Suppose that Black, not content with his manipulation of Jones, decides to manipulate Jane too by implanting a responsibility attitude for mechanisms of kind $K$ in her mind through electronic stimulation. Jane might, in that case, see herself as the agent of, and an apt target for the reactive attitudes generated by, the behaviour issued by $K$-type mechanisms; however, she will not have guidance control over that behaviour in that her responsibility attitude was not formed in an appropriate manner.
More generally, this proviso means that in order for an agent to have guidance control over the behaviour issuing from a MRR mechanism, they must have formed their responsibility attitude towards that behaviour ‘off their own back,’ as it were, and as a response to their experiences of that mechanism’s operation.

2. THE ANTI-AKRASIA CHIP
As I noted above, my concern with Fischer and Ravizza’s analysis of guidance control lies with the ownership condition. It strikes me that it gets things backwards, so to speak: typically, recognising that we were are in control of various actions and events—that we are actually guiding them through their occurrence—provides us with a reason for viewing ourselves as the agent of those actions or events; on Fischer and Ravizza’s view, however, an agent has to view herself as an agent in order that she can have control.

If guidance control really is to be identified with actually guiding some action, as the initial gloss suggests, then the account is open to various counterexamples in which agents take responsibility for MRR mechanisms and yet, intuitively, do not actually guide the behaviour issued by those mechanisms. Take the following case and its modification:

**AAC 1.0**

A robotics company designs and begins manufacturing an ‘anti-akrasia chip’ (henceforth AAC), a neural implant designed to help weak willed individuals act on their judgements about what they ought to do (where they might otherwise succumb to various temptations). The chip is surgically inserted into the brain and, once activated, begins monitoring the agent’s mental states. If the agent recognises a

---

49 I use ‘weak willed’ here to mean incontinent: failing to act in accordance with what one judges best.
sufficient reason to \( \varphi \) but does not have the strength of will to act in accordance with that reason, the chip will electronically stimulate their brain, take control of their motor-functions and force them to \( \varphi \). In this way, the agent will always act in accordance with what they judge they have most reason to do.

John is a particularly weak willed individual: regularly acting contrary to his judgements about what he has most reason to do when those judgements are accompanied by conflicting desires or attitudes.\(^{50}\) As soon as he hears about the AAC he goes to have it installed, and by the mid-afternoon John’s actions are in perfect accord with his normative judgements; the majority of these actions issuing from his newly active AAC mechanism, which operates as follows:

i) John recognises that there is sufficient reason to \( \varphi \);

ii) if John does not form an intention to \( \varphi \), or his intention to \( \varphi \) is not effective in bringing him to action at the appropriate time, the AAC electronically stimulates John’s motor cortex; and

iii) the AAC’s activity causes John to \( \varphi \) at the appropriate time.

This mechanism is potentially MRR: John’s AAC actions are issued by the AAC mechanism as a direct response to his reasons-recognition (which we can assume is operating in a regular manner) and he would do otherwise if he recognised a sufficient reason to do so.

Suppose, also, that John comes to form a responsibility attitude for his AAC behaviour, thereby making it his own: His peers treat him as the agent responsible for

\(^{50}\) This is not intended to imply that John is incapable of acting in accordance with his judgements: an agent who is always, or almost always, weak willed might simply fail to be the kind of agent that, intuitively, can have control over and be morally responsible for their actions. On the contrary, John is able to act in accordance with his judgements—his akratic actions are not compulsive—he is simply weak: more often than not he lacks the oomph to do what he thinks is best and is too easily seduced by other temptations. Thanks to an anonymous referee for pushing me on this point.
his AAC behaviour and he too, fully aware of the link between his judgements and resulting behaviour, comes to think of himself as the agent of that behaviour. That being the case, his AAC behaviour would be issued from one of his own MRR mechanisms.

Is John actually guiding his AAC behaviour? That is, is he guiding his actions with the kind of control required for moral responsibility? I think not: Suppose that a week before having the AAC installed John had been unfaithful to his wife. Some time after the chip’s activation, John is feeling guilty and recognises that he has sufficient reason to come clean to his wife; nonetheless, he does not want to confess, and in the absence of the AAC he would not even entertain doing so; rather, he would commit himself to keeping this infidelity a secret. However, with his AAC mechanism fully operational, the recognition that there is a sufficient reason to come clean activates the chip, which in turn causes John to confess the whole affair in sordid detail.

John’s AAC behaviour, it seems, can still constitute his acting against his will (or at least, against the will that he wants): John is internally divided in a similar way to Frankfurt’s unwilling addict, the difference being that whilst the addict acts on a desire he does not identify with, John is caused to act by a judgement that he does not want to have made. In each case, the agents are moved to action by forces that they do not identify themselves with and therefore, I submit, just as the unwilling addict is not properly in control of (and actually guiding) his behaviour, neither is John properly in control of his AAC behaviour. If that is the case (and my description of the AAC mechanism as MRR is appropriate) then taking responsibility for a MRR mechanism cannot be sufficient for having guidance control over the behaviour issued by that mechanism.

51 (Frankfurt, 1971)
There is, however, a quick and easy answer to the problem posed by the above case (AAC 1.0); specifically, the AAC mechanism does not meet the conditions for MRR: As noted in §1 above, a mechanism’s being MRR involves an agent’s intentionally acting in the actual and any alternate sequences. In this case, however, John does not seem to do anything at all, let alone act intentionally: the AAC acts on his behalf. In that case, the AAC will fail to be MRR, and so, even if John takes responsibility for it in the appropriate manner he will not thereby gain guidance control over the behaviour it issues.

Fortunately, it is simple enough to modify the AAC case such that John’s AAC behaviour is intentional.

AAC 2.0
The robotics company responsible for the development of the AAC upgrade the software running on their chips to version 2.0. Now, when an agent recognises a sufficient reason to φ, rather than stimulating the motor cortex directly, the AAC stimulates the user’s brain in such a way as to cause the formation of an effective intention to φ. This involves two elements: i) if he has not already formed one himself, the mechanism implanting in John an intention to φ, and ii) the mechanisms ensuring that John maintains and acts upon his intention to φ at the appropriate time. We can imagine that the first element is achieved through direct stimulation to the relevant part of John’s brain, whilst the second element is achieved by whatever means Black ensures that Jones maintains and acts upon his intention to kill the mayor.

The AAC 2.0 mechanism will now operate in the following manner:

i) John recognises that there is a sufficient reason to φ;
ii) if John does not form an intention to $\varphi$ by the appropriate time, the AAC causes in John the formation of an intention to $\varphi$;

iii) the AAC ensures that John maintains his intention to $\varphi$ and that it is effective in bringing about his $\varphi$-ing; and

iv) John intentionally $\varphi$’s at the appropriate time.  

With this modification in place the AAC 2.0 mechanism meets the criteria for MRR: holding fixed its operation across a set of possible worlds $W$ with the same physical laws as the actual world and in which there exists sufficient reason to do otherwise, (i) John’s reasons-recognition across the members of $W$ gives rise to an understandable pattern (remember we are assuming that John’s reasons-recognition is perfectly average); and (ii) in at least one of the possible worlds in $W$ in which the John recognises a sufficient reason to do otherwise, John intentionally does otherwise for that reason.

Supposing that John again takes responsibility for the AAC 2.0 mechanism, he will meet the conditions for guidance control; however, if John was not actually guiding his behaviour in AAC 1.0, I maintain that we have no reason to suppose that he will be actually guiding the behaviour issued by the AAC 2.0; so, if John does enjoy any kind of control over his behaviour, it certainly is not the ‘actually guiding’ kind enjoyed by Sally in Driving Instructor. That being the case, I submit that either i) taking responsibility for the behaviour issuing from a MRR mechanism is not sufficient for having guidance control over that behaviour; or ii) having guidance

---

52 Assuming a broadly causal theory of action according to which one’s actions are intentional just in case they are appropriately caused by one’s reasons.

53 In fact, the AAC 2.0’s operation will cause John will do otherwise in every possible world in which there exists a sufficient reason to do otherwise, thereby making the AAC mechanism strongly reasons-reactive. I take it that there is no issue with this given the wording of WRR (would react in at in at least one world…)
control over some $x$ does not amount to having actually guided $x$ (as Fischer and Ravizza’s initial gloss would have us understand it).

For the remainder of the paper my talk of `John’s AAC mechanism’ can be understood as referring to the modified 2.0 case.

3. OBJECTIONS AND REPLIES
Assuming for the time being that guidance control does amount to actually guiding some action, let’s formalize my argument from the previous section as follows:

1. John’s AAC mechanism is MRR;
2. John takes responsibility for the behaviour issued by his AAC mechanism;
3. John does not have guidance control over the behaviour issued by his AAC mechanism; and
4. therefore, taking responsibility for a MRR mechanism is not sufficient for having guidance control over the behaviour issued by that mechanism.

In the following three sub-sections I’ll consider a number of objections to each of these premises: that John’s AAC mechanism is not genuinely MRR in §3.1; that John does not, and perhaps cannot, appropriately take responsibility for his AAC mechanism in §3.2; and that John does in fact have guidance control over his AAC behaviour in §3.3.

Finally, in §3.4. I’ll consider some of Fischer’s responses to other critiques, in particular from Stump (2002) and Long (2004), and whether they might be applied to AAC 2.0.

3.1 The AAC mechanism is not MRR
An obvious line of objection would be to simply deny premise (1), that John’s AAC mechanism is genuinely MRR. If it is not then, whether he takes responsibility for it or not, John cannot be exercising guidance control through its operation.
Remember that, in order to be MRR, the AAC mechanism needs to involve both (i) John’s reason-recognition across a set of possible worlds in which there exists sufficient reason to do otherwise giving rise to an understandable pattern (where some of the reasons recognised are moral); and (ii) John’s doing otherwise in at least one of the possible worlds in which he recognises a reason to do so, and doing so for that reason in the appropriate sense.

As stipulated in the example, the AAC is linked directly into John’s normal reason-recognition faculties: When he recognises a reason to φ, it electronically stimulates his brain so as to produce in him an effective intention to φ. So, assuming that John himself is RRR and can recognise that certain reasons exist for acting in this way or that, the AAC mechanism will also be RRR (and will therefore meet condition (i)).

In that case, a plausible objection to (1) will have to show that the AAC mechanism fails to meet condition (ii): that it does not involve John’s intentionally acting otherwise for the reasons that he recognises in the alternate sequences.

Such an objection could take one of three forms:

i) John’s behaviour is not properly intentional;
ii) John’s behaviour is causally deviant; or
iii) John does not act for a reason in the relevant sense.

I shall take these points in turn:

First, because it is the AAC that causes John’s intention to do otherwise in the alternate sequences, a defender of the Fischer-Ravizza account might claim that his actions are not properly intentional: An agent, we might think, cannot by caused to have a genuine intention by a source external to their own psychology. If John’s
actions are not intentional, in the actual or alternate sequences, then the AAC mechanism is not MRR.

This particular line of objection does not appear to be open to the Fischer-Ravizza account, however: as noted above, Fischer and Ravizza rely on Frankfurt cases such as Jones and Black to motivate their distinction between guidance and regulative control. In doing so, they endorse the possibility of agents being caused to act intentionally by entities external to their own agency: In order for Jones and Black to involve Jones’s being unable to do otherwise, Black needs to cause Jones to intentionally kill the mayor in the alternate sequences; otherwise, we could maintain that Jones can do otherwise as his killing the mayor intentionally and killing the mayor unintentionally through coercion or stimulation are different act types (if the latter counts as an action at all). That being the case, and given the stipulation that the AAC ensures that John acts on his intention in the same manner that Black ensures Jones kills the mayor, it looks unprincipled for a defender of the view to argue that the AAC mechanism cannot cause John to intentionally φ.\textsuperscript{54}

Second, it might be argued that, even though his behaviour is caused by his recognition that he has certain reasons to act, John’s AAC behaviour turns out to be causally deviant; that is, whilst his behaviour is caused by his reasons, John’s AAC behaviour is not caused by his reasons in the right kind of way.

Consider the following case from Donald Davidson that illustrates this kind of deviation between reason and action:

\textsuperscript{54} It might be objected that in Jones and Black all that Black does is force Jones to act on an intention that he had already formed (and was about to abandon), whereas in the AAC case, the device creates a new intention in John. This difference is easily removed, though: we could stipulate that John had intended to confess to his wife about his affair, say, but later abandoned that intention, so the AAC’s activation simply forces him to act on an intention that he himself formed. Still, due to the way that the AAC’s operation fits in-between reason recognition and action it will count as MRR, and still John will not have control.
Nervous Climber: A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold. (Davidson, 1973, p. 79)

The climber in this case certainly drops the man because he wants to rid himself of the weight and danger associated the man he is holding and knows that he can do so by loosening his grip on the rope; however, his action is not guided by those reasons, or by him, in the right sort of way: his action is brought about by those reasons somewhat accidentally and so, we might think, he does not really act at all when he loosens his hold. If something similar is going on in the AAC case then it looks like the AAC mechanism will not meet the requirements for MRR after all: if John’s AAC actions are causally deviant, then they are not actions at all (according to the standard casual theory of action, at least).55

As above, however, the Fischer-Ravizza account does not appear to be entitled to this objection given its ownership element: causal deviance occurs whenever the relevant behaviour to some rationale is produced by the wrong kind of mechanism; more precisely, by a mechanism that is not expressive of the agent’s own agency or guidance. But, according to Fischer and Ravizza’s own conditions, the AAC mechanism is of precisely the right kind for producing morally responsible action: it is weakly reactive to the presence of reasons and produces the right pattern of behaviour across a range of possible worlds; what’s more John has made it his own by taking responsibility for it. If after all that it is still the wrong kind of mechanism to

55 Thanks to an anonymous referee for this point.
produce non-deviant action, we are owed some explanation of casual deviance that does not simply terminate in an agent’s ownership of the mechanism issuing in their behaviour.

From another perspective, the problem of causal deviance just is a problem of guidance: An agent’s behaviour is caused by and in accordance with their reasons, but they do not genuinely guide that behaviour. Now, Fischer and Ravizza do not make explicit whether their account is supposed to apply only to non-deviant action, but given this link between guidance and deviance it would be surprising if it was completely silent on such cases. What’s more, if it turns out that a mechanism can only be MRR if it issues in non-deviant (guided) actions, there may be some circularity in the account: a mechanism can only be MRR if it issues in non-deviant actions (those guided by the agent), and an action is only guided by the agent (non-deviant) when it issues from one of their own MRR mechanisms.

Finally, a similar objection might be put forward in terms of acting for reasons: whilst John’s behaviour is performed for reasons in the sense that it is prompted by his recognition of certain reasons to act, John does not really act for those reasons in the relevant sense for intentional action, and if he does not act for a reason in the relevant sense then the AAC mechanism is not MRR.

Unfortunately, the sense of acting for a reason relevant to intentional action is notoriously difficult to pin down. Fischer and Ravizza themselves seem to endorse Robert Audi’s account of acting for a reason, according to which it is enough for an agent’s having φ’d for a reason r that he would give r as the reason for his action if he were asked for an explanation (Fischer and Ravizza 1998, p. 64);\textsuperscript{56} and, as it is John’s recognition of reasons qua reasons for action that activates the AAC mechanism, it

\textsuperscript{56} See (Audi, 1986).
certainly seems possible for John to cite the reasons he recognised as the reasons for his AAC behaviour (knowing, as he does, how the AAC mechanism works).\textsuperscript{57}

Alternatively, we might think that an action is done for a reason \textit{in the right way} only when the mechanism by which the reasons are translated into action is properly the agent’s own. However, according to Fischer and Ravizza’s own conditions, John \textit{makes} the AAC mechanism his own by taking responsibility for it.

For an objection of kind (iii) to work, it will need to provide a different account of acting for reasons that rules out John’s case, and until such an account has been provided it seems unprincipled to insist that John’s AAC behaviour is not performed for a reason.

All of these objections seem to suggest that there is some other notion of ownership at play in the analysis of MRR: an agent needs to take responsibility for a MRR mechanism in order to exercise guidance control, but that mechanism needs to be the agent’s own mechanism in some other sense too, or issue in actions that are properly the agent’s own, in order that it be MRR. It is not at all obvious, though, what this secondary notion of ownership, call it \textit{ownership*}, turns out to be and how to distinguish it from the primary ownership granted by taking responsibility. What’s more, the requirement that agents have ownership* over their mechanisms in order that they be MRR threatens to make taking responsibility somewhat redundant: if an agent needs to own* a mechanism and the behaviour it issues in order for it to be MRR, then they already seem to have ownership over that mechanism and their actions in a sense relevant to control and responsibility; it is simply not obvious, in that case, what more there is for taking responsibility to achieve.

\textsuperscript{57}Interestingly, returning to the previous objection on causal deviance, whilst John plausibly would cite the reasons he recognised as the reasons for his actions, the climber in Davidson’s case plausibly \textit{would not}, making this look less like a standard case of causal deviance.
3.2 John has not taken responsibility for the AAC mechanism

The second line of objection would be to reject (2), that John has taken responsibility for the AAC mechanism. If he doesn’t have the relevant responsibility attitude then he can’t have ownership of the mechanism and therefore will not have guidance control over the behaviour it issues.

Such an objection might plausibly be based upon the appropriateness constraint that Fischer and Ravizza place on taking responsibility: As noted in §1 above, a responsibility attitude must be developed, in an appropriate manner, on the agent’s evidence (1998, pp. 213).

This constraint, they acknowledge, is “intended (in part) to imply that an individual who has been electronically stimulated to have the relevant view of himself… has not formed this view of himself in the appropriate way” and therefore lacks guidance control despite meeting all of other the requirements (1998, pp. 235-6). Due to the manner of the AAC’s operation, perhaps it will be contended that John has not, and perhaps cannot, develop his responsibility attitude, in an appropriate manner, on the evidence that he has.

Fischer and Ravizza do not offer an analysis of this appropriateness condition, so it is hard to assess why John’s taking responsibility might be inappropriate. Presumably, it will because either: (a) the process by which he formed the attitude was inappropriate; (b) the attitude he formed was inappropriate given the evidence he had; or (c) the evidence itself was inappropriate to base a responsibility attitude on.

Of the three options, (a) and (b) look the most plausible, especially if we consider the type of case the condition was introduced to prevent. In response to (a), John forms his attitude by coming to see himself as the originator of his AAC behaviour (on the basis that it is his judgements that end up causing his actions) and
on the basis of the reactive attitudes of his peers in response to his behaviour. This seems like an appropriate process given what Fischer and Ravizza claim goes on during a child’s moral education. At least, nothing is obviously wrong with forming it in this way.

In response to (b), his responsibility attitude itself looks appropriate given that it was based upon his peers treating him as though he was morally responsible for his AAC behaviour. However, he may also be aware of his strong desires to act contrary to his better judgement (his desire not to tell his wife about his infidelity, for instance), and we might think that this awareness makes it less plausible that his responsibility attitude is appropriately based upon all the evidence he has: Perhaps if one strongly desires to not-A, then it is inappropriate to see oneself as responsible for any A-ing that one might produce.

This objection cannot be right, though: we quite often act responsibly whilst having strong desires to act otherwise and, what’s more, in many of those situations it will be entirely appropriate to hold a responsibility attitude towards ourselves. Think of any case in which an agent overcomes a strong fear: for instance, James is terrified of spiders and yet faces his fear and intentionally picks up a tarantula and allows it to walk across his palm; his strong desire not to pick up the spider ought not prevent him from being able to see himself as responsible for this behaviour; after all, he conquers his fear and deserves praise for doing so. Neither, in that case, should John’s strong desires not to do what he judges best make it inappropriate for him to take responsibility for his AAC mechanism.

58 We could even change the example to stipulate that the AAC was implanted in John’s brain at birth so that he develops his responsibility attitude as part of his moral upbringing.
59 Thanks to an anonymous referee for pushing me on this point.
This leaves (c), that the evidence itself was inappropriate. This looks like the least plausible reading of Fischer and Ravizza’s condition: ‘in an appropriate manner’ suggests that it concerns the process by which the attitude was formed, and not the evidence that it was based upon. However, perhaps it will be maintained that developing a responsibility attitude on evidence in the appropriate way involves one’s developing their responsibility attitude on appropriate evidence.

Potentially, the dissonance that John experiences between what he wills and what he actually does might make his taking responsibility in this case seem inappropriate. Intuitively, his experience of his AAC behaviour ought to tell him that he is not actually guiding that behaviour. John’s peers are wrong to treat him as if he is responsible for his AAC behaviour and John himself is wrong to base a responsibility attitude on his first-hand experience of AAC behaviour: it is not genuine evidence of control.

Unfortunately, this response is not available to the Fischer-Ravizza view: taking responsibility is conceptually prior to being in control, so being in control cannot be a condition on appropriately taking responsibility. If John does not have guidance control because he cannot appropriately take responsibility, and cannot appropriately take responsibility because he is not actually guiding his behaviour (and so does not have genuine evidence of guidance control), then the analysis turns out to be circular.

Unless we can give a non-circular reading of Fischer and Ravizza’s appropriateness condition, and no such reading is particularly forthcoming, taking responsibility simply cannot be sufficient for having guidance control over a moderately-reasons responsive mechanism.
3.3 John does have guidance control over his AAC behaviour

A certain kind of response, which we might call the *hard-headed* option, would be to deny (3) and assert that, in fact, John *does* have guidance control over his behaviour.

As before, there are a number of lines that this kind of objection could take:

(i) the AAC gives John control by allowing him to overcome his akratic tendencies;
(ii) the AAC acts as a prosthetic, allowing John to guide his behaviour more effectively;
(iii) John would have guidance control by the ‘tracing method’ (to some extent); or
(iv) guidance control was not supposed to be an analysis of any pre-theoretical notion of control or guiding, anyway.

Again, I take these points in turn.

First, someone might argue that, because the AAC prevents John from acting akratically it actually helps him to remain in control by allowing him to remain continent and exercise self-control over his behaviour (to bring it in line with his judgements).

I think we can dismiss this kind of initial objection quickly: If all that self-control and continence mean is that an agent’s behaviour is correctly aligned with their judgements, then this does not look to be the same as guidance control, it is certainly not the kind of control exercised by Sally when guiding her car.

Second, we might think that, due to the way that it integrates with John’s brain, the AAC should not be thought of as an external manipulator but, rather, as a prosthetic device for helping John to connect his reasons with the appropriate actions (the device is certainly marketed in this way, at any rate). The chip, in the example, serves as a
conduit between John’s judgements and his resulting intentions and actions, and, once it is installed and fully integrated with his brain, John’s neural psychology will be functionally identical to any ‘normal’ (non-manipulated) strong-willed agent: he will recognise reasons and translate those into the relevant intentional actions. So, if normal strong-willed agents are capable of actually guiding their behaviour, then John ought to be capable also.\footnote{Thanks to an anonymous referee for this point.}

This is an extremely tricky area and there certainly is a case to be made for the AAC being a kind of prosthetic; however, if we are careful to distinguish between prosthetics that enable an agent to exercise guidance and prosthetics that guide on an agent’s behalf I think it will become obvious that the AAC really does prevent John from guiding his behaviour.

Consider a variant of the ACC called ‘the Deliberator.’ Like the AAC, the Deliberator is a neural implant that monitors the reasons that an agent recognises, along with their relative strengths, and, when the time for a decision arrives, feeds those inputs into an algorithm to determine what the agent has most reason to do. Unlike the AAC, however, the Deliberator only produces a judgement about what the agent ought to do, which the agent is then free to act upon or ignore. We can further imagine that the Deliberator takes into account all of the agent’s values and characteristics when coming to a decision so that it produces exactly the judgement that they would have come to, were their deliberative faculties functioning at their very best, but in one hundredth of the time.

As a prosthetic device, the Deliberator allows for an agent to more effectively exercise guidance over their actions: it speeds up the slow organic deliberation of normal agents by replacing it with quick and accurate electronic calculations, and it...
allows agents who are unable to effectively deliberate on their own (perhaps due to some neurological disorder) to function as a psychologically standard agent. Importantly, the Deliberator does not cut the agent out of the picture altogether: the agent receives a judgement that they can then choose to act upon or veto if it is not to their liking.

Contrast this to the AAC case: The AAC not only deliberates for John, it also forms an intention to act and ensures he sticks to it, so he has no chance to veto any of the decisions it makes on his behalf. The AAC insulates John’s deliberation and intention formation processes from his conscious will, cutting him out of the picture altogether. It is still his reasons-recognition that starts the process off, but from that point on he is a bystander to the process, and there is a good deal more to guiding one’s actions than having them conform to one’s reasons in a regular pattern: one needs to get involved in this process and mediate between the reasons that one has for acting, adding one’s own motivational force to whichever motive one favours (where this is not necessarily the motive that one identifies as being best justified). The AAC, I submit, is a prosthetic that guides on the agent’s behalf and therefore one that removes, rather than enables or enhances, an agent’s ability to guide her behaviour. It does not matter how integrated the prosthetic becomes, if it subverts an agent’s will in the process leading to their actions, it can only remove, rather than enable, that agent’s ability to effectively guide their behaviour.

Third, we might think that the AAC behaviour John exhibits is a good candidate for being treated to what Fischer and Ravizza refer to as the ‘tracing method.’

Roughly, whilst a particular action may not issue from a MRR mechanism, Fischer and Ravizza claim that we can nonetheless look for guidance control “at
various places along the way to the action... More specifically, we can look for [MRR] in the formation of the relevant trait, its retention, or its expression” (1998, pp. 87–8).

In John’s case he has no control over whether he retains the trait—you suppose that he maintains his judgement, much to his despair, that he is better off with the chip than without it, and so cannot intentionally have the chip removed due to its continued operation—but he did have control over its formation, and we might think that he has some control over its expression also: if he actively does what he judges best himself, before the AAC kicks in, then the mechanism will not operate, so he can prevent himself from losing control by doing what he judges best as soon as he so judges; also he might be able to avoid doing things he does not want to do just by trying not to think about them and thereby not become aware of what he has sufficient reason to do.

The thing to note in response to this objection is that the tracing method is supposed to establish that an agent is morally responsible for some action which issued from a non-MRR mechanism, not that they have guidance control over that action’s performance. Fischer and Ravizza’s own examples of the tracing method focus on agents who are not in control of their behaviour and yet are morally responsible for what they do due to their having exercised guidance control in getting themselves into their current situation.61

In John’s case, his AAC behaviour does issue from a MRR mechanism which he takes responsibility for, so Fischer and Ravizza’s conditions imply that he ought to have exactly the same kind of control over his AAC behaviour that Sally exercises over her car. The conditions entail, that is, that John is actually guiding the AAC behaviour, not just that he is accountable for it. This objection therefore seems to be

---

misplaced: The tracing method suggests only that John may be accountable for his AAC behaviour, not that he is actually guiding it.

This point also demonstrates that moral responsibility and control can come apart. We might think of John as morally responsible for his AAC behaviour given that he got himself into this situation by his own choice. However, this in no way entails that he must be actually guiding his AAC behaviour, putting some pressure on Fischer and Ravizza’s claim that guidance control is necessary and sufficient for moral responsibility.

A final hard-headed objection that I shall consider might run as follows: John does have guidance control but this is just a placeholder for the position one is in if one’s actions issue from a MRR mechanism for which one has taken responsibility. That is, whilst it might be true that John is not actually guiding his AAC behaviour, and nor is he in control of that behaviour in a commonplace or folk-sense, all that is simply irrelevant since guidance control was not supposed to be an analysis of any pre-theoretical conception of control or guidance anyway.

Guidance control, according to this line, is a term relevant only to working out whether an agent is morally responsible and should not be thought of as an analysis of our folk concept of control.

It seems unlikely to me that this is what Fischer and Ravizza intend by ‘guidance control’ given their initial gloss; however, if this is the case, and guidance control is just a placeholder for the state one is in when one can be held morally responsible then (i) it’s not clear why Fischer and Ravizza even bother to talk in terms of control: it would be clearer to simply give all of their claims in terms of moral responsibility (for example: ‘an agent is morally responsible when their behaviour issues from one of their own MRR mechanisms’); and (ii) guidance control is really
nothing like the control that Sally has over her car: Sally’s case displays a kind of
control, perhaps more fundamental than guidance control, which amounts to actually
guiding some process and that intuitively is relevant to moral responsibility. If
guidance control does not capture this notion then my example may well fail to show
that the conditions Fischer and Ravizza stipulate for having guidance control are
insufficient; however, once we drive a wedge between guidance control and actually
guiding, it’s not clear why we should want an analysis of moral responsibility in terms
of guidance control.

The intuition behind Fischer and Ravizza’s account was that agents need not be able to do otherwise in order to be morally responsible for their behaviour: they need only to have been the one who actually guided their behaviour. If guidance control does not amount to actual guidance then it is not clear how it is related to moral action and why having guidance control over an action would entail being morally responsible for its performance.

To sum up, if guidance control is supposed to capture an agent’s intentionally guiding some action (as Sally does), then John does not have guidance control over his AAC behaviour, and in that case the conditions stipulated by Fischer and Ravizza for having guidance control are insufficient; if, however, guidance control does not amount to actually guiding some action, then it is not clear why we should favour an analysis of moral responsibility in terms of guidance control over the more commonplace ‘actually guiding’: the latter would appear to be far more informative.

### 3.4 Individuation, inputs and fair opportunities

In this final sub-section I should like to move away from direct rejections of the premises of my argument to look at some of Fischer’s responses to critiques by other authors, in particular his (2004) reply to Stump and (2010) reply to Long. Both of
these responses suggest a tension between the manner in which Fischer intends the
guidance control account to be applied and its official conditions; examination of this
tension reveals that there may be extra conditions on having guidance control, not
explicitly stated in Responsibility and Control (1998), which the AAC mechanism
might fail to meet.

Stump’s (2002) worry is much the same as my own: that the Fischer-Ravizza
account allows for agents to have guidance control over mechanisms that clearly
involve a high degree of manipulation so long as those mechanisms are suitably
responsive to the agent’s reasons. She imagines a case (based upon Robert Heinlein’s
The Puppetmasters) in which an agent, Sam, has had his mind taken over by an
intelligent alien ‘master’ as part of a wider scheme by that alien’s race to conquer
Earth. When the master takes over Sam’s body, it takes his consciousness ‘off-line,’
leaving it to run pretty much as it always does but removing from it the ability to affect
his behaviour. It is the master’s consciousness alone that determines how Sam acts
(2002, pp. 47-8). She goes on:

Since it is crucial to the alien plan that their taking over human beings
be undetected in the early stages of the invasion, they are careful to make the behavior of people like Sam correspond to the behavior Sam
would normally have engaged in had he not been infected with the alien.
So when, under the control of the alien, Sam does A, it is also true that if there had been reason sufficient for Sam in his uninfected state to do
not-A, the alien would have brought it about that Sam in his infected state did not-A. In this case, then, Sam acts on a mechanism that meets
Fischer and Ravizza’s condition for being strongly reasons responsive. (2002, pp. 47-8)
Stump then imagines that the master alien reveals itself to Sam and convinces him to take responsibility for this mechanism, thereby making it his own (2002, pp. 49-50). From then on, whenever he is being controlled by the alien master, Sam is acting on his own MRR mechanism; clearly, though, Sam is not the one guiding his behaviour.

Fischer’s response to this is that, of course, if you individuate the mechanism on which Sam acts as broadly as ‘manipulation by an external source,’ then it will turn out to be moderately reasons-responsive; but the Fischer-Ravizza account’s proposed way of dealing with manipulation cases is to individuate the relevant mechanism in a far more narrow manner: ‘manipulation of this specific sort,’ for instance (Fischer, 2004, pp. 152-3). The relevant mechanism to hold fixed in Sam’s case, then, is not ‘manipulation of Sam by the alien master,’ but something like ‘manipulation of Sam’s brain such that he is caused to A;’ and with that narrower mechanism held fixed Sam will A irrespective of any reasons to not-A in alternate sequences.

Stump’s alien master case shares many of the features of my own AAC case, save for one important feature: in her case, an agent is manipulated by the actions of another agent; in my AAC case, the process leading to John’s actions does not involve any agents other than John himself. Nonetheless, a similar response might be given to AAC 2.0: of course the AAC comes out as MRR if you individuate the mechanism issuing in John’s behaviour so broadly as: ‘manipulation by the AAC in accordance with John’s reasons.’ However, when the AAC forces John to intentionally A, we should assess the reasons-responsiveness of the mechanism issuing in his behaviour by holding fixed something like ‘the AAC manipulating John’s brain in such a way that he intentionally As.’ If this narrower mechanism is held fixed then it is clear that John will be incapable of doing anything other than A-ing; the description of the mechanism being held fixed includes that John is forced to intentionally A, so no
matter what reasons exist for not A-ing the AAC will manipulate John’s brain as though there is a sufficient reason to A and force him to intentionally A. That being the case, the AAC mechanism is not MRR and John does not guide the behaviour it produces.

It is difficult to know exactly how to respond to this kind of objection because, as Fischer himself acknowledges (2004, pp. 166-7), the Fischer-Ravizza account does not contain an explicit account of mechanism-individuation and so there is no principled reason for holding fixed the narrow rather than the broad manipulation mechanism in the AAC case, this is simply left up to intuition.

The obvious response is to simply restate the above admission: there is no principled reason why we should not hold the broad rather than the narrow AAC mechanism fixed: it is not at all obvious, even if left to intuition, that the narrower mechanism is the relevant one for assessing John’s AAC mechanism; after all, remember that we can describe the AAC case such that the device be integrated with John’s brain to such a degree that it is functionally identical to the deliberative faculties of a ‘normal’ strong-willed agent (and still, John would not guide his behaviour). Unfortunately, butting heads does not get us very far.

I think, though, that there is some indication of a principle at work in this objection, and by drawing it out we can see, again, that it is the ownership element of the guidance control analysis that is at issue.

First, consider the following question: if we must hold fixed the narrower mechanism in the AAC case, why ought we not hold fixed an analogously narrow mechanism in cases of normal un-manipulated action? That is, when John’s AAC causes him to intentionally A, we are told that we ought to hold fixed the narrowly individuated mechanism type:
n-AAC ‘the AAC manipulates John’s brain such that he As’

rather than the more broadly individuated mechanism type:

b-AAC ‘the AAC manipulates John such that he acts in accordance with his reasons-recognition.’

Whereas, when a normal non-manipulated agent intentionally As we ought not hold fixed the narrowly individuated mechanism type:

n-NORM ‘the agent’s deliberative faculties operate such that she As’

but instead hold fixed the broadly individuated mechanism type:

b-NORM ‘the agent’s deliberative faculties operate such that she acts in accordance with her reason-recognition.’

If we were to hold fixed n-NORM in cases of normal un-manipulated action, because the description of the mechanism’s type is so narrow as to include that a specific action is performed, no deliberative mechanism would turn out to be MRR: in any alternate sequence, no matter what the strength of reasons for not A-ing, n-NORM will issue in the agent A-ing.

Why should the two cases be treated differently, though? Potentially, because the AAC case involves manipulation, whereas the ‘normal’ case does not. However, to draw this distinction, between a manipulation and non-manipulation case, before considering whether an agent has taken responsibility for the mechanism issuing in their behaviour suggests that there is already some notion of ownership at play here: a case involves manipulation only in as much as it involves an agent’s behaviour being brought about, at least in part, by a mechanism that is not their own. So in order for
this objection to hit its target, John must lack some different, perhaps more fundamental, kind of ownership over his AAC mechanism to the kind granted by taking responsibility; as in §3.1 above, call this ulterior notion *ownership*.

This objection might plausibly be applied to Stump’s case because in her example the mechanism by which Sam acts involves the actions of *some other agent*; namely, the alien master. In that respect, there is an intuitive way of distinguishing between Stump’s case and that of a ‘normal’ agent: in her case the mechanism leading to action contains the embedded actions of the alien master as *a part of its operation*, whereas in the ‘normal’ case the mechanism issuing in action involves no such embedded actions.\(^6\) We might, then, offer the following principle to the Fischer-Ravizza account for mechanism individuation:

\[ (\text{EMBED}) \quad \text{whenever a mechanism’s description involves the performance of a particular action } A, \text{ in holding fixed that mechanism we must also hold fixed the performance of that very same action } A \text{ in all alternate sequences.} \]

Given this principle, Sam would not do otherwise in any of the alternate sequences in which there existed a reason to do so as, due to the embedded actions of the alien master, we would need to hold fixed the narrowly individuated mechanism that involved the alien master’s causing him to act in a particular way.

It is here that AAC 2.0 departs from Stump’s alien master example: like the mechanism in the ‘normal’ case, the AAC mechanism contains no embedded actions: the AAC is not an agent, and it does not act; it merely plays a functional role in the

\[^6\text{It is important that we distinguish between the action that a mechanism issues in and the process by which that mechanism operates: An agent’s } A\text{-ing might be issued by their practical deliberation mechanism, but } \text{issuing-} in-A\text{-ing is not thereby a part of the process by which their practical deliberation mechanism operates (else it would only be able to issue in } A\text{-ings).} \]
processing of John’s reasons-recognition into action. If all that mattered for ownership was taking responsibility, then we would not be in a position to say whether John’s case, or the ‘normal’ case, involved manipulation at the outset. What’s more, once we remove all mention of manipulation from the descriptions of n-AAC and b-AAC’s operation, both the narrowly individuated n-AAC and n-NORM and the broadly individuated b-AAC and b-NORM will conform to the same narrow and broad mechanism types:

n-MECH     $K$ operates such that $S$ As
b-MECH     $K$ operates such that $S$ acts in accordance with $R$

(Where $K$ is the kind of mechanism operating and $R$ is some reasons-responsive faculty)

To distinguish between the two cases then, without first assuming that the ‘normal’ agent owns* their deliberative faculties whilst John does not own* the AAC mechanism, simply looks ad hoc.

If there is a notion of ownership* at play here, though, it is not at all obvious that one also need to take responsibility in order to guide one’s behaviour. Surely, if a mechanism is psychologically one’s own and it is suitably responsive to reasons, it simply does not matter whether one thinks of oneself as an agent when that mechanism operates. Without a more worked out explanation of ownership* we are not in a position to say, but there is certainly the threat of making taking responsibility redundant by introducing some more fundamental notion of ownership.

To sum up, it is not clear whether the Fischer-Ravizza account is entitled to this defence, given that it includes no explicit method for individuating mechanisms. Our best bet for identifying some principle at work here is to focus on the fact that the
AAC case involves manipulation not normally present in practical deliberation; this, however, suggests that there is some hidden ownership condition in the Fischer-Ravizza account, determined prior to taking responsibility, which in turn suggests that the official ownership condition on guidance control is insufficient.

Moving on to Long’s (2004) worry, Long considers a case in which a manipulator feeds inputs into an agent’s MRR mechanism right before they come to a decision in order to change the output to their liking.

For instance, Block might want Schmidt to vote for Hitler to be given supreme power over Germany; sensing that Schmidt is about to vote against Hitler having this power, Block sets to work: adding new inputs to the very same mechanism that operates in the actual sequence (suppose that these come in the form of reasons for voting in favour of Hitler) so that when the time comes Schmidt’s normal deliberative faculties issue in his voting for Hitler.

Schmidt, in this case, appears to act on his own MRR mechanism, and so to have guidance control over his behaviour according to the Fischer-Ravizza account, but, clearly, he is being manipulated by Block (and is not really responsible for his actions) (Long 2004, pp. 157-9).

Fischer’s response to this case is as follows:

The suggestion in question is that the implantation or manipulative induction of “reasons” is done immediately prior to the choice and subsequent behavior, and that there is thus no reasonable or fair opportunity for Schmidt to reflect on or critically evaluate the new input in light of his standing dispositions, values, preferences, and so forth.

I contend that when “inputs” are implanted in a way that does not allow for a reasonable or fair opportunity for the agent to subject
those inputs to critical scrutiny in light of his or her normative orientation, then such manipulation does indeed remove moral responsibility. Such manipulation typically “changes the mechanism.”

(2004, pp. 180-1)

Schmidt is not morally responsible for his voting for Hitler, because the mechanism on which he acts is not the one that he has taken responsibility for: the addition of extra reasons immediately before his decision without the opportunity to filter these reasons through his normative orientation changes the kind of mechanism issuing in his behaviour.

Perhaps a similar complaint can be lodged against my AAC case: The AAC produces intentions that are appropriately based upon John’s reasons to act, but because it also ensures that he acts upon these intentions it does not provide him with a fair and reasonable chance to filter those intentions through his normative orientation. That being the case, John ought not to be held morally responsible for his AAC behaviour, and does not have guidance control over it.63

Again, it is not clear whether the Fischer-Ravizza account is entitled to this kind of response given its official conditions. This requirement that the inputs to one’s mechanisms be filtered through one’s normative orientation is presumably part of guidance control’s ownership element; but it is not at all obvious how this requirement is supposed to fall out of the official conditions for appropriately taking responsibility for a mechanism, so this appears to be another extra condition on having guidance control not explicitly stated in Responsibility and Control (1998). Adding this condition might help to alleviate the worry raised by the AAC case, but only by

63 Thanks to an anonymous referee for this point.
suggesting that there is more to owning a mechanism than *taking responsibility* for it, which has been my argument throughout the paper.

In any case, it is not clear that the AAC mechanism should be conceived of in the same way as Long’s manipulation case: the AAC example is set up in such a way that the AAC does not simply add an input to John’s normal deliberative mechanisms; rather, it replaces those mechanisms and, once he takes responsibility for it, becomes *his* mechanism for translating reasons into action. The intentions it produces do not need to be filtered through John’s normative orientation because, once the AAC mechanism has been made his own, its operation just *is* the filtering of reasons through his normative orientation. Of course, the intention could be re-filtered through John’s normative orientation and deliberative faculties, but this would simply involve re-running it through the AAC mechanism.

To sum up, if we conceive of the AAC mechanism as adding inputs into John’s deliberative mechanisms, this response may well show that John does not have guidance control over his AAC behaviour; it only does so, however, by adding extra conditions onto the account of appropriately taking responsibility and thereby showing that the official account is insufficient. If, on the other hand, we conceive of the AAC as a mechanism of John’s into which inputs are fed, the response does not seem to apply at all.

4. SUMMARY AND PROSPECTS
I have argued that taking responsibility for a MRR mechanism is not sufficient for an agent’s having guidance control over the behaviour issuing from that mechanism. An attitude on the part of the agent, as the AAC case shows, is not able to make the relevant kind of difference to whether an agent is actually guiding a particular action.
Responses to this case tend to either force the Fischer-Ravizza account into circularity or further demonstrate that taking responsibility, as it is described in *Responsibility and Control* (1998), is insufficient for mechanism ownership (by introducing further constraints on making a mechanism one’s own); indeed, many of the potential objections to the AAC case, as well as some of Fischer’s responses to manipulation cases, seem to suggest that there is some ulterior notion of ownership at work in the analysis, as above call this *ownership* *, that is not conditional upon taking responsibility and which prevents the analysis from applying to cases such as AAC 2.0. Until this condition is made more explicit we are not in a position to determine whether John’s case is a genuine case of guidance control. In any case, I think that this suggests that the ownership element included in the official conditions for guidance control is not suitable for an account of actually guiding some action.

The problem I’ve raised for Fischer and Ravizza’s account largely arises out of their treatment of moral responsibility and guidance control as interchangeable concepts. Having a responsibility attitude may be necessary for being morally responsible for one’s behaviour; indeed, it may be that unless one understands what it means to be a moral agent and sees oneself as such, one is not in a position to be held accountable for one’s actions; what’s more, control in the form of ‘actually guiding some action’ is plausibly a necessary condition on being morally responsible for what one does. However, whilst the link between moral responsibility and control might prove informative for this reason, they are distinct concepts: there are situations in which they come apart (such as the AAC case, if we think that John is morally responsible for his behaviour), and as such they deserve distinct approaches.

Interestingly, if, as I suggest here, we treat guidance control and moral responsibility separately, the two ownership conditions mentioned above can plausibly
be attributed two distinct roles: ownership* determines whether an agent has guidance control over a mechanism (and is therefore guiding their behaviour in the way that Sally guides her car), whilst taking responsibility determines whether an agent is morally responsible for a particular action that they had guidance control over. It is only when guidance control and moral responsibility are treated interchangeably that we seem to need only one ownership condition for mechanisms, so developing the account in this way removes the threat of redundancy facing taking responsibility as well as presenting a plausible way of dealing with the AAC case and any similar cases: John does not have ownership* over the AAC mechanism so he does not actually guide the behaviour it produces; it is then inconsequential whether he mistakenly takes responsibility for that mechanism.

Note, finally, that this development also resolves the potential circularity in John being unable to appropriately take responsibility for his AAC mechanism: he genuinely is not guiding the AAC behaviour, in that he does not own* the AAC mechanism, so it would be inappropriate for him to take responsibility and gain moral-ownership of that behaviour.

Obviously, in order for this development to succeed, some work needs to be done to make explicit just how one gains ownership* over a mechanism. The AAC case itself suggests that what is missing from John’s situation is the ability to veto the decisions made on his behalf by the AAC.

In any case, control in the sense of actually guiding some action is a notion that anyone interested in moral responsibility, and action more generally, ought to be interested in analysing. It is a mistake, I submit, to start with moral responsibility and work back to control; we should instead start with an analysis of control and see what further conditions are required to arrive at moral responsibility.
REFERENCES


Abstract: Manley and Wasserman's recently proposed analysis of dispositions, '(PROP)', suggests that we take a proportional approach to analysing dispositions, according to which an object is disposed to M when C just in case it would M in a suitable proportion of all C-cases (cases in which it undergoes circumstances of kind C). This move is largely motivated by their critique and rejection of the 'getting specific' strategy which, in contrast, takes a restrictive approach to analysing dispositions. I argue that, understood as a purely proportional approach, (PROP) is unable to deal with dispositional ascriptions made at extremely restrictive contexts and will be susceptible to cases of what I call 'proportionality swamping.' To avoid this problem, (PROP) must be read as taking a 'dual approach' to analysing dispositions: a synthesis of the proportional and restrictive approaches. I first flesh out a weighting system for C-cases gestured at by Manley and Wasserman and then show how incorporating this into their account allows us to easily amend their analysis so as to make it genuinely both proportional and restrictive. In the final sections of this paper I explore a potentially fatal problem for (PROP) and consider Manley and Wasserman's suggested approach to dealing with it. Their suggestion involves the generation of a slightly different version of (PROP) to the one I have been attacking and I go on to argue that even this alternate version is susceptible to proportionality swamping. I conclude that no matter which version of (PROP) we prefer, it is best interpreted as taking a dual approach to analysing dispositions.

What role does context play in the analysis of dispositional ascriptions?

It is relatively old news that context has some part to play in settling the meanings of dispositional terms—one and the same object might truthfully be fragile
to a giant whilst simultaneously non-fragile to a human, taken in context of their individual strengths, just as that same object might be light for one but heavy for the other—however, just how context enters into dispositional analyses, and therefore what kind of contextualist account we end up with, remains a matter of some contention.

Manley and Wasserman’s recent work on linking dispositions and conditionals (2007 & 2008) has brought two contrasting contextualist approaches to analysing dispositions into sharp focus: what I shall call the proportional approach and the restrictive approach. Briefly, a proportional approach holds that an object is disposed to $M$ when $C$ just in case it would $M$ in some contextually determined proportion of cases in which it is subjected to circumstances of kind $C$; a restrictive approach, on the other hand, holds than an object is disposed to $M$ when $C$ just in case it would $M$ in every one of a contextually restricted set of cases in which it is subjected to circumstances of kind $C$.

In this paper I consider Manley and Wasserman’s recently proposed analysis, (PROP), and argue that if it is understood as a purely proportional approach it turns out to be unworkable; in fact, I shall demonstrate, in order to generate the strongest reading of (PROP) we must construe it as taking a dual approach to analysing dispositions: one that incorporates both a proportional and a restrictive element.

I take my starting point with Manley and Wasserman’s rejection of purely restrictive approaches before putting a pair of novel problem cases to a reading of (PROP) on which it is takes a purely proportional approach. These cases, I shall argue, demonstrate the need for a restrictive element to (PROP) and I go on to show how a weighting system that Manley and Wasserman suggest for individual cases might be fleshed out to allow for this feature in their analysis.
In the final two sections of this paper, I explore a potentially fatal problem for (PROP), concerned with how we might go about comparing infinite numbers of $C$-cases, along with Manley and Wasserman’s response to this worry, which involves the development of a slightly different version of (PROP) to that which I discuss in sections 1-3. I shall go on to demonstrate that this alternate version too, if indeed possible, requires a restrictive as well as a proportional element and therefore conclude that (PROP) is best interpreted as a dual approach analysis.

1. REJECTING THE PURELY RESTRICTIVE
Manley and Wasserman (2008) have recently proposed the following analysis of dispositions:

$(PROP) \quad N \text{ is disposed to } M \text{ when } C \text{ if and only if } N \text{ would } M \text{ in some suitable proportion of } C\text{-cases } (2008: 76)$

A $C$-case, they explain, can be understood as “a fully specific scenario that settles everything causally relevant to the manifestation of [a] disposition.” (Manley and Wasserman 2007: 72). So dropping a glass, say, on Earth from one metre up onto a solid surface with a Shore durometer measurement of 90A, through a substance with a density of 1.2 kg/m$^3$ constitutes a particular $C$-case associated with that glass's disposition to break when dropped---with any change in those parameters generating a distinct $C$-case.

A glass is fragile, according to (PROP), just in case it would break in a suitable proportion of all dropping cases,$^{64}$ where just what counts as a `suitable' proportion of cases will depend “not only on the dispositional predicate involved but also on the

---

$^{64}$ Assuming here that the conventional disposition ‘is fragile' ought properly be translated into overtly dispositional locution as ‘is disposed to break when dropped.’
context of utterance” (2008: 76). In this way, they claim, “we can think of the [context of utterance] as providing the standards for ‘fragility’ by establishing a requisite proportion of $C$-cases in which [the glass] would break (2008: 76). Call this a proportional approach to analysing dispositions.

(PROP) is largely motivated by Manley and Wasserman's critique of what they call the `getting specific' strategy, which is gestured at by Lewis (1997) as a possible response for proponents of the simple conditional analysis to the problem of dispositional maskers. Briefly, the simple conditional analysis is typically given as:

\[
\text{SCA} \quad N \text{ is disposed to } M \text{ when } C \text{ iff } N \text{ would } M \text{ if } C.
\]

As is well known, this analysis fails for cases in which objects have certain of their dispositions masked by the presence of external factors. Consider, for example, Johnston's (1992: 233) example of a glass protected by internal packaging: the glass, being perfectly ordinary, is fragile and disposed to break when dropped; however, the presence of internal protective packaging entails that if the glass were to be dropped, it would not break---and would therefore lack the disposition to break when dropped according to SCA.\(^{65}\) Intuitively, though, the packaging does not remove the glass's disposition to break when dropped: it merely masks that disposition.

The getting specific strategy offers the proponent of SCA a possible response to this problem: Of course a glass is not disposed to break when dropped whilst containing packaging specially designed to prevent it from breaking when dropped, but that was hardly what we had in mind when we described it as disposed to break when dropped. When we make a dispositional ascription, according to the getting specific strategy, we really mean to attribute a highly specific disposition to an object

\(^{65}\) See also Bird's (1998) 'antidote' cases for a similar problem.
with our more general dispositional language. So, for instance, when we describe a
glass as disposed to break when dropped, we really mean to attribute to it a highly
specific disposition---to break when dropped from a certain height, onto a certain kind
of surface, through a certain medium, in the absence of packaging materials, and so
on---where the exact details of that disposition's stimulus conditions are determined
by the context of utterance.

Once the context of a dispositional ascription has been fully taken into account
we will arrive at a maximally specific set of stimulus conditions that ought to rule out
the presence of all dispositional maskers. In effect, by getting specific we narrow down
and restrict the set of C-cases at which we will be assessing a dispositional attribution
(perhaps even to just one C-case) and focus on only those cases that are contextually
relevant. In contrast to the proportional approach taken by (PROP), call this a
restrictive approach to analysing dispositions.

Whilst initially plausible, purely restrictive approaches like the getting specific
strategy quickly run into serious trouble: as Manley and Wasserman point out, a
disposition can be masked in precisely the kind of circumstances that we would expect
an object that is disposed in such-and-such a way to manifest that disposition. Consider
their example of a crystal glass with a strong spot, or what they call a reverse Achilles’
heel:  

If you tap this glass with your finger it will shatter. If you blow on it directly,
it will crack. But the glass has a reverse Achilles’ heel: it can withstand a surprisingly
strong force, provided that the force is applied at exactly the right angle and at exactly
the right point (2008: 465).

---

66 Or perhaps, ‘from a certain range of heights, onto a certain range of surfaces, through a certain
range of mediums... etc.’

67 See Manley and Wasserman (1998: 67–8) for their example of a straight-forward Achilles' heel.
I stick with the reverse case for continuity with my other examples.
Now, according to the getting specific strategy, in any given context `this crystal glass is disposed to break when dropped' will attribute to the glass a highly specific disposition—to break when dropped from a certain height, through certain medium, and so on. However, it is quite possible for the glass to be dropped in exactly those circumstances and land on its strong spot and therefore not break. Clearly the glass is disposition— to break when dropped from a certain height, through certain medium, and so on. However, it is quite possible for the glass to be dropped in exactly those circumstances and land on its strong spot and therefore not break. Clearly the glass is disposed to break when dropped in any normal sense of that phrase; however, because it would not do so in every case in which the specific stimulus conditions were met, on the current analysis it would come out as not being so disposed.68

Conversely, (PROP) deals easily with this kind of case, as well as the more commonplace place masking cases: because we need only concern ourselves with whether an object manifests a disposition in a suitably large proportion of all cases, rather than in every case in the restricted set of those relevant to the context of utterance, the few cases in which an object $O$ fails to manifest a disposition $D$ due to $D$'s being masked need not rob that object of its disposition. The masking cases represent just a small number of $C$-cases amongst many and, so long as $D$ is manifested in a suitable proportion of those other cases in which no masks are present, (PROP) will give the result that $O$ has $D$ (the same also applies to cases in which a disposition is finked69 or mimicked70).

Additionally, (PROP) allows us to account for gradability of dispositional properties, a feature not present in standard conditional accounts: just as it makes sense to say that some object $x$ is larger than some other object $y$, it makes sense to say that $x$ is more or less fragile than $y$. This is easily dealt with by a proportional account: we

68 If the crystal glass is not disposed to break when dropped due to the mere possibility, however small, of its strong spot preventing it from breaking, then it is not clear that anything will count as fragile: surely it is at least possible of any, if not most, fragile objects that they be dropped and land just right so as not to break.
69 See Martin (1994) and Lewis (1997).
70 See, for instance, Lewis's (1997) 'hater of Styrofoam' example.
can say that $x$ is more fragile than $y$, for instance, just in case $x$ manifests fragility in a
greater proportion of fragility relevant $C$-cases than $y$. This also allows us to talk of
the *degree* to which an object has a particular disposition in terms of the proportion of
$C$-cases in which it manifests that disposition.

A purely restrictive approach, then, should be rejected in favour of a more
proportional approach like (PROP) which not only deals with the Achilles’ heel type
cases that a purely restrictive approach stumbled upon, but also provides the extra
theoretical benefits of accounting for gradability and thereby comparability of
dispositions.

2. STEEL, SCHMEEL
To what extent, though, should we interpret (PROP) as a purely proportional
approach?

Manley and Wasserman are not explicit about whether they intend (PROP) to
involve any restriction over $C$-cases; in most of their discussions of (PROP) and its
application the suggestion seems to be that we look at the proportion of manifestations
across all $C$-cases, and in their (2011) response to Choi (which I discuss further at the
end of §3) they seem to suggest that a restrictive element in their account would be
unmotivated, given that it already deals with all of the typical problem cases facing
dispositional analyses. However, there is nothing that explicitly rules out there being
a restrictive element to (PROP), and in this section I shall argue that it is just this
feature that is needed to generate the most plausible reading of Manley and
Wasserman’s analysis.

First, let us establish a reading of (PROP) according to which it is purely
proportional in its approach:
(PROP-PURE) \( N \) is disposed to \( M \) when \( C \) iff \( N \) would \( M \) in some suitable proportion of all possible \( C \)-cases.

Now, whilst (PROP-PURE) will successfully deal with cases of finking, masking, mimicking and Achilles’ heels, it will struggle to account for cases that involve a dispositional attribution being made at an extremely narrow context and in this section I shall put a pair such cases to (PROP-PURE) to demonstrate this point.

Manley and Wasserman are, in fact, aware of something similar to this problem and whilst they gesture towards a weighting system for \( C \)-cases to overcome it, they do not say, in any detail, how this response should be realised. In this section I sketch a weighting system for \( C \)-cases along the lines Manley and Wasserman suggest and go on to argue that, even with such a weighting system incorporated into it, (PROP-PURE) cannot deal with the problem cases I raise. What we need, I shall argue, is a reading of (PROP) on which it takes a dual approach to analysing dispositions, and in the following section I demonstrate how the weighing system developed for (PROP-PURE) can be modified to allow a reading of (PROP) that is genuinely both proportional and restrictive in its approach to analysing dispositions.

First, consider the following case: a researcher at a facility cooled to absolute zero might truthfully describe a steel rod he is carrying as fragile---due to steel's tendency to become brittle at extremely low temperatures. However, only a minuscule proportion from the whole set of \( C \)-cases (in which the steel rod is subjected to the kind of stress relevant to fragility) involve the rod's breaking. (PROP) can easily accommodate this by stipulating that the context of the utterance---having been made in the super-cooled research facility environment---sets the standards for fragility in the facility very low by drastically lowering the proportion of \( C \)-case manifestations needed in order that an object be fragile.
Suppose now, though, that the facility also houses a number of rods made from 'schmeel,' an imaginary alloy with almost identical properties to steel save for that, where steel becomes fragile at extremely low temperatures, schmeel becomes fragile at extremely high temperatures; and let us further suppose that schmeel breaks in exactly the same proportion of fragility relevant \(C\)-cases. It would seem that in the research facility context—where the proportion of \(C\)-case manifestations required for fragility is extremely low—schmeel comes out fragile according to (PROP), and \textit{as fragile as steel no less} (Call this case \textit{steel, schmeel 1}).

This looks obviously wrong: the steel rods in the facility will almost certainly break when subjected to stress, whilst the schmeel rods are practically indestructible (within the facility's super cooled environment) and, intuitively, what we want is for the steel to come out fragile and schmeel to come out as robust. What we are interested in when assessing the researcher's utterance, this seems to suggest, is whether or not the rods will break when subjected to stress \textit{in the research facility} rather than in some drastically different scenario.

Manley and Wasserman anticipate two similar cases to \textit{steel, schmeel 1} which prompt them to suggest the introduction of a weighting system for \(C\)-cases. The first case concerns a possible problem for extrinsic dispositions:

Whether a castle is vulnerable (or disposed to be conquered) depends on its environment, including the lay of the land and the position of the enemy. For example, two castles that are intrinsic duplicates might differ in whether they are vulnerable because one is on the front lines, while the other is deep in friendly territory... if we treat all \(C\)-cases equally, we will find that the two castles just mentioned will count as equally vulnerable. This is because whether they would be conquered
will be evaluated at all the same C-cases... Clearly, it matters most whether the castles are conquered in nearby C-cases... in general, for extrinsic dispositions, the closer the C-case, the more it matters.\(^7^{1}\)

(2008: 78)

Here, then, we can see that when assessing an objects dispositions, we need to weight the C-cases in accordance with their closeness to the actual world, with the closest C-cases receiving the highest weighting (for extrinsic dispositions, in any case).

The second case they anticipate concerns a possible worry that might arise out of the context-dependent nature of our dispositional utterances:

Take two objects. Suppose that humans, with our paltry strength, are incapable of breaking the first, but that it would take little or no effort for a giant to break it. The second object, however, is strange—it will break under very light pressure, but it is impervious to the destructive efforts of giants who are incapable of applying light pressure. Now, it may well be that, given a suitable notion of proportion, these two objects break in an equal proportion of cases involving the application of stress. The trouble is that we are likely to call the second object ‘fragile’ but not the first, while the giants will have opposite inclinations. Clearly we are not only concerned with the proportion of stress-inducing situations in which an object would break, because some of those situations matter more to us than others. And these situations are different from the ones that matter to giants. (2008: 78-9)

\(^7^{1}\) This example originates in McKitrick (2003).
In this case, $C$-cases need to be weighted by their relevance to the context of utterance: clearly what kinds of cases matter to humans, and therefore which cases turn out to be contextually relevant, is dependent upon which cases humans have access to. Accordingly, Manley and Wasserman submit that “for humans, $C$-cases involving low-stress situations are weighted more heavily; for giants, the reverse holds true” (2008: 79). Importantly, Manley and Wasserman add that they “do not want to say that our own context ignores the high-stress situations: they can matter to a thing’s degree of fragility, but they just matter less for us than do low-stress situations” (2008: 79).

So whilst $C$-cases receive a weighting in accordance with their contextual relevance, we are still assessing dispositions at every $C$-case; none are ruled out by context outright.

Taking these two weightings into account, an object’s being fragile, say, is no longer a case of its simply manifesting fragility in a suitable number of $C$-cases: some $C$-case manifestations matter more than others and merely counting cases will not bring this out. Instead, the degree to which an object has a disposition can be represented by a fraction corresponding to the sum of all of the weighted manifestation $C$-cases over the sum of the total set of weighted $C$-cases. Call this dispositional strength.

Letting $w_\phi$ stand for the weighting of $\phi$ (where $\phi$ is a $C$-case), $c_\phi$ for the closeness of $\phi$ to the actual world and $r_\phi$ for the contextual relevance of $\phi$ to the current circumstances, we can generate the following equation for working out the weighting of a $C$-case:

$72$ I take no stance here on whether ‘fragility’ is an extrinsic or intrinsic disposition. Manley and Wasserman suggest that the closeness weighting is for extrinsic dispositions only; however, it is not clear, to me at least, that the closeness of the case in which an object manifests a disposition might not matter for intrinsic dispositions also. If fragility is intrinsic, and there is some problem involved with weighting intrinsic dispositions in terms of closeness, then it is easy enough to drop the closeness weighting from the analysis when intrinsic dispositions are involved.
\[ w_\phi = c_\phi r_\phi \]

where \( 0 < c \leq 1 \) and \( 0 < r \leq 1 \)

A closeness value of 1 here would signify that the C-case in question occurs in the actual world, with that value decreasing as we move further away into more distant worlds. Similarly, a relevancy value of 1 would signify that the C-case in question was fully relevant to the conversational context at hand, with that value decreasing as we move further away from contextually-relevant scenarios. Note that there are no 0 weightings for cases: a 0 value for closeness would mean that the case did not occur at any possible world, and allowing C-cases to have a 0 relevancy value would mean we were no longer assessing dispositions at all.

Once we know the weight of individual C-cases we can calculate the strength of a disposition using the following formula: Let DS stand for a disposition's strength, \( a_i \) for any C-case (where that case could involve either manifestation or non-manifestation of the relevant disposition) and \( m_j \) for a manifestation C-case.

\[ DS = \frac{(w_{m_1} + w_{m_2} + w_{m_3} + \ldots + w_{m_n})}{(w_{a_1} + w_{a_2} + w_{a_3} + \ldots + w_{a_k})} \]

where \( \{m_1, m_2, m_3, \ldots, m_n\} \subseteq \{a_1, a_2, a_3, \ldots, a_k\} \) \( n \leq k \)

We now can now generate two readings of (PROP-PURE) that incorporate this weighting system:

---

73 Where \( c \) is greater than 0 and less than or equal to 1 and \( r \) is greater than 0 and less than or equal to 1.

74 According to this system, the maximum weighting for a C-case is 1, with the minimum weighting approaching 0. This is to reflect the original intuition that we might had that we can simply add up all of the C-cases and see what proportion of those cases resulted in manifestation of the relevant disposition. I make no claims about whether this is the only or best way to calculate a C-case's weighting, this merely seems the simplest and most intuitive method.

75 Where the set containing \( m_1, m_2, m_3, \ldots, m_n \) is contained within or equal to (is a subset of) the set containing \( a_1, a_2, a_3, \ldots, a_k \), such that \( n \) is less than or equal to \( k \).
(PROP-PURE1)  \( N \) is disposed to \( M \) when \( C \) if and only if \( N \) would \( M \) in some contextually set proportion of \( C \)-cases, where the above function determines \( N \)'s dispositional strength for \( M \)-ing when \( C \); or

(PROP-PURE2)  \( N \) is disposed to \( M \) when \( C \) if and only if \( N \) meets some contextually set dispositional strength for \( M \)-ing when \( C \), where the above function determines \( N \)'s dispositional strength for \( M \)-ing when \( C \).

Put simply, according to (PROP-PURE1) we count up cases to see whether an object has a disposition and then calculate that disposition's strength; whereas according to (PROP-PURE2) we calculate an object's dispositional strength for a given disposition and then use this to determine whether that object has the disposition in question.\(^{76}\)

The first option is closer to the letter of (PROP-PURE), and so to Manley and Wasserman’s own proposal, but has the awkward consequence that an object \( A \) which lacks a disposition \( D \) may have a higher dispositional strength for \( D \) than another object \( B \) which actually has that disposition: for instance, suppose that \( A \) manifests fragility in a few extremely close \( C \)-cases whilst \( B \) manifests fragility in a great number of extremely distant \( C \)-cases, such that \( \text{DS}_{(A, \text{fragility})} > \text{DS}_{(B, \text{fragility})} \);\(^ {77}\) suppose also that \( A \) does not manifest fragility in a high enough proportion of \( C \)-cases to be fragile, according to (PROP-PURE1), whilst \( B \) does. In that case, whilst \( A \) is more

---

\(^{76}\) It may sound odd to talk of the strength of \( N \)'s disposition to \( M \) when \( C \), in cases where \( N \) may lack the disposition to \( M \) when \( C \) (if it fails to meet the standards for having that disposition in the context in question). This need not be too troubling once we realise that we are talking about the degree to which an object has a particular disposition: a concrete block, for instance, can clearly be fragile to a particular degree (generally a very low one) without actually being fragile at a given context.

\(^{77}\) The dispositional strength of \( A \)'s fragility is greater than the potential dispositional strength of \( B \)'s fragility.
fragile than $B$ (having a higher DS for fragility), $B$ is fragile whilst $A$ is not. For this reason I think (PROP-PURE1) turns out to be unworkable and for the remainder of the paper I'll understand (PROP-PURE) as working in the manner outlined by (PROP-PURE2).

With this weighting system in place, the vulnerable castles and giants cases cease to pose any real problem for (PROP-PURE): although the castles are conquered in exactly the same number of $C$-cases, the castle on the front lines (call this castle $A$) is conquered in much closer $C$-cases than the castle deep in friendly territory (call this castle $B$) and so at the very least castle $A$ is more vulnerable than castle $B$; though whether either of the two castles actually turns out to be vulnerable will depend on the contextually set DS threshold for vulnerability.

In the Giants case, $C$-cases will be weighted according to the context of utterance; so, in the context of being a human, $C$-cases in which the second object breaks when struck will be weighted a great deal higher than $C$-cases in which the first object breaks when struck. That being the case, for humans, the second object will turn out to be fragile whilst the first object turns out to be non-fragile (and *vice versa* for giants).

Perhaps the above strategy can be applied to my *steel, schmeel* case: the cases in which the steel breaks when dropped by the researcher are closer and more contextually relevant than the cases in which the schmeel breaks when dropped, so the steel rods have a much higher fragility DS than the schmeel rods. We could therefore hold that either:

1. Steel and schmeel both meet the contextually set DS threshold for fragility, so both are fragile but steel is more fragile than schmeel; or
Steel, but not schmeel, meets the context set DS threshold for fragility, so steel is fragile whilst schmeel is non-fragile and again steel is more fragile than steel.

The second option seems to be closer to our intuitions about steel and schmeel, but in either case we get the result that steel is more fragile than schmeel.

Unfortunately this solution turns out to be short-lived: suppose we tweak the example slightly so that schmeel will break when dropped (from a suitable distance, finking and masking cases aside) at any temperature other than absolute zero (call this steel, schmeel 2). Now, because (PROP-PURE) has us assess dispositions at every C-case, schmeel will have a higher fragility DS than steel due to the potentially infinite number of fragility C-cases in which it manifests fragility far outweighing the more relevant cases in which it does not; call this a case of proportionality swamping. Given that it remains almost indestructible within the research facility, however, I maintain that within this context there is no sense in calling it more fragile than steel or, in fact, fragile at all! To see this more clearly consider the following case:

Suppose that scientists discover that wood will only burn when in oxygen due to the presence of a certain subatomic particle X that is pervasive in our solar system. Suppose further that it is discovered that particle X is not found anywhere else in the universe, and hence that wood would not burn when heated in oxygen anywhere other than within our system.

Would this discovery cause us to stop thinking of wood as disposed to burn when heated in oxygen? It seems highly doubtful. To echo Choi (2008: 24-6) on similar cases, we would still need to distinguish between objects that would burn when heated in air from those that would not in our ordinary circumstances, and given our
ordinary circumstances involve the pervasiveness of $X$, wood must obviously be placed into the first category.

Similarly, in what becomes the ordinary circumstances for the researchers in the research facility steel must obviously be placed into the category of things which are fragile, whilst schmeel into the category of things which are not.\textsuperscript{78} Granted the environment in the $X$ particle example is a great deal larger and shared by many more people, but that doesn't appear to be a principled reason for treating the contexts generated by these two environments any differently.

In fact, the $X$ particle example hints at the answer to the problem that steel, schmeel poses for (PROP-PURE): not all stimulus condition cases are relevant when assessing the dispositional properties of an object: some are less relevant than others, as per the suggested weighting system, but some are simply not relevant at all. In the $X$ particle case, the discovery that $X$ does not exist outside of our solar system (and hence that at a near infinite number of cases in which we heated wood in oxygen it would not burn) does not affect our thinking of wood as being disposed to burn in air because stimulus cases at which $X$ is not present never played a part in our assessment of wood's disposition to burn in the first place. Similarly, our intuitions about the fragility of steel and schmeel within the research facility, I submit, are based upon our ruling out certain $C$-cases as irrelevant to the context at hand.

(PROP-PURE) falls down, then, because it is committed to assessing dispositions at every $C$-case: No matter how we weight $C$-cases,\textsuperscript{79} we will always be able to make the set of contextually relevant $C$-cases arbitrarily small and the set of

\textsuperscript{78} Note that it need not be the being in the research facility environment that is important here. Rather, we might think of the intentions and aims of the speaker as generating a particular context for their utterances which then determines which $C$-cases turn out to be relevant to assessing the truth and falsity of those utterances.

\textsuperscript{79} Where $C$-cases can only be assigned non-zero weightings.
contextually non-relevant $C$-cases arbitrarily large to yield the kind of proportionality swamping illustrated by *steel, schmeel*. We therefore need to modify (PROP) so that contextually irrelevant cases can be excluded from our assessment of an object's dispositions at that context completely.

3. A DUAL APPROACH TO ANALYSING DISPOSITIONS

Ironically, it is a restrictive aspect of the kind that Manley and Wasserman began by criticising that is needed by (PROP) in order to deal with *steel, schmeel* type cases. Helpfully, the formulas for calculating dispositional strength and $C$-case weighting that I discuss above provide us with a simple way of accomplishing this: if we allow the relevancy weighting on $C$-cases to take a 0 value when a case is contextually irrelevant, then, due to the way that $C$-case weighting is derived by multiplying a cases’s closeness by its contextual relevance, that $C$-case would itself receive a weighting of 0, effectively excluding it from the set of cases at which we assess an object's dispositions within that context.

With this development, along with the incorporation of the weighting system for $C$-cases and move to talking in terms of dispositional strength rather than proportions of cases, we are in need of a restatement of (PROP). I suggest the following:

(PROP) \(N\) has a disposition $D$ to $M$ when $C$ iff $N$ meets a suitable dispositional strength for $D$ across a range of contextually relevant $C$-cases.

As before, what counts as a suitable dispositional strength will be determined by the conversational context, with that context additionally determining the $C$-cases at which the dispositional ascription is to be assessed. The move to talking in terms of meeting a certain dispositional strength (brought by the incorporation of the weighting
system) along with the restricting of $C$-cases to those that are contextually relevant is a significant change to the wording of the original account, though one that, I submit, is compatible with the original statement of the analysis. It also constitutes a far stronger reading of (PROP) than might initially have been taken: instead of having a solely proportional account, we now have an account that is genuinely both proportional\footnote{Note that although the account is no longer stated in terms of proportions of cases, but instead in terms of dispositional strength, it is still a proportional approach in that dispositional strength is determined in large part by the proportion of manifestation cases that an object displays.} and restrictive in its approach and that ought to make good on the original intuitions driving both (PROP) and the getting specific strategy.

With this reading, we are now equipped to deal with both versions of steel, schmeel: In the first case the small amount of cases in which schmeel breaks when dropped will most likely be ruled out entirely in the facility context, meaning that only steel will come out fragile. In the second case, there may be some relevant $C$-cases in which schmeel breaks when dropped (given that we might think that at least some non-research-facility temperatures very close to absolute zero are relevant to whether an object at the facility is fragile), but due to their weighting, and the fact that their numbers are limited by context, steel will come out more fragile than schmeel and schmeel will fail to be fragile at all.

Read as dual approach account, (PROP) retains its ability to avoid the numerous counterexamples to earlier conditional approaches: whilst some cases may now be ruled out by context alone, the cases of finking, masking and Achilles’ heels that remain contextually relevant will constitute just a small proportion of the overall cases at which a disposition is assessed.

It is worth noting that a similar account is sketched by Choi (2011b) in his response to Manley and Wasserman’s critique of the getting specific strategy.
According to Choi, it is not context that rules out the presence of dispositional masks and finks (etc.), but the context independent nature of dispositionality itself (Choi 2011b: 1169; see also Choi 2011a). As such, “it is an option open to the proponents of (SCA) to develop much the same proposal [as Manley and Wasserman] with respect to the semantic elements [of dispositional ascriptions] saturated by the context of ascription.” (Choi 2011b: 1170). In effect, this would result in an approach that was both proportional and restrictive (though the restrictive aspect would be dealt with by the nature of dispositionality rather than contextual factors). Manley and Wasserman respond that this approach is insufficiently motivated given that (PROP) already deals with finks, masks, Achilles’ heels, and so on (Manley and Wasserman 2011: 1215). 

_steel, schmeel_ type cases, I submit, are the motivation for taking a reading of (PROP) on which it takes a dual approach that is both proportional and restrictive.

Suppose that someone responds to this reading: ‘but schmeel really is more fragile than steel in _steel, schmeel_ 2, given that its fragility (as given by the formula above) is higher than that of steel across all C-cases, it is just that it would flout Gricean maxims of conversation to describe it as such.’ That is, whilst it is pragmatic for the researcher to describe the steel as more fragile than the schmeel, this has no effect on actual dispositional properties of the two materials. So, when the researcher says that steel is more fragile than schmeel what he says is useful but literally false.

A simple reply to this is that it just looks _ad hoc_ considering that, by adopting (PROP), we have already endorsed contextualist truth conditions for dispositional ascriptions. It seems unprincipled, that is, to reject (PROP) in favour of (PROP-PURE) in light of _steel, schmeel_ 1 and 2. After all, we can still maintain that schmeel is more

---

81 Importantly, we could not rule out intrinsic mimics and masks such as Achilles’ heels and reverse Achilles’ heels in the same manner: we cannot assess the fragility of a glass with a strong spot in the absence of that glass’s strong spot; to do so would simply be to assess an different glass (one without a strong spot).
fragile than steel in certain contexts: Suppose that after hours in the facility's bar the researcher says to his intern “funny thing, schmeel is actually way more fragile than steel; I've handled it before I got here and it broke at the slightest touch.” According to this dual approach reading of (PROP) the researcher's utterance in the bar and previous utterance in the facility are both true because his interests and intentions are different in each case, thereby altering the context of his assertions: in this later utterance a much wider range of C-cases is taken as relevant to assessing the fragility of schmeel.

To be clear, In this section I have sought to give a reading of (PROP) that is both proportional and restrictive in its approach and have argued that such a ‘dual approach’ is necessary given the various problem cases discussed above. This is not intended as a replacement for, or a competing theory to, Manley and Wasserman’s thesis, but rather as a suggestion for the strongest reading of (PROP) given the wording of their analysis.

4. COUNTING CASES
The intention in the previous two sections has been to demonstrate that (PROP) requires a restrictive element in order to avoid the kind of swampi ng effect that occurs in Steel, Scheel 2; however, it might be argued that my more detailed analysis of (PROP) and (PROP-PURE) mischaracterises the way that these analyses are supposed to operate, and therefore that my Steel, Schmeel cases miss their mark.

More specifically, I have characterised (PROP) as involving the ‘counting up’ of the numbers of manifestation and non-manifestation C-cases in any given scenario and then the feeding of those numbers into a formula designed to calculate dispositional strength. That is, I have stated (PROP) as involving the comparison of
sets of $C$-cases in terms of their cardinality, or the numbers of their elements; call this the *counting* version of (PROP).

In fact, interpreting (PROP) in this way opens the analysis up to a rather devastating problem: given that we are dealing with potentially infinite numbers of $C$-cases, it is not obvious that we can actually compare the sizes of two sets of $C$-cases in terms of their cardinality as both will contain an infinite number of elements.

Manley and Wasserman acknowledge this problem and suggest that rather than trying to compare sets of $C$-cases in terms of their cardinality, we should make use of some alternate measure of set size to calculate our proportions of manifestation (2008: 79-82). Call any such version of (PROP) a *measured* version of the analysis.

Now, whilst it might be conceded that my *steel, schmeel* cases are effective in demonstrating that the counting version of (PROP) needs to take a dual approach in its analysis of dispositions, this version may turn out to be unworkable, in any case, and abandoned in favour of a measured version that might account for *steel schmeel* cases where the counting version could not. More work needs to be done, therefore, to demonstrate that (PROP) needs to be interpreted as taking a dual approach to analysing dispositions.

In this section I shall outline the problem with comparing sets of $C$-cases by appealing to their cardinality, and argue that without a method for restricting $C$-cases to a finite number, the counting version of (PROP) appears to be unworkable. In the following section I shall address Manley and Wasserman’s proposal that we appeal to some alternate measure of set size and argue that, irrespective of the measure we use to compare sets of $C$-cases, (PROP) will still need both a proportional and a restrictive element.
First let us consider the kind of problem that (PROP) faces: suppose that a glass made from element Z will break in every fragility relevant C-case that is situated within the Earth’s atmosphere but nowhere else in the universe. In what proportion of C-cases will the Z-glass manifest fragility? The formula in §2 above suggests that we represent that proportion as a fraction equivalent to the sum of the manifestation C-cases divided by the sum of every fragility relevant C-case (appropriately weighted by relevance and closeness where appropriate). But there are are infinite number of possible fragility cases for us to consider, so the denominator of that fraction will contain an infinite string of weighted C-case additions, making its value tend towards infinity.

In a little more detail, we can re-write the formula for calculating dispositional strength as follows:

\[ DS = \frac{x}{y} \quad | \quad x = \sum_{i=1}^{n} w_{m_i} \text{ and } y = \sum_{i=1}^{n} w_{a_i} \]

For finite values of \( n \) this would generate a well-formed fraction expressing a determinate proportion of C-cases. However, as we are dealing with an infinite number of C-cases, \( y \) will equal the sum of an infinite set of weightings \( (y = \sum_{i=1}^{\infty} w_{a_i}) \) and, as such, will have no determinate value but rather a limit of \( \infty \). This, in turn, will give \( \frac{x}{\infty} \) a limit of \( 0 \) \( \left( \lim_{y \to \infty} \left( \frac{x}{y} \right) = 0 \right) \). So, in our DS equation, as the value of \( y \) approaches \( \infty \), the value of \( \frac{x}{y} \) will tend towards 0, making the dispositional strength effectively zero.\(^8\)

\(^82\) Read: \( DS = \frac{x}{y} \) where \( x \) = the sum of all values of \( w_{m_i} \) for all values of \( i \) between 1 and \( n \), and \( y \) = the sum of all values of \( w_{a_i} \) for all values of \( i \) between 1 and \( n \).

\(^83\) To see this consider what happens as we repeatedly increase the denominator of \( \frac{1}{10} \) by a factor of 10:
According to the DS formula above, then, the proportion of $C$-cases in which, say, a $Z$-glass breaks when struck will be infinitesimally small and, to make matters worse, this will be the case for every possible value that $x$ can take. There will be little sense, therefore, in saying that a glass made from element $Z$ is more fragile than a glass made from element $Y$, say, in that the formula will appear to attribute to them the same dispositional strength for fragility (arbitrarily close to 0).

We might think that this problem can be avoided via the contextual restriction of $C$-cases that I suggest in the previous sections: If dropping cases are restricted to those the average human is likely to encounter, we might think we cut down the infinite number of conceptually possible cases to a finite subset that we can then feed in to our DS formula. However, taking just one variable and limiting the range of cases that we will look at along that variable’s dimension, we will still have an infinite number of cases to deal with. Consider the height from which an object is dropped:

Suppose we say that in a certain conversational context $K$, fragility $C$-cases are limited to all and only those in which an object is dropped from 1 meter onto the ground, and suppose further that we take dropping an object from exactly 1 meter as the first $C$-case in $K$, $a_1$, with subsequent cases involving dropping an object

\[
\frac{1}{10} = 0.1 \\
\frac{1}{100} = 0.01 \\
\frac{1}{1000} = 0.001 \\
\frac{1}{10000} = 0.0001 \\
\ldots
\]

As the denominator grows larger, the fraction's value moves closer to zero. As the denominator approaches infinity, therefore, the fraction's value will approach 0.
increasingly closer to the ground. Our problem is that for any point \( p_0 \) that we pick between our 1 meter limit, \( l \), and the ground, there will be a mid-point \( p_1 \) between \( p_0 \) and \( l \), and a further mid-point \( p_2 \) between \( p_1 \) and \( l \), and a further midpoint \( p_3 \) between \( p_2 \) and \( l \), and so on, so we will never arrive at a dropping height for our second case, \( a_2 \).

For instance, if we take our limit \( l \) and 99cm above the ground as our two initial points, we can find their midpoint at 99.5cm, and then a second midpoint between \( l \) and 99.5 at 99.75, and then at 99.875, 99.9375, 99.96875, 99.984375, and so on \textit{ad infinitum}. In fact, for any two points that we pick, there is an uncountably infinite number of points between them,\textsuperscript{84} and therefore an uncountably infinite number of \( C \)-cases to consider.

Perhaps even more worrying is that the same holds for the manifestation cases: Suppose that a glass will break if dropped in all cases (though not exclusively in these cases) in the following range: between a height of 99cm and 1 meter, onto a surface with a Shore durometer measurement of 89A to 90A, through a substance with a density of 1.19 kg/m\(^3\) to 1.2 kg/m\(^3\). Inside of those limits are an uncountably infinite number of dropping cases, and so the glass breaks in an uncountably infinite number of \( C \)-cases. That being the case, both \( x \) and \( y \) in the \( DS \) formula given above will tend towards infinity, in that their value will be derived from the summation of the weightings of an uncountably infinite number of \( C \)-cases:

\textsuperscript{84} A set is uncountably infinite if it has an infinite number of elements which cannot be put in a one-to-one relation with the set of natural numbers. In our case, because there are an infinite number of \( C \)-cases along spatial dimensions, for any two \( C \)-cases that we pick there will be a third \( C \)-case that ought to be placed in between them (if organising \( C \)-cases by their dropping height). As such, we cannot place our \( C \)-cases in a one-to-one relation with the set of natural numbers: if ever we paired two \( C \)-cases as 1 and 2 there would be an infinite number of \( C \)-cases in-between them that we would have skipped over.
With the formula set up in this way, it is unclear whether \( \frac{x}{y} \) can even be conceived of as expressing a proportion at all as the value of \( \frac{\infty}{\infty} \) is simply undefined.

We have the complication here that both \( x \) and \( y \) are the sums of uncountably infinite sets of weighted C-cases, so perhaps it might be suggested that \( x \) and \( y \) will turn out to represent infinites of differing cardinalities. That is, whilst the set of relevant C-cases and the subset of those cases that result in manifestation of the relevant disposition are both uncountably infinite, we might think that the subset is of a smaller cardinality than its superset.

Aleph numbers are used to represent the cardinality of such infinite sets, where \( \aleph_0 \) represents the cardinality of the set of all natural numbers then \( \aleph_1 \) for the next larger cardinality, then \( \aleph_2, \aleph_3, \aleph_4 \) and so on. We might suppose, then that if \( x = \aleph_h \) and \( y = \aleph_j \) such that \( h < j \), then \( \frac{x}{y} \) will still express the proportion of C-cases in which an object manifests a particular disposition.

This solution is, again, heavily flawed, however: First, it is far from obvious whether the set of all relevant C-cases and its subset of manifestation C-cases have a different cardinality: the two sets potentially have the same number of elements, unintuitive as that might be, just as the set of natural numbers and the set of even numbers both have a cardinality of \( \aleph_0 \); and second, whilst multiplication and addition of aleph numbers are definable operations (such that, for instance, \( \aleph_\alpha + \aleph_\beta \mid \alpha < \beta = \)

\[ DS = \frac{x}{y} \] where \( x = \sum_{i=1}^{\infty} w_{m_i} \) and \( y = \sum_{i=1}^{\infty} w_{a_i} \).  

---

85 Read: \( DS = \frac{x}{y} \) where \( x = \) the sum of all values of \( w_{m_i} \) for all values of \( i \) between 1 and \( \infty \), and \( y = \) the sum of all values of \( w_{a_i} \) for all values of \( i \) between 1 and \( \infty \).
\[\aleph_\beta \times \aleph_\alpha \mid \alpha < \beta = \aleph_\beta\], division of aleph numbers is an undefined operation, so having \(DS\) equal such an operation is nonsensical.

Unless some plausible account of infinite proportionality can be provided, and I cannot even begin to offer one here, a proportional analysis such as (PROP), understood in the way outlined above, cannot succeed without limiting \(C\)-cases to a finite number and, as I’ve indicated above, we cannot do so with contextual restriction alone.

We might try to restrict ourselves to a finite number of cases by making the cases that we look at themselves more coarse grained. So, rather than looking at maximally specific \(C\)-cases, suppose that we concentrate on cases characterised by more general descriptions of stimulus conditions, call them \(G\)-cases, that contain within them a range of \(C\)-cases. So, for instance, a particular fragility \(G\)-case (taking dropping height as the only variable for example’s sake) might be ‘dropping an object from \(h\) centimeters above the ground where \(100 \geq h > 99\)’, with the next distinct \(G\)-case being ‘dropping an object from \(h\) centimetres above the ground where \(99 \geq h > 98\).’ By continuing in this fashion we could generate 100 distinct dropping cases between 1 meter and the ground, none of which overlap and which jointly contain all of the \(C\)-cases over the same range of heights. Given that we are counting only the more coarse grained \(G\)-cases, we would not need to count the infinite numbers of \(C\)-cases that they contain in order to calculate the proportion of manifestations, we could simply count up the \(G\)-cases and feed them into out \(DS\) formula as normal.

There is a rather large complication with this strategy, though: not all of the \(C\)-cases contained within the more general \(G\)-cases will be of just one type (manifestation or non-manifestation), so we will not be able to appropriately identify and count up which \(G\)-cases involve manifestation and which do not. That is, unless an object
behaves in the same way in every C-case contained within the range of a particular G-case, we will be unable to say whether that G-case is a manifestation or non-manifestation case.

Remember that the problem for the getting specific strategy was that an object’s disposition can be masked even in the perfect circumstances for manifesting that disposition. A similar problem will arise for G-cases: an object dropped from between 99 and 100 centimetres might break on some occasions but not on others due to variations in the precise details of the case: the medium it falls through, the surface it lands on, the angle at which it hit, its landing on its strong spot, etc. Even restricted to a narrow range, these variables could plausibly make the difference between an object’s breaking when dropped and its remaining intact and it will therefore be unclear how we should go about categorising G-cases.

If we go with an all-or-nothing approach, such that a G-case is a manifestation case iff every C-case contained within its range is a manifestation case, then we will find that most objects are non-fragile: there are an uncountably infinite number of C-cases contained within each G-case, remember, and all it will take is one in which an object doesn’t break when dropped for the G-case itself to be categorised as a non-manifestation case.

We could try to significantly narrow the range of our G-cases so as to avoid this issue, so that for instance a single G-case might cover the range of dropping cases between 1,000,000,000 and 999,999,999 nanometres above the ground, but it is far from obvious that this will resolve the worry: there will be an infinite number of C-cases in any range that might fall under a G-case description so there is a danger that the same worry will crop up again; what’s more, restricting G-case ranges to whatever level rules out this particular problem seems rather ad hoc.
Alternatively we might try to say that a $G$-case counts as a manifestation case iff a suitable proportion of the $C$-cases within its range are manifestation cases; but of course this would take us back to the problem discussed above of trying to make sense of proportions of uncountably infinite numbers of $C$-cases.

The counting version of (PROP), then, appears to be unworkable: we either attempt to compare sets of $C$-cases in terms of their cardinality, which cannot work for the uncountably infinite cases we have to deal with, or we try to restrict ourselves to a finite number of more coarse grained cases and invite the problems faced by purely restrictive approaches back into the analysis.

5. MEASURING $C$-CASE-SPACE

In the previous section I identified a potentially fatal problem facing (PROP) and suggested that if the analysis is reliant upon comparing sets of $C$-cases in terms of their cardinality, then it may ultimately turn out to be unworkable. That is not so say that some suitable response could not be generated along the possible lines of response that I suggested, but with no solution particularly forthcoming this problem represents a serious threat to the plausibility of (PROP).

However, Manley and Wasserman’s discussion of this problem (2008: 79-82) suggests that this is really only an issue for what I have called the counting version of (PROP), according to which we compare $C$-case set sizes in terms of their cardinality, and that we can avoid it entirely by finding an alternative measure for $C$-case set size:

Sometimes there is a natural substitute for comparison of cardinality.

For example, take the intuition that on a real line from 1 to 100 metres, there are in some sense ‘fewer’ points between 1 and 2 metres than there are between 2 and 100 metres; and that a point selected at random
from the line is far more likely to be selected from the second interval.

These intuitions can be appeased, but not by comparing the cardinal number of points in each interval. Instead we can appeal to a natural measure on the set of points on the line, one that corresponds to the intuitive notion of length. This gives us a non-arbitrary standard by which to compare the ‘size’ of sets of points—at least, those that correspond to intervals on the line. (2008: 79-80).

Briefly, in mathematics, a measure on a set $A$ is a generalisation of the concepts of length, area and volume that allows us to systematically assign a number to every suitable subset of $A$ which can be interpreted as those subsets’ sizes.

So, allowing $X$ to be the set of all the points on a 10 metre line (of which there are an uncountable infinite number), we can decompose $X$ into ten equal subsets, $A$ to $J$, of the points from 0 to 1 metres, 1 to 2 metres, 2 to 3 metres, and so on (see fig.1). The measure on a set must be countably additive, so that the measure of a ‘larger’ set must equal the sum of all the measures of its ‘smaller’ subsets (providing

---

86 We can avoid any overlapping of sets by allowing each set to contain its lower boundary point but merely approach its upper boundary. So, for instance, $A$ which is bounded by 0 and 1 metres, would contain the point at 0 and all of the points up to but not including 1 metre.
that the larger set can be decomposed into a countable number of subsets); so the size, in terms of the measure of length, of $X$ will equal the sum of the sizes of its subsets $A$ to $J$, with each of those subsets being a tenth the size of $X$; in other words, the proportion of points in $X$ that are members of the subset $A$, for instance, will be $\frac{1}{10}$.

What we need, then, is an analogous measure, to that of length in the above example, for sets of $C$-cases; importantly, one that does not assign uncountable size to subsets of $C$-cases so that they can be compared and their relative proportions calculated.

As Manley and Wasserman indicate, however, whether there is an analogous measure for the sizes of $C$-case sets is not at all obvious:

The problem for (PROP) is that it is far from a trivial issue whether there is any analogous non-arbitrary measure on the set of $C$-cases that will allow us to justify our intuitions of proportion among its subsets. Because we must combine the several factors of height, density, etc., and there may be many equally good ways to do this, the worry is that we will end up with too many equally natural measures on the set of $C$-cases. And these competing measures would yield conflicting proportionality results. (2008: 80)

Assuming that there is some suitable measure for the sizes of $C$-case sets, I think that (PROP) will require the kind of contextual restriction that I have argued for in sections 1 to 3. In the remainder of this final section I shall sketch a potential measure for $C$-case sets and subsets and then demonstrate that, even if such a measure is possible, a measured version of (PROP) will still be at risk of allowing the kind of proportion swamping illustrated by steel, schmeel 2 if it is interpreted as taking a purely
proportional approach. I shall therefore conclude that, aside from the worries discussed in this and the previous section, to generate the strongest reading of (PROP) we must interpret it as taking a dual approach to analysing dispositions.

First, let’s consider how we might generate a potential measure for C-case sets: if we assume that every disposition has just one C-case type associated with it, where each C-case type has a set number of variables along which its tokens can vary, we can generate a case-space in which to map all of the potential C-cases associated with a particular disposition by allowing each variable of the C-case type to be represented as a (generalized) spatial dimension of that case-space. So, a C-case type with n variables would generate an nth dimensional case-space in which all of the possible C-cases of that type could be mapped.

Now, rather than counting up the elements of sets of C-cases, and of their subsets of manifestation C-cases, to compare them, we can calculate the nth dimensional volume (call this $n$-volume for short) of the manifestation regions of our case-space and compare that to the $n$-volume of the entire case-space. Our DS formula, then, will now involve dividing the $n$-volume of all the manifestation regions of a case space by the $n$-volume of the total case space.

With such a measure the fact that there are an infinite number of points from which to drop an object from 1 metre to the ground, say, is no longer an issue for calculating proportions of manifestation because we are only interested in the $n$-volumes of certain regions of measurable case-space, not the infinite number of C-cases that might populate those regions.

We need to represent every causally relevant variable of a C-case type as a dimension of our case-space, note, because otherwise we would end up with regions of case-space that we could not determinately identify as containing manifestation or
non-manifestation cases. Take the presence of finks as an example: If we fail to represent the presence of a dispositional fink as one of our dimensions, there will be no differentiating between cases in which a fink is present and those in which one is not, as such cases would be mapped to the same point in our case-space. Because the presence of a fink can mean the difference between a C-case resulting in manifestation and its not, it would therefore be indeterminate whether those points were manifestation cases or non-manifestation cases, in turn making it indeterminate whether the regions they populate are manifestation regions. If we are unable to identify which regions of case-space involve manifestation and which do not, we cannot calculate the proportion of cases in which an object would manifest a given disposition.

Of course, C-cases also need to be weighted in accordance with their closeness (for extrinsic dispositions) and contextual relevance. This can be achieved by representing relevance and closeness as extra variables in our nth dimensional case-space. To see how this would work consider the following example:

Suppose that we play a simple game with a 10cm piece of string in which we each roll dice in order to determine, by some rule or other, who owns which parts of the string. The smallest segment one can own is 1cm, with all larger segments being multiples of that smallest amount. The game is over when all of the string is owned and we then count up how much string we each own to determine the winner. The value of the string is effectively one dimensional, in that it is directly equivalent to the length of any given segment. So if at the end of the game I own the segments from 0 to 4 cm and you own the segments from 4 to 10cm, you will win with a score of 6 to 4 (see fig.2).
Now suppose that we make the game a little more interesting by weighting the string so that the segments closer to one end are worth more than those closest to the other. We might do this by stipulating that the string at 0cm is worth 10, and the string at 10cm is worth 0, where the value of the string declines in a linear fashion between these two end points. Now in order to calculate a player’s score we need to work out the area under the line from (0,10) to (10,0) that are theirs, in accordance with which segments of the string they own.

The value of the string is now two dimensional, in that it is determined both by the length of the segment owned and the weighting of that particular segment; so, supposing that, again, the game ends with me owning the segment from 0 to 4cm and you owning the segment from 4 to 10cm. With the new weighting system, I will win the game with a score of 32 to 18 (see fig.3).

Essentially, we want our weighting system for C-cases to operate in much the same way as the weighting on string segments in our game above. By including an extra dimension in our nth dimensional case-space for contextual relevancy (and closeness for extrinsic dispositions), we can weight particular regions of the case-space and make them matter more in our dispositional analysis than others. For instance, in figure 4, letting $x$ and $y$ stand for any two variables of a 2-dimensional...
case-space and \( w \) for the contextual relevance of the cases within that space on a scale of 0 to 1, we can generate a 3 dimensional case-space in which certain regions, in this case towards the centre, are weighted more heavily than others: the regions of the case-space that are more contextually relevant end up with much larger \( n \)-volumes than contextually irrelevant regions and so will count more in our calculations of dispositional strength, effectively making some kinds of \( C \)-case matter more to an object’s having a certain disposition than others.

Before proceeding any further it is worth noting a few worries we might have about this sketch, and any potential measures of \( C \)-case set size:

First, we have assumed that every disposition has just one \( C \)-case type associated with it, but it is far from obvious that this will always, or ever, be the case; fragility, for instance, appears to have at least two characteristic stimulus condition cases: dropping and striking. The two kinds both involve the putatively fragile object being subjected to some kind of impact, but they have substantially different variables: dropping height vs speed of strike, hardness of surface vs mass of striking implement, and so on. Now, whether the different kinds of \( C \)-case types that might be associated
with a disposition can be reconciled into just one C-case type remains to be seen, but it is certainly not clear how we would go about plotting these different kinds of case within the same case-space, and less obvious still how we would go about combining the results from two different case-spaces for the same disposition.

Second, C-case sets may turn out to be non-measurable; that is, their structure might be so complex that no meaningful measure can be applied to them. For instance, as noted above, in order to construct our case-space for a C-case type, we would need to compare variables such as height, density, hardness and so on, but it is not clear that all of these variables will be commensurable; take the variables of ‘dropping height’ and ‘ambient temperature’ in fragility C-cases for two objects, Manley and Wasserman note:
To compare the ‘size’ of the sets of C-cases in which these objects would break, we need a way to compare the size of height intervals to the size of temperature intervals. For example, how does a range of 10 degrees Celsius compare to a range of 1 meter? Is there a non-arbitrary standard that decides the matter in a fashion that would match our intuitions of proportions of C-cases? (2008: 80)

Perhaps more worrying is how we would go about comparing dropping height, say, to the presence of dispositional finks or masks? How would we, for instance, interpret a fink’s presence as a dimension in our case space? Is it as simple as a 0 to 1 scale of not present or present, or do we need a scale that covers the different kinds of finks and their position in relation to the putatively disposed object? Alternatively, do we need a different dimension for every conceivable kind of fink, mask and mimic that might causally affect the relevant objects manifesting a given disposition? If C-case sets are simply non-measurable then the measured version of (PROP) is a non-starter.

Finally, the creation of an nth dimensional case-space from the variables of a C-case type is reliant upon the assumption that each C-case type has a finite number of variables. If there are an infinite or uncountable number of variables to a given C-case type then we will simply be unable to construct a case-space in which to plot all of the C-cases for the relevant disposition. Take the presence of finks as an example once again: if the kind of fink is relevant to a C-case, not merely the presence of some fink or other, and there are an infinite number of variations of finking cases, we would need a case-space with an infinite number of dimensions which simply could not be modelled.

These worries are substantial, and should make us seriously consider the plausibility of a measured version of (PROP); however, if such a version is available
along the lines I have sketched above, I submit that it will still need to take a dual approach to analysing dispositions if it is to be successful.

Manley and Wasserman motivate the idea of using measure theory to assign sizes to the sets of $C$-cases (PROP) has to deal with by drawing an analogy between the uncountably infinite number of $C$-cases in any given set and the uncountably infinite number of points on a real line 100 metres in length. The intuition, remember, was that, although we cannot make a comparison in terms of their cardinality, the set of points between 0 and 1 metre on the line is smaller than the set of points between 1 and 100 metres on the line according to the natural measure of length.

This analogy breaks down, however, when we consider that many of the variables of a $C$-case type will translate into an infinite scale in our case-space. Take dropping height as an example: we will no longer have the worry that there are an infinite number of dropping heights between 1 metre and the ground, say, because we can use the height dropped from, or the $n$-volume of the relevant region of case-space, as our measure of set size; however, there are still an infinite number of increasingly large heights from which an object can be dropped (or perhaps we should say ‘released’), the majority of which occur in the vacuum of space (and so would presumably not result in a manifestation of fragility).

A more appropriate analogy, then, would be an infinitely long line extending along the positive $x$ axis of a graph with its starting point at the origin, call it $X$. Again, there are an uncountably infinite number of points along $X$, and also in-between any two points that we might choose along it, so we would not be able to compare segments of $X$ in terms of their cardinality. Using the measure of length we would be able to compare discrete portions of $X$ with one another and calculate which segment was larger (and so which set of points along each segment was larger according to the
measure of length). However, seeing as there would be an infinite number of segments, we would not be able to calculate what proportion of $X$ any one segment of the line represented: what proportion of an infinitely long line is 0-1cm? To calculate this, we would need to divide the segment length by the total length of the line which, as discussed in the previous section, would give a proportion approaching 0.

Even weighted for contextual relevance or closeness (if appropriate) we would end up with the same problem. In figure 5, for instance, where we have just two variables, $x$ and a weighting dimension $w$, the heavily weighted regions (from 0 to 5 on the $x$ axis) initially appear to constitute a larger proportion of the total $n$-volume of the case-space, but assuming that $x$ extends to $\infty$, the un-weighted regions of the case-space ($x > 5$) will have a greater $n$-volume (an infinitely greater $n$-volume, in fact). That being the case, we would i) be faced with the potential problem of proportional swamping illustrated by steel schmeel 2, and ii) unable to actually calculate the
proportions of any discrete $n$-volume of the case-space to the whole space, making it impossible to properly run the analysis.

In order for there to be a proportion of manifestation at a context $K$ we must contextually restrict C-cases, so that certain regions of case-space are ruled out altogether.

Happily, this is again simple enough to realise: just allow the weighting dimension (or just the contextual relevance dimension) to hit zero, rather than just approach it. In figure 6, the weighted region of the case-space ends at $x = 10$, so whether $x$ extends towards $\infty$ is immaterial to the proportions of manifestation in the relevant regions of the case-space, as irrelevant regions would have 0 $n$-volume and not figure in our calculations of dispositional strength.

**Figure 6**
CONCLUSIONS

I have argued that in order to generate the strongest reading of (PROP) we must construe it as a contextualist account that takes a dual approach to analysing dispositions. Whilst a proportional approach such as (PROP) certainly should be preferred to a purely restrictive approach like the getting specific strategy, the possibility of proportion swamping cases such as steel, schmeel 2 mean that a purely proportional approach cannot hope to succeed. What we need is a reading of (PROP) according to which it has both a proportional and a restrictive element, which I have demonstrated is easily achieved by allowing contextually irrelevant cases to be given zero weighting in our calculations C-case proportions.

I have further argued that a version of (PROP) which uses some alternate measure (to their cardinality) to determine the sizes of sets of C-cases will also be susceptible to the probability swamping problems illustrated by steel, schmeel 2. Irrespective of which version of (PROP) we prefer, then, in order to generate the strongest reading of the account, we must interpret it as taking a dual approach to analysing dispositions.

REFERENCES


RELEVANT ABILITIES

Abstract: In this paper I defend dispositional compatibilism from a recent attack by Ann Whittle: Whittle argues that the dispositional accounts of abilities adopted by compatibilists such as Fara, Vihvelin and Smith, in order to defend the principle of alternate possibilities (PAP) against Frankfurt cases, latch on to what she calls a global rather than a local reading of abilities. She goes on to argue that as such global abilities are irrelevant to free and moral action, the dispositional compatibilist accounts fail to rescue PAP. I argue that in the first instance this objection misfires as it rests upon misconstruing PAP as making a claim about responsibility for outcomes, when in fact it is concerned only with responsibility for actions; I go on to demonstrate, that a modified version of PAP concerned with responsibility for outcomes can be defended by adopting a contextualist account of abilities built upon Manley and Wasserman’s contextualist analysis of dispositions.

Recently, compatibilists such as Fara (2005), Vihvelin (2004) and Smith (1997; 2003), the so called new dispositionalists, have revived the dispositional approach to abilities in order to defend the principle of alternate possibilities (henceforth PAP), from the challenge posed by Frankfurt style cases originating in Frankfurt’s (1969) ‘Alternate Possibilities and Moral Responsibility.’

A Frankfurt case, roughly, is one in which an agent freely performs some action, is intuitively free and responsible in so acting, and yet, due to the presence of some counterfactual or conditional intervener, is unable to act in any other way. According to the new dispositionalists what Frankfurt cases show is not that PAP is false but, rather, that we need a more sophisticated account of abilities, one based upon

87 This term is due to Clarke (2009).
an updated analysis of disposition; once we have adopted such an account, we will see that agents in Frankfurt cases really are able to do otherwise and hence that PAP is not threatened by such cases. Call this position *dispositional compatibilism*.

More recently, Clarke (2009) and Whittle (2010) have both independently offered a response to the new dispositionalists’ defence of PAP. I concentrate solely on Whittle’s response in this paper.

Whittle argues that the dispositional accounts of abilities suggested by the new dispositionalists latch on to what she calls a *global* rather than a *local* reading of ‘ability to do otherwise’ and further that upon considering a number of other Frankfurt type cases, we can see that such *global* abilities are simply not relevant to free and responsible action. That being the case, whilst global abilities might be retained in the presence of counterfactual interveners, no ability relevant to free and moral action is retained and the new dispositionalists’ defence of PAP therefore fails.

In this paper I provide two responses to Whittle’s discussion of dispositional compatibilism: first, in §2 I shall argue that Whittle’s discussion of Frankfurt cases is too simplistic and that once we are careful to distinguish between i) the action performed and ii) the outcome of each case we will see that there is no problem involved with PAP focussing upon *global* abilities to do otherwise: Whittle’s argument against dispositional compatibilism rests upon contrasting an agent’s responsibility for an action in one case with an agent’s lack of responsibility for some outcome in a different case; because PAP is focussed upon freedom and responsibility for actions, and not events, this argument misses its mark. In fact, I shall demonstrate, if we concentrate only on the actions performed in each of Whittle’s cases, we will see that the agents involved are all morally responsible and all have the global ability to do otherwise.
Second, in §3 I shall consider the plausibility of Whittle’s response to a modified version of PAP, which I label PAP*, according to which an agent freely brings about some event only if she could have brought about some other event. Against PAP* Whittle’s argument seems to hit its mark, however I shall argue that, due to an ambiguity in the notion of a local ability, a solution is left open to the dispositional compatibilist to defend this modified principle: I argue that rather than having to choose between having PAP* focus upon the necessity of local abilities assessed at the closest possible worlds to an agent’s actual circumstances and global abilities assessed at the majority of most, or all, worlds, PAP* can focus upon what I shall call relevant abilities which are assessed at a set of contextually relevant worlds to the case at hand. I shall then demonstrate how this resolves the issues Whittle raises for the dispositional compatibilist treatment of Frankfurt-type cases.

Finally, in §4 I argue, from the previous sections, that there are two main criteria that an account of abilities must meet in order to be suitable for grounding dispositional compatibilism: i) it must provide a solution to the various problem cases facing traditional dispositional accounts (discussed in §1 below), and ii) it must provide the right outcomes in the various Frankfurt cases discussed throughout §§1-3. I shall then provide a rough sketch for an analysis of relevant abilities, based upon Manley and Wasserman’s (2008) recent account of dispositions, and demonstrate that such an analysis can meet criteria (i) and (ii). I shall conclude that, in fact, both PAP and PAP* are better thought of as focussing upon relevant abilities, rather than either local or global.

First, though, in §1 I shall give a brief exposition of the new dispositionalists approach to defending PAP along with Whittle’s response.
1. THE NEW DISPOSITIONALISM

Traditionally, compatibilists such as Moore (1912) and Ayer (1954) have defended PAP against the threat posed by determinism by appealing to conditional analyses of abilities.

Briefly, PAP-compatibilists hold that an agent’s freely $A$-ing, or sometimes just being morally responsibility for some action $A$, requires that agent to have alternate possibilities open to them when they $A$. More formally, let’s say:

**PAP** An agent’s $A$-ing at $t$ is free only if they are able to do otherwise than $A$ at $t$.\(^{88}\)

Whilst perhaps intuitively quite plausible, this principle comes under serious threat from the possibility that our universe is causally deterministic: Supposing that Embla’s $A$-ing at $t$ is causally determined by the laws of nature and previous states of the universe, then there is a definite sense in which Embla cannot do otherwise than $A$ at $t$. If all of Embla’s actions are similarly caused, and given the truth of PAP, it seems that Embla is incapable of acting freely, even when her actions are proximally caused by, and thereby expressions of, the beliefs and desires that she most identifies herself with.

This result strikes many as unacceptable: even if our actions are causally determined, so long as they are caused in *the right kind of way*, we might think, we can still act freely and be responsible for those actions.\(^ {89}\) Accordingly, such thinkers adopt one of roughly two solutions: Some compatibilists abandon PAP and assume

\(^{88}\) I put PAP in terms of free action, though a corresponding reading could be generated in terms of moral responsibility: an agent is morally responsible for $A$-ing at $t$ only if they are able to do otherwise than $A$ at $t$.

\(^{89}\) In fact, we might take the view that we can *only* act freely when our actions are causally determined, as only those actions will be capable of expressing our agency.
that free action does not require an ability to do otherwise,\textsuperscript{90} whilst other, PAP-compatible, hold on to PAP but develop an account of abilities that is compatible with determinism.

The traditional move for PAP-compatible was to adopt a conditional analysis of abilities; for instance:

**ABLE** $S$ is able to $A$ iff $S$ would $A$ if $S$ were to try to $A$.

Effectively, ABLE treats abilities to the same simple conditional analysis that was given to dispositions, where some pro-attitude of the agent’s towards their action, ‘trying’ in the above formulation, plays the part of the stimulus condition for the disposition:

**SCA** $N$ is disposed to $M$ when $C$ iff $N$ would $M$ of $C$.

Call this the *dispositional approach* to analysing abilities.

According to ABLE, an agent’s being able to do otherwise than $A$ at $t$ is not dependent on their genuinely having two alternate possible futures open to them, given just one causal history, but rather upon the truth of a certain subjunctive conditional: if they were to try to $B$, where $B$ consists in some way in not $A$-ing, then they would $B$. As such, determinism no longer poses a threat to PAP: an agent can be able to do otherwise, even if there is only one course of action that they ever can perform given their causal history, so long as given a suitably different causal history they would do otherwise.

Unfortunately, this response, specifically the adoption of the simple conditional analysis of abilities, is open to several devastating counterexamples: First,

\textsuperscript{90}See, for instance, (Fischer & Ravizza, 1998) for a semi-compatible account that takes PAP to be false.
being a reworking of SCA, ABLE is open to all of the counterexamples that have troubled conditional accounts of dispositions.

Ask, for instance, might find himself in the presence of sorcerer who would remove his ability to jump should he so try, making him unable to jump according to ABLE; still, as Ask is a perfectly normal human being, we might think that he obviously is able to jump and that it is only due to the presence of an ability fink, manifested by the sorcerer in this case, that he would not succeed in jumping if he tried. Similarly, Ask might find his ability to jump masked by his being stood in a vat of honey. Ask is plausible still able to jump in this case, even though he would not succeed were he to try. Finally, a toddler might mimic the ability to jump five feet in the air by having an adult lift them up that high whenever they try to jump; quite clearly the toddler does not possess this ability, but ABLE will give the result that they do.91

Second, an agent might well perform some action A if they were to try to A, but be unable to so try and thus intuitively be unable to A. Take the following example:

Comatose John: John has suffered a head injury and is currently in hospital in a comatose state. His central nervous system is intact, so he would walk if he were to try but, being in a coma, John is unable to try to walk.

According to ABLE, John is able to walk: he would walk if he were to try; but clearly this is not an ability that we would want to attribute to him: there is a plain and obvious sense in which he is unable to do anything at all! Call this an impossible stimulus condition case.

91 For further discussion of these problems see (Martin, 1994) and (Lewis, 1997) on finking, (Johnston, 1992) on masking (see also Bird’s (1998) antidote cases), and Lewis’s (1997) ‘hater of Styrofoam’ example for mimicking.
In addition to these problems for ABLE, PAP itself faces a challenge from Frankfurt cases, which appear to show that an ability to do otherwise is simply irrelevant to free and moral action; take the following version of Frankfurt’s (1969) original case:

*Black and Jones*: Black, an evil scientist, implants a device in Jones’s brain. If Jones wavers in his intention to kill the Mayor, the device will be activated, forcing Jones to remain faithful to his original intention. As it turns out, however, Jones murders the Mayor and the device remains inactive.\(^2\)

The threat to PAP here is obvious: Jones appears to be acting freely when he kills the mayor and intuitively seems fully responsible for his behaviour; however, if we take the case at face value, Jones was unable to do otherwise according to ABLE, given the presence of the Black’s device: if Jones had tried to do otherwise, Black’s device would have intervened and caused him to carry out his plan to kill the mayor regardless. That being the case, Jones is free and responsible without having the ability to do otherwise and PAP is straightforwardly false.

Classic dispositional compatibilism fails in its defence of PAP, then: Frankfurt cases *prima facie* appear to refute the principle, and what’s more there are independent reasons to reject the naïve conditional account of abilities on which its defence is based: finking, masking and mimicking cases all show the simple conditional account underpinning (ABLE) to be inadequate.

Recently, however, the new dispositionalists have argued that what such cases show is not that an ability to do otherwise is irrelevant to free and moral action but,\(^{92}\)

---

\(^2\) This version of the Black and Jones case comes from (Whittle, 2010, p. 8).
rather, that we need a more sophisticated analysis of dispositions underpinning our account of abilities. Once such an account has been adopted we will avoid finking, masking, mimicking and impossible stimulus condition cases and see that Jones, for instance, really is able to do otherwise in a sense relevant to free and moral action.

Vihvelin and Smith, for instance, propose roughly that an agent is able to A if she has a certain bundle of dispositions, where those dispositions are grounded in some intrinsic properties of hers.\textsuperscript{93} Returning to \textit{Black and Jones}, because the presence of Black does not alter any of Jones’s intrinsic properties, his abilities to consider and respond appropriately to reasons remain intact and he therefore \textit{is} able to do otherwise when he kills the mayor.

Alternatively, Fara argues that an agent is able to A just in case they are in the habit of A-ing when they intend to A.\textsuperscript{94} Again, the presence of Black does not remove Jones’s habit of doing otherwise when he so intends—habits can be retained even when they fail to be manifested—and so he is able to do otherwise even though Black would prevent his doing so were he to try.

In both cases, Black acts as a \textit{masker} rather than \textit{remover} of Jones’s ability to do otherwise and so PAP is no longer under threat: Jones can maintain his ability to do otherwise even if he has no opportunity to express it.

More recently, however, Whittle (2010) has argued that i) this new variety of dispositional compatibilism fails to appropriately respond to the problems facing ABLE (finking, masking, etc.) and ii) whilst Jones may well retain some kind of ability to do otherwise in the presence of Black, this is not the kind of ability that looks to be relevant to free and moral action: Black clearly removes \textit{some} ability to do otherwise

\textsuperscript{93} See (Vihvelin, 2004) and (Smith 1997 & 2003); see also (Lewis, 1997) for his revised conditional analysis of dispositions upon which this approach is based.

\textsuperscript{94} See (Fara, 2008); see also (Fara, 2005) for his habitual analysis of dispositions on which his ability account is based.
from Jones, so why should we think that the ability that he maintains is the kind PAP
claims to be necessary for freedom and responsibility?

Whittle asks us to consider the new dispositionalists position in terms of a
distinction between what she calls local and global abilities: a global ability is one had
in a wide range of situations, even in the absence of the opportunity to express that
ability; for instance: John may be able to walk even though he is currently tied up,
Sally may be able to sing even though her aunt paralyses her with fear, and Sam may
be able to swim even though he is nowhere near water. These global abilities are a
kind of general capacity to act, grounded in in an agent’s succeeding in acting in the
relevant way in a majority, or large number, of possible cases in which they try. Such
abilities are not lost when the particular circumstances one finds oneself in do not
permit their expression: Roger Federer does not lose the ability (in the global sense)
to play tennis whenever you take away his tennis racquet, merely the opportunity to
exercise that ability (Whittle, 2010, pp. 2-3).

A local ability, on the other hand, is one had in a particular set of circumstances
at a particular time with the corresponding opportunity to express that ability. So, John
lacks the ability-to-walk-when-he-is-tied-up, Sally lacks the ability-to-sing-in-her-
aunt’s-presence, and Sam lacks the ability-to-swim-when-not-submerged-in-water; in
each case, the use of hyphens indicate that the circumstances are part of the
characterisation of the ability in question (Whittle, 2010, pp. 2-3).

In spite of their differences, local and global can be thought of as two ends of
a single gradable scale: on the local end we assess whether an agent would A, for
instance, if they were to try in just one maximally specific set of circumstances,

---

95 Whittle makes this suggestion in her description of ‘fairly local abilities’ (2010, pp. 4-5) and in her
talk of abilities being ‘more local.’ (2010, p. 9)
whereas on the global end we assess whether an agent would $A$ if they were to try to $A$ in a suitable number of all possible scenarios. This scale can be nicely brought out by Manley and Wasserman’s analysis of dispositions.

Briefly, Manley and Wasserman (2008) have proposed that we analyse dispositions as follows:

**PROP** $N$ is disposed to $M$ when $C$ iff $N$ would $M$ in a suitable proportion of $C$-cases

A $C$-case, here, is a “stimulus condition case” (Manley & Wasserman, 2008, p. 74), or more loosely a circumstance case, so that for every maximally precise description of a set of circumstances $C$ there is a corresponding $C$-case in which those circumstances are exemplified. If an ability to $A$, more generally, is understood as a disposition to $A$ when one tries to $A$, we can assess an ability to $A$-in-$C$ by considering whether an agent would $A$ when they so try in a suitable proportion of all cases in which circumstances $C$ obtain.

Importantly, because our descriptions of circumstances can be more or less finely-grained, we can refer to more or less $C$-cases depending on the precision of our $C$ description: A maximally precise description of an agent’s circumstances would pick out just one $C$-case that exemplifies every feature of that description, whereas a less precise description would pick out a greater number of $C$-cases: all of those that exemplify *at least* the circumstances described.

A maximally local ability, or what Whittle follows Austin (1956) in calling an “all-in” ability (2010, p. 4), would be one in which the $C$ description is maximally precise and is therefore assessed at only one $C$-case; we can then move towards more global abilities by relaxing the precision of our $C$ description and thereby allowing the ability to be assessed at a greater number of $C$-cases, all the way up to a maximally...
global ability which is assessed at all $C$-cases. An ability $x$ is \textit{more local} than some other ability $y$ just in case it is assessed at fewer $C$-cases than $y$ (its $C$ description is more precise), whilst $x$ is \textit{more global} than $y$ just in case it is assessed at fewer cases than $y$ (its $C$ description is less precise).

Returning now to \textit{Black and Jones}, the dispositional compatibilists position turns out to equate to the following: whilst Jones obviously lacks the local ability to do-otherwise-in-the-presence-of-Black, he maintains a global ability to do otherwise whilst in Black’s presence: If Jones tries to do otherwise in the exact circumstances in which he finds himself, he will fail—Black will prevent him from doing so; however, if we widen our assessment of Jones’s abilities to consider all $C$-cases in which he tries to do otherwise we will find that he generally does do otherwise when he so tries. So, Jones \textit{is} able to do otherwise in the global sense in the presence of Black and if this is the sense of ability that PAP latches on to then it looks to have been rescued.

The new dispositionalists, then, can be interpreted as adopting an updated account of abilities that casts them as global:

\textbf{D-ABLE} \quad \text{an agent } S \text{ is able to } A \text{ just in case she would } A \text{ if she so tried in a suitable proportion of all } C\text{-cases.}

Also, we can interpret the new dispositionalists as adopting a slightly modified version of PAP that targets these this kind of ability:

\textbf{D-PAP} \quad \text{an agent freely } A \text{'s at } t \text{ only if they have the global ability to do otherwise than } A \text{ at } t.

Unfortunately, this response to Frankfurt cases turns out to be unworkable: Whittle asks us to contrast Black and Jones with another Frankfurt style case, this time where
the agent retains their global ability to do otherwise but lacks freedom and responsibility for the ensuing events all the same:

**Bound Ben**: Ben, an excellent swimmer, has been forcibly bound to a chair. He watches helplessly as a child drowns in a lake.

(Whittle, 2010, p. 10)

It looks obvious here that Ben is not responsible for the child’s drowning, and nor does he appear to let the child drown freely. However, if we look at all C-cases in which he tries to swim (or perhaps tries to save a drowning child) we will find that, being an able swimmer, he is successful in the majority of those cases in doing so; in that case, Ben has the global ability to save the child despite being tied to the chair—tying him up does not remove his capacity to swim, after all. So the global ability to do otherwise looks to be irrelevant to Ben’s freedom: whether or not he has a general capacity to save the child is beside the point when evaluating his current circumstances; rather it is his local ability-to-do-otherwise-in-his-circumstances, or more accurately his lack of that ability, that is pertinent to his freedom and responsibility: it is precisely because he cannot save the child in his very circumstances that make us think that he is not responsible for their drowning.

Could PAP latch on to a local notion of ability instead? Perhaps fairly obviously, no: it was precisely this kind of local ability that Jones lacked in *Black and Jones*.

We need the same kind of ability to link up with the agent’s freedom, or lack of freedom, in all of the cases under consideration and, as we have already seen, Jones is unable-to-do-otherwise-in-his-current-circumstances, though he *is* free and
(plausibly) morally responsible for his actions, so again if PAP latches onto this local sense of ability it is straightforwardly false.

The dispositional compatibilist is left with a dilemma, then: PAP cannot latch onto the notion of local ability, because Jones is free and responsible and yet lacks the local ability to-do-otherwise-in-the-presence-of-black, but cannot latch onto the global notion either in that Ben is not free or responsible and yet has the global ability to do otherwise. On this basis, Whittle concludes that dispositional compatibilism fails in its response to Frankfurt cases and in establishing that an ability to do otherwise is required for freedom.

Now, it is important to flag up here that PAP provides necessary conditions for free or moral action, not sufficient. So there may be a problem with Whittle’s argument: granted, Ben is able to do otherwise in the global sense and is not morally responsible or free, but this may be for any number of other reasons. Neither PAP nor the new dispositionalists were suggesting that an agent is free or morally responsible whenever they have an ability to do otherwise, just that such an ability was necessary for free or moral action.

It might appear then, that Whittle’s argument relies upon a mischaracterisation of PAP. However, I think her point here is that without any substantial difference between Black and Jones and Bound Ben, we have no principled reason for treating them any differently: Ben’s situation follows the Frankfurt case formula set out in Black and Jones almost exactly: an agent (Jones/Ben) acts in such a way (shoots the mayor/does not save the child) such that a certain event is caused to come about (the mayor dies/the child drowns), had they tried to do otherwise they would have been prevented from succeeding (by Black/by the bindings). Clearly, if any kind of ability to do otherwise is relevant to Ben’s moral responsibility in Bound Ben, it is the local
ability to do otherwise: in this case the local ability to save the child which Ben lacks.

But, that being the case, why is it not a local ability targeted by PAP, and why is it the
global ability that we should think of as being relevant to Jones’s freedom and
responsibility.

We might, at this point, try to point to some feature of the two cases that can
suffice to show that they are not really as similar as Whittle would have us believe. Notably, Jones performs an action whereas Ben refrains from acting and we might
think that this has an effect on our intuitions about the case and the agents’ abilities in
them.

Whittle does not directly deal with this difference, however she does make the
following point regarding the discounting of Black’s device in Jones and Black due to
its inactivity:

…this response risks greatly underestimating the importance of
absences in our causal systems. Why assume that the inactivity of the
device has no consequences for Jones’s abilities and so can be safely
ignored? The inactivity of a person’s pacemaker will affect what that
person is able to do. The inactivity of a sprinkler system may affect a
plant’s ability to flourish when its owner goes away. The inactivity of
a mugger affects your ability to walk home safely. Although the causal
status of absences is controversial, no one will deny that, in some sense,
they are causally significant factors. So the mere fact that Black’s
device is not activated does not suffice to show that it has no effect
upon Jones’s abilities. (Whittle, 2010, p. 13)
Perhaps this same line of reasoning could be used to reject treating the two cases differently in respect of Jones’s action in one and Ben’s inaction in the other; clearly in some cases inaction can be an important factor, consider the following modification of *Black and Jones*:

*Black and Schmidt*: Schmidt learns of a bomb that has been planted in the mayor’s house which will go off at 6pm tonight. Schmidt thinks about reporting the bomb to the authorities, but given that he dislikes many of the mayor’s policies, he decides it is more in his interest to let him be killed by the bomb. Unbeknownst to Schmidt, Black has planted a device in his brain to ensure that he doesn’t intervene with the bomb’s going off. If Schmidt wavers in his intention to let the bomb go off, Black’s device will activate and force him to remain faithful to his original intention. As it turns out, however, Schmidt does not have a change of heart, the bomb goes off and the device remains inactive.

In this case, Schmidt is unable to do otherwise than remain inactive (with respect to the bomb), and his inaction is surely necessary for the mayor’s being killed in much the same way that Jones’s activity is required in *Black and Jones*. What’s more, like Jones’s shooting the mayor, Schmidt appears to be morally responsible for his refraining from altering the authorities to the bomb. Perhaps, in that case, we *can* treat Jones’s action and Ben’s inaction as similar enough for Whittle’s point to go through.

In fact, I think there are several inescapable differences, aside from this, between *Black and Jones* and *Bound Ben* which affect our intuitions about the responsibility of Jones and Ben for various features of the two cases, some of which I
shall go on to discuss in §3. However, I think that PAP can be defended even if we
grant the above point about the similarity of action and in-action in the two cases, as I
shall argue in the next section.

2. ALTERNATE POSSIBILITIES FOR ACTION

There are two complications with Whittle’s argument in the previous section, the first
of which, that I shall discuss in this section, concerns the need to appropriately
distinguish between two main elements of each of the Frankfurt type cases discussed:
the action performed by the relevant agent, and the event that constitutes the outcome
of the case. The second complication, which I shall turn to in the following section,
concerns a possible ambiguity in the distinction between local and global abilities.

Each of the Frankfurt type cases that Whittle discusses is made up of roughly
two halves: an action and a resulting event, and these need to be carefully identified
so that we can appropriately consider what the agents in each case are potentially
responsible for.

In Black and Jones, the action element of the case is Jones’s shooting the
mayor, with the outcome of the case being that that the mayor dies. Now, it might be
tempting to confuse these two elements by describing Jones’s action as ‘killing the
mayor’ or the outcome as ‘the mayor’s being killed by Jones.’ In fact, it might be
somewhat contentious even whether Jones does, in fact, kill the mayor: if killing
involves causing to die, then on certain accounts of causation Jones does not cause the
mayor to die nor kill him: according to a counterfactual account of causation, for
instance, Jones does not cause the mayor to die, because if he were to choose not to
shoot the mayor Black’s device would force him to shoot him as he originally
intended. I think, therefore, that we should stick with ‘Jones’s shooting the mayor’
for the action in Black and Jones, and ‘the mayor’s dying’ as the outcome.
When we come to *Bound Ben* we find that the case is somewhat under-described: we are told that Ben watches helplessly as a child drowns, but what action does he perform? Does he struggle against his bonds in an effort to break free and save the child? Or does he sit and watch the child without so much as trying to save them?

I suggest that we move away from *Bound Ben* to focus instead on two modified cases: *struggling Sam* and *indifferent Ian*.

*Struggling Sam:* Sam, a competent swimmer, has been forcibly bound to a chair and watches helplessly as a child drowns in a nearby lake. He struggles desperately against his bindings but cannot break free to save the child.

*Indifferent Ian:* Ian, a competent swimmer, has been forcibly bound to a chair and watches as a child drowns in a nearby lake. He does not struggle against his bonds, in fact he is quite content to let the child drown and therefore makes no attempt to try save him.

With these two modified cases we are in a position to identify the action and outcome of each case: in *Struggling Sam*, the action Sam performs is ‘struggling against his bonds’ with the outcome of the case being that ‘the child drowns’; in *Indifferent Ian*, the action, to the extent that refraining from acting can be considered an action, that Ian performs is ‘refraining from struggling against his bonds’ with the outcome of the case again being ‘the child drowns’.

Now, the problem arising out of this complication for Whittle is that her argument relies upon comparing Jones’s freedom and responsibility for his act of shooting the mayor, with Ben’s responsibility for ‘the child’s drowning’: in her discussion of *Bound Ben*, Whittle asks: ‘[is] Ben morally responsible for not
preventing the child’s death?” and later in that same passage, she states that “what we should be considering when judging whether he was morally responsible for the child’s drowning is whether he instantiates the fairly local ability-to-swim-when-bound-to-a-chair” (my emphasis) (2010, p. 10). The argument against the relevance of global abilities, rests upon the fact that whilst the two cases show no significant differences, where Jones is morally responsible for shooting the mayor, Ben is not morally responsible for the child’s drowning.

‘The child’s drowning’, or ‘Ben’s not preventing the child from drowning’ is an event, though, and not an action. It is clearly true that in their respective cases Ben, Sam and Ian do not prevent the child from drowning, but this is a description of the case, and not of the actions that they perform. As already seen above, the actions that they perform in each case are struggling against their bindings, or refraining from struggling against their bindings.

Now, PAP is a principle concerned only with actions—an agent freely As at t only if they are able to do otherwise than A at t—and that being the case, what we should be comparing is Jones’s shooting the mayor, Sam’s struggling and Ian’s not struggling against his bonds. In each case, the agents have the global ability to do otherwise: Jones has the capacity not to shoot the mayor, even in the presence of Black, Sam has the capacity to not struggle and Ian has the capacity to struggle. What’s more all of the agent’s appear morally responsible for these actions: we already know that the intuition in Black and Jones is that Jones is morally responsible without having a local ability to do otherwise; in struggling Sam and indifferent Ian, Sam and Ian not only have the global ability to do otherwise but also the local ability to do otherwise: Sam could refrain-from-struggling-Whilst-bound-to-the-chair, and Ian could struggle-against-his-bonds-Whilst-bound-to-the-chair, and we surely would want to hold them
morally responsible for their actions: we would want to say, in all likelihood, that Sam is to be praised for struggling against the odds, whilst Ian is to be criticised for not even trying to save the child (especially considering that he does not try because he actively approves of the child’s drowning).

Now whether or not Jones is responsible for the event ‘the mayor’s being killed’, or Sam or Ian are responsible for ‘the child drowning’ is both a matter of contention and immaterial to the plausibility of PAP. Even if it were established that Jones is responsible for the mayor’s death, whereas Sam and Ian are not responsible for the child’s death, this has no impact on the plausibility of PAP which is only concerned with responsibility for actions.

3. ALTERNATE POSSIBILITIES FOR OUTCOMES
Given the argument in the previous section, Whittle’s argument will not hit its mark against a version of PAP that targets only actions. However, we can imagine a corresponding principle to PAP that focusses upon an agent’s responsibility for events, call it PAP*:

\[
\text{PAP*}: \text{An agent is responsible for some event } E's \text{ occurring only if they were able to bring about some other distinct event } E_i.\
\]

There is some limited plausibility to PAP* if we have already endorsed PAP: if we suppose that being responsible for some event’s occurring requires you to have caused that event to occur through some freely performed action of yours, the performance of which requires an ability to do otherwise, then being responsible for an event’s occurring, too, will require an ability to do otherwise.

Now, interpreting the new dispositionalists as defending PAP*, the argument from §1 against this defence is on target: Jones is at least plausibly responsible for the
outcome of *Black and Jones*—the mayor’s dying—whereas Ben, Sam and Ian are all clearly not responsible for the child drowning in their respective cases, given that they were unable to bring about any other outcome in the local sense.

Let us return to the workings of Whittle’s argument against the new dispositionalists. There is a complication in the local/global terminology: ‘local’ is ambiguous between ‘a specific set of circumstances’ and ‘the specific set of circumstances in which an agent actually finds themselves.’ On the first reading, I can have many local abilities right now—the ability-to-swim-when-I-am-in-a-pool, the ability-to-play-tennis-when-i-am-on-a-tennis-court, the ability-to-play-the-piano-when-I-am-sat-at-one, and so on—even though I am not currently the circumstances that would allow me to express those abilities. In the second sense, I lack the ability to do all of these things—swim, play tennis, play the piano—because the circumstances local to me involve being on a train which, unfortunately, has neither a swimming pool, a tennis court nor a piano.

Local, then, can mean, on the one hand, a detailed specification of the scenario at which we assess an ability, and thereby a greater level of restriction on C-cases; for instance, when assessing my ability to swim-in-a-pool, our assessment is localized to all and only those C-cases in which I am in some kind of pool. This, reading of local corresponds to the gradable reading of local and global already discussed above.

On the other hand, local can refer to the circumstances local to an agent, as in: the very circumstances in which they find themselves. So when assessing my local ability to swim we need to assess my ability to swim-in-my-very-circumstances.

We could draw this distinction in terms of the localisation of the C-cases that we must assess: are they localised to a C description, or are they localised to the agent that has the ability?
A problem with Whittle’s assessment of dispositional compatibilism arises out of this ambiguity in that whilst she introduces the notion of local ability under the first interpretation, she then appears to apply it to the two cases according to the second:

In characterising the concept of a local ability, she quotes Mele on ‘specific practical abilities’:

…an ability an agent has at a time to A then or to A at some specified later time. (2010, p. 2)⁹⁶

She goes on:

If determinism threatens anything, it threatens only Mele’s “specific practical” abilities, Berofsky’s “token” abilities, or what I shall call local abilities — the ability-to-walk-in-circumstances-\(C\). (Whittle, 2010, pp. 2-3)

Nothing here suggests that the circumstances-\(C\) have to be the circumstances that an agent is actually in; and, of course, we should expect agents to have these kinds of abilities when not in the circumstances that characterise them: I am able to swim-in-a-swimming-pool irrespective of whether I am in a pool currently.

Later, in discussing the application of Manley and Wasserman’s (PROP) in bringing out the distinction between local and global, Whittle says:

…an object \(O\) has the “all-in” local disposition to \(D\)-in-\(W\) (where \(W\) is a particular type of \(C\)-case) if \(O\) instantiates a property complex in virtue of which \(O\) would (standardly) \(D\)-in-\(W\)…So the idea is that the most local of dispositions concern what the object is likely to do in just

⁹⁶ This passage is a quote from (Mele, 2003, p. 447)
one type of C-case, which specifies every variable concerning the state of the object, its circumstances, and the stimulus it is subject to. (2010, p. 4)

And then subsequently:

…an object \( O \) has the (fairly) local disposition \( D \) if \( O \) instantiates a property complex in virtue of which \( O \) would \( D \) in a suitable proportion of a range of C-cases, in which certain circumstances in that range of C-cases are held fixed. (2010, pp. 4-5)

Again, in both of these cases, the type or types of C-cases that are held fixed are not designated as the circumstances in which the relevant agent actually finds themself; indeed the suggestion is that this could be any type of C-case that we are holding fixed.

However, when we get to the treatment of Jones and Ben, Whittle appears to treat local abilities in the second sense as those only had in one’s actual circumstances, thereby presenting a false dichotomy of global abilities on the one hand and abilities local to the agent’s circumstances on the other. For instance, of Jones she says:

We might agree that since there are many occasions in which, were Jones to try to do otherwise, he would succeed, Jones does have the global ability to do otherwise. But we might insist that what is at issue is whether Jones has the opportunity to act otherwise in this situation. In other words, whether he has the local ability-to-do-otherwise-with-the-device-present. (2010, p. 8)

Then, again, talking about Ben later in the paper:
Ben has the global ability to swim since, given a suitable range of \( C \)cases, he usually swims if he tries. But this global ability is intuitively irrelevant to the question at hand. Given that Ben is bound to a chair, what we should be considering when judging whether he was morally responsible for the child’s drowning is whether he instantiates the fairly local ability-to-swim-when-bound-to-a-chair. (2010, p. 10)

Now, it is indeed true that Jones has the global ability to do otherwise whilst lacking the local ability-to-do-otherwise-in-the-presence-of-Black, and that, likewise, Ben has the global ability to swim whilst lacking the local ability-to-swim-when-bound-to-a-chair; however, we simply needn’t agree that these are the only two kinds of ability that could be relevant to their responsibility. In Jones’s case for instance we needn’t go so far as a fully global ability to find an ability that he can retain in his current circumstances: whilst Black’s presence precludes a local ability to do otherwise in the second sense, it does not preclude Jones having a local ability to do otherwise in the first sense: there are many specific circumstances in which, if Jones were to try, he would succeed in doing otherwise—for instance, those which involve the absence of Black. It might be, then, that the relevant local ability to Jones’s acting freely, is not one centred on his current circumstances, but localised to some other description.

To avoid further confusion over the meanings of ‘local’ and ‘global’ I suggest we reserve these terms for the gradable abilities centred on an agent’s actual circumstances—so a maximally local ability would be had in the exact circumstances that an agent is in at any given time with abilities getting more global as we relax the description of the agent’s circumstances and include more \( C \)-cases in our assessment of their abilities—and use Mele’s (2003) terminology of ‘specific’ and ‘general’ abilities to refer to abilities localised to specific \( C \) descriptions—so a maximally
specific ability would be one had in just one C-case, but where that C-case need not exemplify the agent’s actual circumstances, with that ability getting more general the greater the number of C-cases at which it is assessed.\footnote{Note that given this reading of specific and general abilities, where they need not be centred on any world in particular, we could have ‘patchy’ abilities that are to be assessed at a range of disconnected worlds with no smooth connections between them (in terms of closeness to one another).}

Now, with Ben, it may well be that his global or general ability to swim is irrelevant to whether the child drowns, but that might simply be because we ought never to have been looking for such coarse-grained abilities anyway: Instead we should have been looking to see whether he had a certain relevant specific ability that was not necessarily his local ability to swim.

As Whittle points out, the relevant ability to Ben’s responsibility appears to be the specific ability to swim-when-he-is-bound-to-a-chair, which just so happens to coincide with his own circumstances (meaning it is also a local ability); for Jones, on the other hand, I contend that the relevant ability to his responsibility is his ability to-do-otherwise-when-reasoning-normally (seeing as Black does not intervene in the actual circumstances), which is perhaps more global than Ben’s ability, but not necessarily a maximally global ability; in any case, the relevant ability to assess Jones’s responsibility is certainly not local.

The intuition behind this response is that the relevant ability to assessing an agent’s capabilities and responsibility may be localised to a different set of circumstances from case to case, and those circumstances will only sometimes be the agent’s actual circumstances.

This solution also picks up on the intuitions of the new dispositionalists: Jones’s specific ability-to-do-otherwise-Whilst-reasoning-normally can be masked by the presence of Black’s device, because it removes the opportunity for its expression.
So, whilst Whittle has demonstrated that global abilities are not sufficient to ground moral responsibility or freedom, she has failed to demonstrate that there is not some kind of relevant specific ability that might be necessary for moral or free action.

4. RELEVANT ABILITIES

From the discussions in the previous sections, we can stipulate two criteria that an account of abilities must meet in order to be suitable for the dispositional compatibilists’ purposes:

1. it must solve or avoid the problem cases facing ABLE and D-ABLE: finking, masking, mimicking and impossible stimulus condition cases; and
2. it must avoid predicating both Jones and Ben with the same kind of ability to do otherwise (for the defence of PAP*).

Can an account of relevant abilities help us to meet these two criteria? I think, yes:

What is important, it seems, in both Jones’s case and Ben’s, is not that they have some fixed local or global ability to do otherwise, but rather that they have a relevant specific ability to do otherwise that constitutes their having a capacity to respond appropriately to reasons in contextually relevant circumstances.

Just as, for instance, if we want to know whether Federer is able to beat Murray in the final we are not concerned with whether Federer would win in worlds in which Murray is a penguin or in which the match is conducted on the moon, so too are we not interested in whether Jones is able to do otherwise only in the presence of Black (given that he does not intervene), nor whether Ben can save the child when not bound to a chair (given that the binds actually play a part in restricting him). Our analysis of relevant abilities, then, needs to make clear that we are assessing the abilities in question only at relevant C-cases:
R-ABLE  

S is able to A iff she would A in a suitable proportion of relevant cases in which she tries to A

This modification of ABLE takes us from global abilities that must be assessed at all possible C-cases, to the relevant specific abilities that I have argued we should focus on.98 We can still maintain the proportional element from Manley and Wasserman’s (PROP): it might be that sometimes the relevant ability need not be had in every relevant C-case: Mickelson is able to make 10 foot putts, we might say, because he does so in a suitable proportion of the relevant cases in which he tries; the fact that he occasionally misses ought not to dissuade us of this, and nor should we have to find some contrived method for excluding these missing cases from our assessment.

Now, consider how R-ABLE deals with our criterion: first, the finking, masking and mimicking problems faced by SCA, ABLE and D-ABLE are no longer any real threat. Being based upon Manley and Wasserman’s PROP, which was designed to deal with such problems for dispositions, R-ABLE is able to easily overcome these worries: the proportional element to R-ABLE along, along with the restrictive element of assessing abilities at only those cases which are relevant to the conversational context at hand mean that finking, masking and mimicking scenarios are rarely going to form a part of our assessment of an agent’s abilities, and when they do they will constitute only a small number of the overall cases assessed. So long as the agent successfully performs the relevant action in a suitable proportion of the non-finking and non-masking cases, they will come out as having the ability in question with the reverse holding for mimicking cases (if the agent fails to A when they try in a suitable proportion of non-mimicking cases, they will lack the relevant ability to A).

---

98 This is the same move that I suggest be made with PROP in ‘Getting Specific with Manley and Wasserman’ (Chapter 2).
With regards to impossible stimulus condition cases, R-ABLE again ought to give us the right assessment. Consider, again, Comatose John: the worry was that John would walk if he were to try, but is unable to try to walk. In a fairly plain sense John is not able to walk, but according to ABLE, and D-ABLE John is able to walk.

R-ABLE is equipped to deal with this: we first need to ascertain what the relevant ability is that we are looking for. That is, asking if John is able to walk in isolation of any kind of circumstances or considerations is incomplete. We need to ask whether John can walk ‘in light of’ some set of circumstances in order to determine which are the relevant worlds at which to assess this ability.99

Suppose, then, that two doctors are discussing whether to fit John with a catheter. One of the considerations in this decision is whether John can walk and thereby get himself to the bathroom when needed. Clearly in assessing this ability, his being in a coma is a relevant factor, so we need to assess it in all and only those worlds in which John tries to walk whilst he is in a coma. Of course though, there are no such worlds—in all worlds in which he is in a coma he is incapable of trying to walk—and so John will not walk in the suitable proportion of C-cases required for having that ability.

Now suppose that the doctors meet a second time to discuss the state of John’s central nervous system after the crash. In considering whether John can walk in this case, they would be considering whether John can walk ‘in light of his bodily injuries.’ This ability might well be assessed at cases in which John is not in a coma, and so it might be true in this context that John is able to walk; but what should be made clear is that this is a special kind of context, and a vastly different one to where we are focussed on the fact that he is in a coma.

99 In this way, R-ABLE treats ‘is able’ in much the way that Elaine Kratzer (1977) suggests we treat ‘can.’
With regards to the second condition (relevant only for the defence of PAP*), we have already seen above how this might work out: in *Black and Jones*, a salient feature of the case is that Black’s device is never activated, so in assessing Jones’s ability to do otherwise (his capacity to respond differently to reasons, remember) we should not be restricting ourselves to cases in which the device is present. Given the way that things actually go, and the mechanism that brings about Jones’s shooting the mayor, the relevant cases appear to be all and only those in which Jones’s deliberates about killing the mayor with or without Black’s device.

In *Struggling Sam*, the ropes binding Sam to the chair play a role in the actual sequence of events leading to the child’s drowning, and so we ought to hold them fixed in any of the worlds that we consider.

Ian’s case in *Indifferent Ian* is slightly more complicated: Ian does not struggle and so the ropes binding him to the chair play no role in preventing him saving the child in the actual sequence. Like Black in *Black and Jones*, the ropes can be triggered by Ian’s trying to bring about a different outcome to the case, saving the child, but will otherwise remain inactive. The difficulty then might be that if we assess Jones’s ability to shoot the mayor in the absence of Black, because he plays no role in the actual sequence leading to him doing so, we ought to assess Ian’s not saving the child in the absence of the ropes binding him to the chair.

However, this would be to, once again, confuse responsibility for an action with responsibility for an event. In assessing Jones’s responsibility for shooting the mayor, we are interested in his capacity to act differently (given the presence of reasons for doing so perhaps), so the presence of Black is not relevant (given that he plays no role in Jones’s deliberation); likewise, if we are assessing Ian’s *action*—not
trying to save the child—the bindings are not relevant: we are assessing his decision not to save a child and thereby his capacity to decide otherwise.

In considering Jones’s responsibility for the mayor’s death and Ian’s responsibility for the child’s drowning, however, we need to consider a different pair of abilities assessed at different sets of worlds. It is not entirely clear how to identify the relevant sets of worlds to these cases, and a challenge remains for those who endorse a contextualist strategy such as R-ABLE to explore by what mechanism a context picks out a particular set of worlds. I submit that there are various disparities between the cases that will make the set of worlds different in each case: we have already recognised that Jones acts whilst Ian fails to act, but consider also that a large part of the Black and Jones narrative is geared towards stressing that it is Jones who sets out to bring about the outcome of the case, the mayor’s death, and that he is the key causal factor in the actual sequence for that event.

In indifferent Ian, on the other hand, the child’s drowning has not been orchestrated by Ian, although he is content to let it happen, rather the case begins with the child already drowning: Ian is introduced into the situation with a sufficient cause existing for the child’s death. So even if Ian’s absence of action can be considered a cause of the child’s death (perhaps in a very loose sense we might say: if he had saved the child the child would not have drowned) it is only a partial cause at best. This difference between the cases, may have some large effect on which cases we are willing to assess the abilities of the two agents at: for Jones, who sets the mayor’s death up himself, we may be more inclined to not confine our assessment of his capacities to act otherwise to just those worlds in which Black is present, perhaps because we are assessing his responsibility for the planned death of the mayor; for Ian, however, who comes to the case with the child already drowning, we are inclined to restrict our
assessment of his ability to prevent the death to those involving his present circumstances, because, I suggest, we are assessing his responsibility for failing to prevent a drowning that was already in progress.

Our first responses to these cases can be quite telling: when we talk about Jones’s responsibility for the mayor’s dying we care more about his actually having fired the shot and having had the capacity, if not the opportunity, to do otherwise, whereas when we talk about Ian’s responsibility for the child’s drowning we are more concerned with his more local inability to make a difference to the situation. I take this as motivation for the contextual approach to abilities and PAP that is suggested by R-ABLE.

Whittle does acknowledge that accounts such as R-ABLE may exist—given that she endorses abundant fine-grained abilities, Jones’s ability to do otherwise can be assessed in numerous ways at numerous worlds—it is just, she thinks that such alternate assessments are not particularly obvious or forthcoming (Whittle, 2010, p. 15); I think that this approach seriously risks ignoring our intuitive responses to these cases and overlooking the contextually sensitive nature of our natural conception of responsibility.

CONCLUSIONS
I have argued that Whittle’s attack on the new dispositionalists’ updated dispositional compatibilism fails: in the first instance her argument misses its mark as it relies upon the agents in Black and Jones and Bound Ben differing in their responsibility for the outcomes of the cases she considers. In fact, because PAP is a principle concerned with responsibility for actions, rather than for outcomes or events, it is not threatened by this argument, and as I demonstrated the agents in these cases both have the global ability to do otherwise and are responsible for their actions.
The dispositional accounts of abilities provided by Fara, Vihvelin and Smith do not, however, answer the challenges presented by finking, masking, mimicking and impossible stimulus condition cases. I have suggested therefore that we adopt the contextualist account of abilities, R-ABLE, built upon Manley and Wasserman’s (PROP) to overcome this worry.

This account not only avoids the various problem cases facing other dispositional accounts, but also allows for the defence of PAP* against Whittle’s original argument.

There is still much to do in order to complete R-ABLE, I have only given a sketch; most pressing, perhaps, is to identify the mechanism by which worlds are selected as relevant to a particular context. Once this has been accomplished though, I am confident that a contextualist account such as R-ABLE will allow dispositional compatibilists to mount an exceptional defence of PAP against Frankfurt cases.

REFERENCES


CONCLUDING REMARKS:

**INTO THE FUTURE**

Each of the papers in this collection has its own concluding remarks, so I shall not repeat too much of what has been said in those section; to briefly sum up: In ‘The Anti-Akrasia Chip’ I argued that the Fischer-Ravizza account of guidance control is in error; either, taking responsibility for a moderately reasons-responsive mechanism is not sufficient for having guidance control, or guidance control does not equate to actually guiding one’s behaviour. I pointed here to Fischer and Ravizza’s focus upon moral responsibility as the root of the issues in their analysis and suggested that more emphasis needs to be placed upon an agent’s ability to effectively intervene in their behaviour.

In ‘Getting Specific’ I argued for a *dual approach* reading of Manley and Wasserman’s (PROP) and demonstrated that such an approach is necessary irrespective of whether we adopt measure theory to compare the sizes of sets of C-cases. Finally, in ‘Relevant Abilities’ I argued that dispositional compatibilists defence of PAP does not necessarily fail: in the first instance Whittle’s argument appears to misfire, and in the second (when considering the modified version of PAP concerned with responsibility for outcomes) we do not have to accept the dichotomy of agents either having local or global abilities. I argued that agents can have *relevant* specific abilities to do otherwise which in turn can ground a compatibilist defence of PAP. I closed by offering a sketchy analysis of this context sensitive notion of ability.

The three papers make some headway in filling in the sketch for an account of agential control that I presented in the introduction; of, course there is still a great deal of work left to do: there is perhaps an entire dissertation’s worth of work to be done
on the awareness element to control. The most pressing element of the account to be completed at this stage, though, is the account of abilities.

The account that I gesture to in ‘Relevant Abilities’, R-ABLE, analyses ‘S is able to A’ as ‘S would A in a suitable proportion of relevant worlds in which she tries to A.’ This effectively equates to the subjunctive condition ‘if S were to A then S would A in a suitable proportion of relevant worlds.’

Now, whilst I argued that this allows us to overcome a particularly devastating problem for simple conditional analyses, that of impossible stimulus condition cases, I did not address another fairly glaring problem for such accounts: specifically, that if a conditional account of ‘S is able to A’ has as its antecedent ‘S tries to A’ it will find it impossible to account for fairly uncontroversial abilities to try: if an agent is able to A iff they would A if they tried to A, then an agent is able to try-to-A iff they would try-to-A if they tried to try-to-A; but what is it to try to try? One either tries, or does not try, there is no sub-trying needed to get one to try to perform an action.

Supposing that such tryings were possible, according to the simple analysis, an agent would be able to try to try only if they would try to try, if they tried to try to try, and they would only be able to try to try to try only if they would try to try to try if they tried to try to try, and so on ad infinitum; We would never end up with a complete account of what it takes to be able to try to perform some action, and hence (because on the simple conditional account being able to A requires that one be able to try to A) we would also never end up with a complete account of what it takes to be able to perform any action. Call this the trying-regress.

Unfortunately, because R-ABLE has ‘trying’ in its antecedent, it too will suffer from the trying-regress: we will be able to account for abilities to act, but never abilities to try to act. Agents obviously are able to try to do things, it was the agent’s
lack of such an ability in the classic coma case that made it so powerful a counterexample to simple conditional accounts, so it would obviously be beneficial if our account of abilities could extend to tryings as well as actions.

The solution to this problem obvious lies in replacing ‘trying’ in the antecedent of the analysis with some other pro-attitude or state, though just what might serve as a suitable replacement remains difficult to say: whatever pro attitude we replace it with, there is a danger of ending up with another regress.\footnote{Suppose for instance that we replaced ‘tried’ with ‘willed’ in our analysis, so that S is able to A iff S would A if S willed that they A’. As with trying we would be unable to develop a satisfactory account of ‘is able to will to A’: S is able to will-that-they-A iff S would will that they-A if they will that they will-that-they-A, and so forth into the regress.}

I think that a possible answer lies in David Velleman’s often overlooked theory of Action, and in closing I should like to briefly sketch how we might modify R-ABLE to avoid this problem.

According to Velleman, agents are motivated by a desire to do what makes sense; or, sometimes, are ‘guided by the aim of doing what makes sense’ or ‘the aim of doing what is intelligible in terms of motives and circumstances.’\footnote{See (Velleman, 2000): this idea of agents being guided by a motive to do what makes sense has its origins in (Velleman, 1989) in which Velleman argues that agents are driven by a desire to know what they are doing. This desire manifests in two ways: first, it motivates an agent to deliberate about what to do, and decide to do only that which they can understand (or is intelligible to them); and second, it motivates an agent not to do anything until there is something that they expect that they will do, so as not to lose self-knowledge by acting without a plan (Velleman, 1989, Chs.1-3). This idea has then been developed in most of Velleman’s later work, especially prevalent in The Possibility of Practical Reason (2000), in which the change is made to talking in terms of being guided by the aim of doing what makes sense, and How We Get Along (2009).} This motive, according to Velleman, can ‘play the role’ of the agent in our story of what happens when someone acts (Velleman, 2000, Ch.6): as agents we get involved in the process leading to our actions by deliberating about what to do—weighing up the reasons we have for acting and then throwing our weight behind the motive that we consider most justified so as to ensure that it is successful in bringing us to action—and this motive would play precisely that role in a reduction of action: motivating deliberation about...
what to do and the weighing up of our various motives so that we might do what is most intelligible, and then adding its own motivational force to whichever motive made the most sense to pursue.

Without going in to too much detail, an agent’s being driven by this ‘agential motive’ will result in their only full-bloodedly\textsuperscript{102} performing those actions that make the most sense, or are most intelligible, to them; in fact, if an agent is free, they will perform whatever action makes the most sense to them whenever doing so makes the most sense.\textsuperscript{103}

This presents us with a unique solution to the trying-regress: if we know that an agent will at least try to perform whatever action makes the most sense to them, then by replacing an agent’s ‘trying to $A$’ in our conditional with ‘$A$’s being most intelligible to the agent’, we should be able to account for abilities without falling into a regress:

\textbf{SENSE} \quad S \text{ is able to } A \text{ iff } S \text{ would } A \text{ if } A\text{-ing makes the most sense to } S

Because we know that an agent will full-bloodedly perform whatever action makes the most sense to them, we also know that if an agent fails to perform an action it is because they are unable to do so. An action’s making sense to an agent, that is, can fill in for the pro attitude that we originally included as the antecedent of our condition

\begin{itemize}
\item \textsuperscript{102} Full-blooded action is Velleman’s term for ‘action \textit{par excellence}, or action at its very best: intentional, uncompelled, full directed by the agent etc.
\item \textsuperscript{103} It is worth noting that an action’s making sense to an agent is to be construed experientially, rather than objectively or cognitively: an action might objectively make sense \textit{for an agent}, without making sense \textit{to} that agent; for instance, it might not make sense objectively for Jane to drink the glass of poison on the table as it will kill her and she wants to avoid dying; however, Jane thinks the poison is water and is extremely thirsty, so drinking it makes sense to her experientially.
\item Additionally, an agent can think that an action makes sense to them (cognitively) without its actually making sense to them (experientially): Jason might think that staying in to work makes the most sense for him, given that his dissertation is due tomorrow; however, the possibility of seeing old friends actually makes going out more intelligible to him than he realises and what he \textit{experiences} as making more sense to him is going out; so, if Jason acts full-bloodedly at all, he will akratically go out with his friends.
\end{itemize}
because we know that if an agent is operating as an agent, they will be guided by the motive of doing what makes sense to initiate a trying in accordance with what makes the most sense to them.

There is no longer a threat regress with SENSE, either: something’s being intelligible to you is not an action, it is a state; it is not something you do, but something you feel or experience; it is not therefore, something that an agent has the even the potential to have an ability to do: we might say that an agent $S$ is able to find some action $A$ intelligible, but we mean by this that it is possible for $S$ to find $A$ intelligible, not that they have the agential level ability to find it intelligible—to perform the finding of A-ing intelligible.

Of course, this assumes that something like Velleman’s theory of agential motivation is correct, and I have not provided a great deal of reasoning here to justify our thinking that it is, and nor can I provide here a sustained defence of Velleman’s thesis here; however, I do think that such a defence is possible.

If this sketch is right, and I believe it to be promising, motivating and defending a Velleman-style theory of agency and then finalising this account of action constitute the next steps in moving towards a theory of control.
BIBLIOGRAPHY


   New York: St. martin's Press.


   Cambridge, MA: Blackwell.


Fischer, J. M. (2002). Frankfurt-style examples, responsibility and semi-
compatibilism. In R. H. Kane, *The Oxford Handbook of Free Will*. Oxford:
Blackwell.
145-177.
Campbell, M. O’Rourke, & H. Silverstein (Eds.), *Action, Ethics, and
117-143.
Research*, 61(2), 467-480.
Press.
Philosophy*, 68(1), 5-20.
157-162.
Find Frankfurt's Argument Convincing. *Philosophical Perspectives*, 10, 403-
17.
Heinaman, R. (1986). Incompatibilism without the Principle of Alternate
63.
Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and
Philosophy*, 1(3), 337-355.


