

J. Lerm

Moral reasons of our own

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Philosophy

UNIVERSITY OF CAPE TOWN

September 2015

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

*For Ross,
still my favourite second person,
the second time around.*

Thank you to my team of supervisors: Professor Thaddeus Metz, Dr Jeremy Wanderer, and Dr Elisa Galgut, with special thanks to Jeremy, for first introducing me to Darwall's endlessly fascinating work, and to Thad, without whose mentorship this project would not have come to fruition. Thank you also to Professor Simon Beck, for being so accommodating.

Thank you to the NRF and to UCT for their funding over 2011 and 2012, and again to the NRF for the funding of 2015.

Contents

Introduction	i
Chapter 1: Darwall's Second-Personal account	1
1. Introduction	1
1.1. The terms of the debate	1
1.2. Second-personal reasons	2
1.3. Normativity and universality	4
2. Realism	6
2.1. Realism and universality	8
2.2. Realism and normativity	9
3. Hobbesian Egoism	12
3.1. Hobbesian Egoism and normativity	12
3.2. Hobbesian Egoism and universality	13
4. Moral reasons of our own	14
5. Neo-Kantianism	17
5.1. A problem with the Neo-Kantian account	18
5.2. The Generality argument	19
5.3. The Analogical argument	21
6. Second-personal accounts	24
6.1. Divine Command theory	25
6.1.1. Problems with Divine Command theory	26
6.1.2. Pufendorf's Point	29
6.2. Morality as Equal Accountability	31
6.2.1. Fichte's Point	33
6.2.2. Darwall's Second-Personal account	35
7. Conclusion	40
Chapter 2: Morality as Equal Accountability	41
1. The <i>non-sequitur</i> problem	42
1.1. The Generality Argument	44
1.2. The Analogical Argument	46
1.3. Pufendorf's Point	50
1.4. The Divine Command theory advantage	54
2. The sceptical problem	58
2.1. Unfounded assumptions	60
2.2. Second-personal intuitions	61
2.3. Validity	65
2.3.1. Validity as grounding in independent reasons	66
2.3.2. Validity as authority	67
2.3.3. Validity as grounding in the Second-Personal standpoint	70
2.3.3.1. The Second-Personal standpoint as contract	74

2.3.3.2.	Problems with the Second-Personal standpoint as contract	75
2.3.3.2.1.	Contractarianism	76
2.3.3.2.2.	Contractualism	77
3.	Conclusion	84
Chapter 3: Mill's Point		86
1.	The "reproaches of his own conscience" objection	92
1.1.	The by-and-large objection	94
1.2.	Restricting the range of moral obligations	95
1.3.	Second-personality as a version of first-personality	97
1.4.	Loosening the accountability-obligation connection	101
2.	The moral community objection	102
2.1.	Discordance	105
2.1.1.	Deriving individual authority from representative authority	106
2.1.2.	Forgiveness as special authority	112
3.	Conclusion	116
Chapter 4: The Sally/Barry counterexample		117
1.	A counterexample to Darwall's Mill's-Point argument	118
2.	The first line of defence: loosening the accountability- obligation connection	119
2.1.	Halving the looseness	121
2.2.	Removing the looseness	123
2.3.	Tightening the connection	125
3.	The second line of defence: rejecting the counterexample	128
3.1.	Relevant descriptions	129
3.2.	Is there a general obligation to keep promises?	131
3.2.1.	A Realist account of the general obligation to keep promises	131
3.2.2.	A Contractualist account of the general obligation to keep promises	132
3.3.	Pluralized address	138
3.4.	Impersonal reactive attitudes	145
4.	Conclusion	153
Chapter 5: The Realist objection		155
1.	Arguments for Realism	156
1.1.	Realist validity	157
1.2.	Realist accountability	161
1.2.1.	An account of accountability	163
1.2.2.	Being "in a position"	166
1.2.3.	Pufendorf's Point	168
1.2.4.	Forgiveness	170
1.3.	Realist symmetry	179
2.	Arguments against Realism	183
2.1.	Reasons of the wrong kind, reasons of the right kind	183
2.1.1.	Response	186

2.2. Realist explanation	191
2.2.1. Response	191
3. Conclusion	196
4. Summary	197
Chapter 6: Does Moral Philosophy rest on a category-mistake?	201
1. 'Does Moral Philosophy rest on a mistake?'	204
2. Category-mistakes	206
3. Realism and Cartesian Dualism	208
3.1. Dualisms	209
3.2. Epistemological model	210
3.3. Analogous criticisms	212
3.4. Inferential knowledge	214
3.5. Origins	215
4. Diagnosing a category-mistake	219
4.1. Cartesian Dualism's category-mistake	219
4.2. Realism's category-mistake	221
4.2.1. The problem of moral responsibility	221
4.2.2. The internalization defence	225
4.2.3. The Realist dilemma	227
4.2.4. Diagnosis	232
5. Ways of behaving	235
Chapter 7: A different approach	241
1. Fichte's argument	243
2. Towards a different approach	245
2.1. The second-personal reason vs. the summons	246
2.2. Morality as a way of relating to people	249
3. Darwall's Second-Personal account	252
3.1. The Second-Personal account as an alternative to Realism	252
3.2. Pufendorf's Point	257
3.2.1. Realist accountability	260
3.2.2. Morality as Equal Accountability	262
3.2.3. Addressing the amoralist's challenge	263
3.3. Mill's Point	267
4. Conclusion	272
References	274

Introduction

A person who deems it more important to avoid the scratching of his little finger than the destruction of the whole world makes a mistake.¹ If, by some strange turn of events, a person were faced with the choice between (1) avoiding scratching his little finger and allowing the destruction of the whole world, and (2) scratching his little finger and avoiding the destruction of the whole world, he would *have to* choose (2). But, why? Who says?

The force of our imagined agent's *having to* opt to save the world is standardly explained by the claim that it would be *immoral* of him to do otherwise. That is, he has a *moral reason* to avoid the destruction of the world, even if it costs him a scratch to his little finger – indeed, even if it costs him much more than this. And, this moral reason *obligates* him; he is *bound* to save the world at the cost of his little finger – not in the sense that he will actually, as a matter of descriptive fact, choose the salvation of the world over the integrity of his little finger, but in the sense that, were he to choose otherwise and sacrifice the world for the sake of his little finger, he would choose *wrongly*, and would subsequently be liable to criticism.

And, notice that this all goes for *anyone*. Although I have described our imagined agent in the masculine, 'he' may be man or woman, young or old, rich or poor. If Sally from up the road would have a moral reason to save the whole world, even at the cost of her little finger, then so would Shaka Zulu, as would Idi Amin, as would Mother Teresa, *etc.*

But, what *is* a moral reason? And, what is it about moral reasons that explains their power over our lives? What is the nature of moral reasons such that they're able to bind us – even bind us into doing things we don't want to do – and, moreover, bind us all equally? These are the questions I aim to address in this project, with particular reference to Stephen Darwall's *The Second-Person Standpoint: Morality, Respect and Accountability*.²

Since the publication of *The Second-Person Standpoint*, the book has received high praise indeed, being described as "exciting",³ "fascinating",⁴ "fecund",⁵ "a contribution of the first importance [...] to moral philosophy",⁶ and even as having "a certain majesty".⁷ Almost all reviewers agree that *The*

¹ This example is taken from Hume 1739, Bk.II, s.3.

² Darwall 2006.

³ Kar 2007, p.881; and Brewer 2008, p.163.

⁴ Lavin 2008.

⁵ Kar 2007, p.881; and Brewer 2008, p.163.

⁶ Yaffe 2010, p.252.

Second-Person Standpoint achieves something “of great significance”,⁸ with this commendation frequently being detailed in terms of the deep and novel challenge the book presents to contemporary metaethics; for instance, here is Samuel Fleischacker:

[*The Second-Person Standpoint*] offers a thorough, intriguing and deeply thoughtful reconception of the entire sphere of morality. And what it has to say is quite new. More than any recent book, [...] Darwall’s book challenges the [metaethical] presuppositions of practically every going moral theory.⁹

I agree with Fleischacker, and many others like him:¹⁰ *The Second-Person Standpoint* is indeed a seminal work. However, almost all reviewers also remark that *The Second-Person Standpoint* suffers serious flaws – and here I agree with (most of) those reviewers also. There is, however, little convergence on what the real problem with *The Second-Person Standpoint* is. The sustained criticism of this present project will be that Darwall fails (on various counts) to give us good grounds on which to accept that there are indeed such things as second-personal reasons in the first place. In later chapters of the projects, I will also mount the novel criticism, inspired by Gilbert Ryle, that (the most natural reading of) *The Second-Person Standpoint* is “category-mistaken”.

Darwall, however, seems to be unmoved by any of the criticisms he has so far received, declaring: “I don’t think any of the criticisms I’ve gotten have made me want to change anything fundamental”.¹¹ Indeed, he has recently gone on to publish a further – not one, but – two books, *Morality, Authority, & Law* and *Honor, History, & Relationship*, both of which build on the groundwork laid by *The Second-Person Standpoint*. The objective of those two more recent books is to demonstrate that there is a “wide range of moral phenomena that the [S]econd-[P]erson[al] standpoint illuminates”,¹² that those issues and ideas alluded to in those two new books’ titles are best understood through a Second-Personal lens. Darwall clearly thinks that the Second-Personal standpoint can do a lot of work for us. Similarly, one major motivation for *The Second-Person Standpoint* appears to be the vindication it offers of Contractualism,¹³ which Darwall is well-known for defending in earlier work. However, both of Darwall’s very recent books, as well as his earlier work on Contractualism, will largely fall beyond the scope of this project, since both depend in some sense on the groundwork laid by *The Second-Person Standpoint*. And, my current, primary interest is in the more fundamental questions of what the Second-Personal standpoint actually is and, concomitantly, why we should

⁷ Fleischacker 2009, p.118.

⁸ Lavin 2008.

⁹ Fleischacker 2009, p.117.

¹⁰ For similar terms of praise, see also Kar 2007, p.881, Korsgaard 2007, p.9, and Brewer 2008, p.166.

¹¹ Darwall 2009, p137.

¹² Fleischacker 2009, p.118 (emphasis removed). (Although, Fleischacker actually makes this remark with reference to *The Second-Person Standpoint*.)

¹³ Darwall 2006a.

think that there is any such thing as it – rather than in any further, advantageous applications of the metaethic that there might be.

My aim in this project, therefore, is to give an account of Second-Personality by which the virtues of its account of the nature of moral reasons are brought into relief, and by which it is clear that *The Second-Person Standpoint* achieves something “of great significance”.¹⁴ I make a start on this in Chapter 1, where I explain how the Second-Personal account achieves what other extant metaethics lack and how it overcomes problems those other metaethics suffer. In particular, I make the novel claim that Darwall’s Second-Personal metaethic presents an account of moral reasons as intrinsic to persons – or, as ‘moral reasons of our own’, from which this project takes its title.

Then, for the following three chapters, Chapters 2-5, I expound my own critique of *The Second-Person Standpoint*, in the light of the positive account I gave of it in Chapter 1. There, I canvass some of the criticisms of Darwall’s work suggested by other philosophers, such as Gary Watson’s¹⁵ convincing critique that the premises of Darwall’s argument for ‘Morality as Equal Accountability’ really don’t entail his desired conclusion. Most of the content of Chapter 2, however, is original argumentation in support of the criticism that Darwall gives us no good grounds on which to accept that there are indeed such things as second-personal reasons. In particular, I argue in Chapter 2 that any intuitions we have in support of the idea that there are such things as second-personal reasons are constrained by the concomitant intuitions that such second-personal reasons, if there are such things, must be *valid* – yet Darwall’s Second-Personal view is unable to supply an acceptable account of what that validity might consist in. Then, over the course of Chapters 3 and 4, I examine one of Darwall’s arguments that has, so far, received little explicit attention in the literature; this argument I call Darwall’s ‘Mill’s Point’ argument, and it is an argument for the conclusion that there are indeed such things as second-personal reasons because we can infer their existence as the best explanation of our practices of holding each other accountable. However, I contend that our practices of holding each other accountable – where this includes holding ourselves accountable and being held accountable by third-person observers who are not the victims of our misdeeds – in fact are not best explained by the second-personal nature of moral reasons. In particular, as I go on to argue in Chapter 5, such practices of holding each other accountable, as well as several other moral phenomena, are in fact better explained by the nature of moral reasons as posited by Realism, not by Darwall’s Second-Personal account. As such, despite an extremely promising start, I argue that we are ultimately without any good grounds to believe that there are in fact such things as second-personal reasons in the first place.

¹⁴ Lavin 2008, as cited above.

¹⁵ Watson 2007.

Then, in Chapter 6, the project undergoes a shift of focus, where I attempt what might be considered a critique of my critique. I turn my attention somewhat away from Metaethics only, and towards the Philosophy of Mind. I see something of a structural similarity between the debates of these two fields, and I try to apply the lessons of the Philosophy of Mind to Metaethics. In particular, I try to apply Ryle's category-mistake critique of Cartesian Dualism to traditional Metaethics, construing Darwall's Second-Personal account as the category-*unmistaken* alternative to traditional accounts of the nature of moral reasons. This entirely novel, and controversial reading of *The Second-Person Standpoint* I develop in the final chapter, Chapter 7, where I argue that it is the book's lingering, implicit commitment to a category-mistaken approach that accounts for the many and various criticisms it has received. This Ryle-inspired reading may not be one that Darwall himself would embrace, but it represents an interesting, original way of salvaging the account from the preceding critique. Furthermore, I think this reading of *The Second-Person Standpoint* is the best way to see it as "a genuine and important advance in moral philosophy",¹⁶ as it has otherwise been praised.

I begin by laying out Darwall's Second-Personal account, and describing the many ways in which it promises to be the best account of moral reasons to date – where several of the advantages I describe are novel and are not themselves touted as virtues of the account even by Darwall himself. I do this by way of contrasting the Second-Personal account with various other extant metaethics. This is the objective of Chapter 1.

¹⁶ Korsgaard 2007, p.9.

1.

Darwall's Second-Personal account¹

1. Introduction

Let us begin to address the question of the nature of a moral reason by first addressing the question, What is a reason generally? Minimally, a reason is a consideration in favour of something;² since my present interest is in practical reasons, this will be a consideration in favour of performing some action (where this covers abstaining from performing another action). This very basic definition of a reason leaves open the question of a reason's defeasibility; that is, the definition allows that if an agent has a reason to Φ as well as a reason to not- Φ , (a) his/her reason to Φ may be entirely defeated by the greater reason to not- Φ , or (b) his/her reason to Φ may simply be trumped but not defeated by the greater reason to not- Φ . It also leaves open the question of (motivational) internalism/externalism; that is, the definition allows that if an agent has a reason to Φ , his/her having a reason entails that (c) he/she is at least minimally (defeasibly) motivated to Φ , or that (d) the agent may be subject to a reason to Φ and yet have no motivation to Φ at all. What even such a basic definition does cover, though, is that reasons are *normative*: if an agent has a(n) (undefeated) reason to Φ and he/she fails to Φ , he/she has done something *wrong*.

1.1. The terms of the debate

Anything more substantial than the above is difficult to say of the nature of reasons without committing oneself to either the 'agent-relative' or the 'agent-neutral' side of one fundamental categorization of reasons. According to Thomas Nagel's definition of this popular distinction, an agent-relative reason is one that "includes[s] an essential reference to the person who has it",³ whereas an agent-neutral reason is one "which does not include an essential reference to the person who has it".⁴ As will become clear, this agent-relative/agent-neutral distinction will be important to grasp the significance of Darwall's Second-Personal account. See the following explanation, provided by Darwall in an interview, of the evolution of his Second-Personal account:

¹ S.2ff of this chapter is the very expanded first half of Lerm 2012.

² Darwall 2009, p.121, referencing Scanlon 1998.

³ Nagel 1986, p.153, following Parfit 1984.

⁴ Nagel 1986, p.152, following Parfit 1984.

What was on the table back in the early seventies, was the question of whether – I’ll use Nagel’s *initial* terms – whether there could be some reasons that were irreducibly “subjective” or whether all reasons must be “objective.” Well those terms got dropped pretty quickly in [favour] of “agent-[centred]” versus “agent-neutral”; then “agent-[centred]” got dropped in [favour] of “agent-relative.” Really, what was at issue there was a discussion whether reasons could be irreducibly egocentric, I think. Or whether they had to be appreciated from some non-egocentric point of view [...]. Anyway, what I didn’t see then, [...] was that there is this other category of irreducibly *second-personal* reasons. If there are irreducibly egocentric reasons, then those are first-personal reasons, and if there are irreducibly objective or impersonal reasons, then those are third-personal reasons. What I came to see was that there is a whole class of reasons that are *second-personal* [...].⁵

That is, Darwall aims to draw our attention to a class of reasons that are neither purely agent-relative nor agent-neutral: second-personal reasons. Insofar as the traditional binary distinction between ‘agent-relative’ and ‘agent-neutral’ reasons obscures the possibility of second-personal reasons, therefore, that traditional terminology is, at best, unhelpful, and, at worst, damaging.⁶ As such, rather than the literature’s standard ‘agent-relative’ and ‘agent-neutral’, I will opt to use alternative terminology. In *The Second-Person Standpoint*, Darwall uses as an alternative for ‘agent-neutral’ ‘state-of-the-world-regarding’;⁷ so, following Darwall’s lead, I will, for short, speak not of ‘agent-neutral’ reasons but of *world-regarding* reasons. And, I will speak not of ‘agent-relative’ reasons but of *person-regarding* reasons, where the definitions of ‘person-regarding’ and ‘world-regarding’ align (respectively) with Nagel’s definitions of ‘agent-relative’ and ‘agent-neutral’.⁸ Furthermore, person-regarding reasons may, in turn, be divided into either *first-personal* or *second-personal* reasons, where first-personal reasons are, as Darwall describes them (above), “irreducibly egocentric”, and second-personal reasons are person-regarding reasons that are not irreducibly egocentric. This taxonomy of reasons, therefore, is as follows:

World-regarding	Person-regarding	
	First-personal	Second-personal

1.2. Second-personal reasons

More should be said about the nature of second-personal reasons, beyond the purely negative definition of them proffered above. In the opening pages of *The Second-Person Standpoint*, Darwall characterizes the second-personal reason as follows. The reader is

⁵ Darwall 2009, pp.122-123.

⁶ Christine M. Korsgaard (1993/1996) has proffered a persuasive critique of the ‘agent-relative’/‘agent-neutral’ terminology, and I have elsewhere tried to make an even more compelling case by expanding upon her argument (Lerm 2013).

⁷ Darwall 2006, p.6.

⁸ Nagel 1986, pp.152-153 (following Parfit 1984), as cited above, and as echoed by Darwall 2009, pp.122-123, cited above.

asked to imagine a scenario in which someone, the 'foot-treader', is standing on your foot. According to Darwall, there are two ways you could respond to the foot-treader: by invoking a world-regarding reason, or by invoking a second-personal reason, and the contrast between these two possible responses will bring into relief the special nature of the second-personal reason. So, firstly, you could:

get [the foot-treader] to feel sympathetic concern for you in your plight, thereby leading him to want you to be free of pain. Were he to have this desire, he would see your being in pain as a bad thing, a state of the world that there is reason for him (or, indeed, for anyone who is able) to change. And he would most naturally see his desire that you be pain-free not as the source of this reason, but as a form of access to a reason that is there anyway. In desiring that you be free of pain, he would see this possible state of affairs as a better way for the world to be, as a possible outcome or state that, as Moore put it, "ought to exist for its own sake" [...]. Were he to credit the way things seem from the perspective of his desire, he would accept a *state-of-the-world-regarding* and *agent-neutral* reason for removing his foot. The reason would not be essentially for him as the agent causing another person pain. It would exist, more fundamentally, for anyone who is in a position to effect your relief and therefore for him, since he is well placed to do so.⁹

That is, one way you might respond to the foot-treader is by drawing his attention to the world-regarding reason to not cause pain, a reason "that is there anyway" (whether he sees it or not) and which you have to get him to notice. Or, says Darwall, you might take quite a different approach:

Alternatively, you might lay a claim or address a purportedly valid demand. You might say something that asserts or implies your authority to claim or demand that he move his foot and that simultaneously expresses this demand. You might demand this as the person whose foot he is stepping on, or as a member of the moral community, whose members understand themselves as demanding that people not step on one another's feet, or as both. Whichever, the reason you would address would be *agent-relative* rather than agent-neutral. It would concern, most fundamentally, his relations to others (and himself) viewed from his perspective within those relations, in this case, that his keeping his foot on yours causes another person pain, causes inconvenience, and so on. The reason would not be addressed to him as someone who is simply in a position to alter the regrettable state of someone's pain or of someone's causing another pain [...]. It would be addressed to him, rather, as the person causing gratuitous pain to another person, something we normally assume we have the authority to demand that persons not do to one another.¹⁰

That is, another way you might respond to the foot-treader is by addressing a *second-personal* reason to him, where second-personal reasons are not reasons that are "there anyway",¹¹ but which, rather, "give a person a distinctive kind of (normative) reason for

⁹ Darwall 2006, pp.5-6, referencing Moore 1903/1993, p.34.

¹⁰ Darwall 2006, p.7 (emphasis added).

See also Darwall 2010a, where the foot-treader scenario is re-presented.

¹¹ Darwall 2006, p.5; as cited above.

acting”,¹² one which is “*agent-relative* rather than agent-neutral”. The crucial point to note here, however, is that a second-personal reason is not a standard agent-relative reason. It is not simply one which, to quote Nagel’s definition again, “includes[s] an essential reference to the person who has it”¹³ – where “the person” is one, singular person; where a reason includes essential reference to just one person, it is a *first-personal* reason. *Second-personal* reasons, on the other hand, are special kinds of person-regarding reasons which include essential reference to *two* people: the addressee (such as the foot-treader in the above scenario) and the addressor (you, in the above scenario). As Darwall puts it, “second personal reasons always derive from agents’ relations to one another.”¹⁴ In later papers, Darwall sometimes adopts Michael Thompson’s terminology of “bipolar obligations”,¹⁵ and I think the term is helpful: just as electricity is *bipolar* in the sense that it can exist only when there are two poles between which current flows, second-personal reasons are reasons existing between two persons when there is a “flow” between the addressor and the addressee.

One further, significant point to note of second-personal reasons is that, according to Darwall, “moral obligation is essentially second-personal”.¹⁶ If this is the case, then Darwall will have changed the face of metaethics forever. Metaethicists have, since the early 1970s at least, believed that reasons are either agent-neutral (world-regarding) or agent-relative (person-regarding) – where the latter is taken to include only *first-personal* reasons. If, therefore, Darwall is able to demonstrate that this old distinction actually obscures a third class of reasons – second-personal reasons – and that one monumental kind of reason – *moral reasons* – are neither first-personal nor world-regarding, but *second-personal*, then this will constitute a truly game-changing achievement.

1.3. Normativity and universality

But, first, what is a moral reason? *Qua* (practical) reason, a moral reason is a normatively forceful consideration in favour of some action.¹⁷ *Qua* moral, however, a moral reason has certain characteristics that distinguish it from its counterpart, non-moral reasons.¹⁸ For instance, over and above normativity, John Cottingham, lists universality, objectivity and

¹² Darwall 2006, p.4 (emphasis removed).

¹³ Nagel 1986, p.153, following Parfit 1984.

¹⁴ Darwall 2006, p.8.

¹⁵ Darwall 2010b, p.1, and 2010c, citing Thompson 2004.

¹⁶ Darwall 2006, p.65.

¹⁷ See s.1, above.

¹⁸ See Darwall 2006, pp.26-27, for his criteria of the distinguishing features of moral reasons.

necessity as criteria of moral reasons.¹⁹ My present discussion will focus on only one of these criteria – the least contentious of them: universality.

According to Darwall, it is “uncontroversial” “that fundamental moral obligations apply equally to all moral agents.”²⁰ Moral reasons are universal. Universality alone will, of course, not be a *sufficient* criterion of moral reasons, because it may well be that some reasons (like, perhaps, the reason to look after one’s health) are universal but are nevertheless non-moral. Universality is, however, surely a *necessary* criterion of moral reasons. As Thaddeus Metz puts it, “For example, it must be true of any person in any place and at any time that it would be wrong for her to torture a baby for fun”,²¹ given that it is a *moral* injunction that one not torture babies for fun. The guiding idea here is that nobody is above morality.²² Morality isn’t just for the soft-hearted or the oppressed or the under-evolved. Nobody, neither the callous nor the powerful nor anybody else, is exempt from morality. As Pope John Paul II puts the point:

*When it is a matter of the moral norms prohibiting intrinsic evil, there are no privileges or exceptions for anyone. It makes no difference whether one is the master of the world or the "poorest of the poor" on the face of the earth. Before the demands of morality we are all absolutely equal.*²³

Morality is universal. If I have a moral reason, say, not to cause undue harm to anyone, then it follows that you too have a moral reason not to cause undue harm to anyone.

My argument of this chapter will be that Darwall’s Second-Personal account is, on the face of it, able to do a much better job than competing metaethics of accounting for both moral reasons’ normativity and their universality. My strategy will effectively be the same as the one Darwall himself employs in *The Second-Person Standpoint*, which is to present Darwall’s Second-Personal account as something like the culmination of various other traditional metaethics that have come before it, retaining what those other metaethics get right, and overcoming what those other metaethics get wrong. My contribution of this chapter is that I present Darwall’s Second-Personal account as the culmination of several more traditional

¹⁹ Cottingham 2005, as referenced by Metz 2013, pp.85-86.

²⁰ Darwall 2010c, p.40 (emphasis removed).

²¹ Metz 2013, p.85.

²² Although, some individuals may be – so to speak – ‘below’ morality. This is because “there exists a set of capacities, which we collect under the term ‘moral agency’, that are both necessary and sufficient to make one subject to moral obligations. Not every human being has these capacities,” and such individuals will not be subject to moral obligation. However, “it is uncontroversial that normal adult human beings do” possess the necessary capacities and so will be subject to moral obligation. (Darwall 2010c, p.40)

²³ Pope John Paul II 1993, s.96

metaethics than he does himself, and as achieving the correct account of moral reasons in ways that Darwall does not himself take into account. I organize the relevant competing metaethics in the following table:

World-regarding <i>Realism</i>	Person-regarding	
	First-personal <i>Hobbesian Egoism</i> <i>Neo-Kantianism</i>	Second-personal <i>Divine Command theory</i> <i>Darwall's Second-Personal account</i>

As I read it, Darwall's Second-Personal account responds to the need for an account of morality as 'intrinsic' to persons and of moral reasons as reasons 'of our own', as this project's title intimates. I go on to demonstrate this by contrasting Darwall's Second-Personal account with world-regarding Realism, with first-personal Hobbesian Egoism and Neo-Kantianism, and with second-personal Divine Command theory. As I will argue, non-second-personal metaethics struggle to account for both the normativity of moral reasons and their universality. Whereas a world-regarding metaethic (Realism) provides a neat account of the universality of moral reasons, it struggles to explain their having normative force over us. And, whereas first-personal metaethics (Hobbesian Egoism and Neo-Kantianism) provide a straightforward account of moral reasons' normative force over us, they struggle to explain the universality of moral reasons. By contrast, I will argue that Darwall's Second-Personal account promises to give us the best of both worlds, and that it does so because it construes moral reasons as reasons of our own. I begin, as Darwall himself does in the opening pages of *The Second-Person Standpoint*,²⁴ with world-regarding accounts.

2. Realism

As Korsgaard notes, "Realism" is something of a confusing label.²⁵ Let me stipulate, therefore, that I have in mind as a Realist the sort of theorist who would recommend the former kind of response to the foot-treader, as described by Darwall (above); that is, one who "would accept a state-of-the-world-regarding and agent-neutral reason for removing his foot", "a reason that is there anyway" for removing his foot from on top of yours.²⁶ This is enough to meet the broad definition of Realism as that metaethic which holds that there are moral facts independently of any person's attitudes towards those moral facts. But note, however,

²⁴ Darwall 2006, pp.5-6.

²⁵ Korsgaard 1996, p.34.

Darwall (2006, pp.235, fn.34, and 292), drawing on Korsgaard (1996), makes some similar remarks regarding the various forms of Realism.

²⁶ Darwall 2006, pp.5-6 (emphases removed), as cited in s.1.2 above.

that Darwall goes on to describe the former response to the foot-treader as follows: upon your complaint, the foot-treader “would see your being in pain as a bad thing, a state of the world that there is reason for him (or, indeed, for anyone who is able) to change” and would see your being free of pain as “as a better way for the world to be”.²⁷ These additional details suggest that Darwall does not have in mind here a Realist in the broadest sense, but instead in the somewhat narrower sense of the *Consequentialist* Realist in particular. This interpretation is bolstered by the fact that Darwall takes as his prime adversary in *The Second-Person Standpoint* G.E. Moore,²⁸ who is famously a Consequentialist Realist.²⁹ Indeed, even in his early characterization of the former response to the foot-treader, Darwall describes the foot-treader as seeing “this possible state of affairs as a better way for the world to be, as a possible outcome or state that, *as Moore put it*, “ought to exist for its own sake” [...].”³⁰ Thus, as a “Realist” I have in mind here the sort of Consequentialist Realist that would recommend the former kind of response in Darwall’s foot-treader scenario.

I note these points because my particular use of “Realism” throughout this project is, admittedly, somewhat misleading, for two reasons. Firstly, such a use of the term clearly excludes *Non-Consequentialist* versions of Realism.³¹ However, it is a fact, as far as I know, that *most* Realisms will indeed be Consequentialist;³² Non-Consequentialist Realisms are atypical. As Consequentialist Realist Richard Boyd notes (with reference to naturalist Realists in particular): “In a certain sense, some version of [C]onsequentialism seems to be the *natural* position for naturalist moral [R]ealists.”³³ As such, for the purposes of contrasting Darwall’s Second-Personal account with Realism, as is the aim of this chapter,³⁴ it is sensible to draw the comparison with the major variety of Realism, namely, Consequentialist Realism. (And, I take it, this is precisely why Darwall himself restricts his focus to Consequentialist Realism too.) Having said all this, however, I do hope that the

²⁷ Darwall 2006, pp.5-6; as cited in s.1.2, above.

²⁸ Darwall 2006, throughout: pp.5, 28, 37, 158 (fn.17), 217 (fn.5), 221, 226–227, 231, 280, 282–283, and 286.

²⁹ Moore 1903.

³⁰ Darwall 2006, pp.5-6 (emphasis added), referencing Moore 1903/1993, p.34.

³¹ Such as those of, for instance, Prichard 1912, or, much more recently, Metz 2013, Ch.5.

³² Here Boyd (2003, p.505) agrees with me, as does Michael S. Moore (2010, p.41, referencing Sturgeon).

For some examples of Consequentialist Realists see Sidgwick 1874/1907 (as referenced by Phillips 2011, p.39, and Hurka 2014, p.87), Moore 1903, Russell 1951 (as referenced by Hurka 2014, p.87) and, more recently, Sturgeon 1985 and 2003, Railton 1986, Boyd 1988 and 2003, Brink 1989, and, somewhat controversially, Parfit 2011.

³³ Boyd 2003, p.506.

³⁴ See s.1.3, above.

points I go on to make about “Realism” do in many cases have broader applicability than just to Consequentialist Realism in particular.

The second way in which my use of “Realism” is potentially misleading, however, is that Darwall does not himself describe the metaethic against which he contrasts his Second-Personal account as “Realist”. As such, I am effectively putting words – or, at least, *a* word – into Darwall’s mouth; but I take this word to be necessary insofar as the practicality of the present discussion demands a *noun*. Darwall proffers only the terms “agent-neutral”,³⁵ which is problematic for reasons described above,³⁶ and “state-of-the-world-regarding”,³⁷ which is anything but efficient and which I have already abbreviated to “world-regarding”³⁸ – and, besides, both of which are merely adjectives. As such, I adopt “Realism” as the noun-form of that which Darwall’s adjectives describe in the former response to the foot-treader. Although it is not Darwall’s own term, the context and his subsequent discussion make it clear that it is indeed (Consequentialist) Realism that he had in mind.

2.1. Realism and universality

We have noted that moral reasons are universal.³⁹ If I have a moral reason, say, not to cause undue harm to anyone, then it follows that you too have a moral reason not to cause undue harm to anyone. But, contrast this universal, moral reason with the *non*-moral reason to, say, complete this dissertation. This is a reason that I am subject to, but it doesn’t on that basis follow that you too are subject to such a reason. It is perfectly possible – in fact, quite likely – that, even though I have a reason to complete a dissertation, you do not. That is, the reason to complete a dissertation is – unlike moral reasons – non-universal.

Now, there seems to be an obvious explanation of the non-universality of non-moral reasons: I have a reason to complete a dissertation because I want to get my PhD; on the other hand, you do not desire (no longer desire) to get your PhD, therefore you have no reason to complete a dissertation. That is, the non-universality of the reason to complete a dissertation is apparently explained by differences between/amongst agents. Of course, I mean here: *relevant* differences between/amongst agents; the facts that I am in Cape Town and you are not don’t explain why I have a reason to complete a dissertation but you don’t, since facts about geographic location aren’t relevant to reasons about dissertations. What

³⁵ Darwall 2006, p.6, as cited in s.1.2, above.

³⁶ See s.1.1, above.

³⁷ Darwall 2006, p.6, as cited in s.1.2, above.

³⁸ See s.1.1, above.

³⁹ See s.1.3, above.

is, however, relevant to reasons about dissertations are facts about desiring to obtain PhDs. Since our desires differ on this score, so do our reasons.

So, non-universal, non-moral reasons apparently depend on features of the person's self, like their desires. That is, they are *person-regarding*.⁴⁰ And, because different selves have different features, like different desires, they will be subject to different non-moral reasons – that is, their non-moral reasons will be non-universal. On the other hand, it seems that moral reasons must be universal because they *don't* depend on features of a person's self. On what, then, do they depend? What gives rise to universal moral reasons? The answer has to be something that will account for the universality of moral reasons, so, arguably, something objective rather than subjective. The natural candidate for what gives rise to moral reasons, therefore, is: the world. Where non-universal, non-moral reasons are person-regarding, universal, moral reasons must be *world-regarding*.⁴¹ Moral reasons must apparently be, or be derivable from, Real states of the objective world. For instance, the moral reason not to cause undue harm obtains universally in the same way that 'Water is H₂O' is universally true: because the world really is such that water is H₂O.⁴² The universality of moral reasons, it seems, is best accounted for by their Real, world-regarding nature.

2.2. Realism and normativity

However, despite the ease with which Realism is able to account for the universality of moral reasons, it faces some difficulties. In particular, it faces the problem of how world-regarding moral reasons could ever come to have any *grip* on us, any normative force over us. To bring the problem into view, contrast the example of the non-moral, non-universal reason to complete a dissertation with that of a moral, universal reason not to cause undue harm. It's clear (or, at least, clearer) how my person-regarding reason to complete my dissertation comes to have normative force over me: it's the product of *a part of me*; namely, my desire to obtain my PhD. That's why I have to do it. And, if this feature of the self – the desire to obtain one's PhD – is not shared by *you*, then the reason to complete your dissertation will fail to have any grip on *you*. And, this is exactly as we think it should be: you *don't* have a reason to write a dissertation if you don't want to obtain your PhD (anymore). So, if you don't have a reason to complete a dissertation in virtue of your not (no longer) desiring to obtain a PhD, how could it be that you could have a *moral* reason not to cause undue harm – independently of any feature of your self (like your desires)? To put

⁴⁰ See s.1.1., above.

⁴¹ See s.1.1., above.

⁴² Metz 2013, p.92.

it another way, if the normative force of our non-moral reasons (like my reason to complete my dissertation) derives from our desires – that is, person-regardingly – then how could it be that moral reasons have any normative force over us, if they are *world-regarding*, as Realism claims?

For a more detailed criticism of basically the same problem, consider Christine Korsgaard's critique of Realism. Korsgaard's critique is that Realism fails to address 'the normative question'. Korsgaard's normative question is just what is otherwise known as 'the amoralist's (or 'Glaucou's') challenge':

Why should I be moral? [...] We are asking what *justifies* the claims that morality makes on us. This is what I am calling 'the normative question'.⁴³

Korsgaard's complaint is that Realism is unable to tell us why we ought to be moral – or, as I have put it, above, that Realism fails to account for morality's normative force over us.

Korsgaard's complaint is that the only kind of response Realism – being world-regarding – can give to the normative question/amoralist's challenge is: That's just the way the world is. "Why should I be moral? Why should I refrain from harming you – especially when harming you will benefit me?" The only answer available to the Realist is: just because. Because the world is such that you ought not to cause undue harm, and that's just the way the world is.

In Korsgaard's words:

You ask the normative question: you want to know whether this terrible claim on you is justified. Is it really true that this is what you *must* do? The [R]ealist's answer to this question is simply 'Yes'. That is, *all* he can say is that it is *true* that this is what you ought to do.⁴⁴

But, of course, this is no answer at all. Pointing out that it's *true* that, say, one ought not to cause undue harm isn't to address the amoralist's question at all. The amoralist's question was *why* he/she ought to be moral – not *whether* causing undue harm is moral or not.

Telling the amoralist that it's true that "The world is such that you ought not to cause undue harm" is like telling the external world sceptic that the universe came to be as the result of the 'big bang' (or, equally, that God created the world). Such a response simply doesn't address the sceptical question. When the Realist asserts that "The world is such that you ought not to cause undue harm", the amoralist can simply shrug and retort, "So what?"

Realism fails to address Korsgaard's normative question/the amoralist's challenge because merely citing facts about the world cannot explain moral reasons' normative force over us.

⁴³ Korsgaard 1996, pp.8-10.

⁴⁴ Korsgaard 1996, p.38.

Other philosophers' reactions to the apparent normative impotence of worldly facts can give us further insight into the complaint. For instance, Korsgaard cites Harry G. Frankfurt, according to whom "the normativity of morality for any given agent is contingent on whether that agent cares about morality [...]."⁴⁵ On Frankfurt's account, a person's reasons are a matter of what he/she *cares* about – which is Frankfurt's technical term for a person's reflectively identifying with her first-order desires.⁴⁶ So, according to Frankfurt,

Even when volitional necessity [i.e. normative force] arises in connection with actions which are required or forbidden by duty, it does not derive from the person's moral convictions as such but from the way in which he cares about certain things⁴⁷

That is, on Frankfurt's account, when a person is subject to moral reasons, it is not his/her "moral convictions" (i.e. his/her beliefs about moral facts) that makes it so; rather, it is that he/she *cares* about what is moral. According to Frankfurt, even if there are such things as moral facts, they only become *reasons* for an agent once he/she cares about them; without this connection to the agent's self, they have no normative force over him/her.

For a similar rationale, we should consider, as Darwall does, Bernard Williams' famous 'internal reasons thesis', "which asserts that all normative reasons for action must be anchored appropriately in the agent's own "motivational set" (be "internal reasons") [...]."⁴⁸ That is, Williams denies that there could be such things as 'external reasons', reasons which are *not* (or could not, by some 'sound deliberative route', be) anchored in the agent's motivational set. This is because – again – such external reasons would apparently lack the characteristic normative force of *reasons*, and *a fortiori*, of moral reasons. Again, the idea is that, without a connection to the agent's self via the agent's motivational set, external reasons have no normative force over the agent.⁴⁹

Thus, although Realism gives us a neat account of the universality of moral reasons, it apparently gives us an account of *external* reasons, to use Williams' phrase: reasons that, as world-regarding rather than person-regarding, have no necessary connection to our motivational sets or other features of our selves and thus have no normative force over us. Now, Realism might try to recover from this by *internalizing* otherwise external reasons – by, for instance, stipulating that the agent *desires* that the victim be pain-free, or that the agent

⁴⁵ Korsgaard 2006, pp.55-56.

⁴⁶ See, for instance, Frankfurt 2004, pp.18, 50.

⁴⁷ Frankfurt 1988, p.90.

⁴⁸ Darwall 2006, p.94, referencing Williams 1979/1981.

⁴⁹ Likewise, we might consider Darwall's own presentation of the world-regarding, Realist response to the foot-treader (2006, p.5; as cited in s.1.2, above), which, charitably, invokes such features of the foot-treader's self as his "sympathetic concern" and "desire that you be pain-free", over and above the reason for him to remove his foot from on top of yours "that is there anyway".

cares about morality.⁵⁰ But, if it's the internal state of desiring or caring that's doing the normative work in such accounts, then they are no longer *world-regarding*. Rather, such accounts would be *person-regarding*, since such internal states as desires, cares, *etc.*, are features of a person's self. As such, Realism fails: it fails to account for the normative force of moral reasons, or, when it attempts to do so, it transforms into a person-regarding metaethic instead, and, in doing so, forfeits the advantage of its world-regardingness: namely, the neat explanation of universality.⁵¹

I turn now to consider a person-regarding metaethic that takes seriously the idea of Frankfurt, Williams, and several others, that the reasons to which a person is subject must bear some connection to his/her self – in particular, to his/her desire to live a non-nasty, non-brutish and long life. That is, I turn now to consider Thomas Hobbes' Egoism.

3. Hobbesian Egoism

For the purposes of this section, I am interested in the general Hobbesian picture, and assume the reader's familiarity therewith. The basic Hobbesian argument,⁵² very briefly, then, is as follows. Life in the State of Nature is 'nasty, brutish and short'. We avoid a nasty, brutish and short life by submitting ourselves to the will of the Sovereign – which, in effect, is to act morally. So, it is in your own interest to submit yourself to the will of the Sovereign and behave morally. That is, on the Hobbesian account, moral reasons come to have their normative force over us by way of our first-personal reasons to avoid nasty, brutish and short lives.⁵³ You ought, say, not to cause undue harm to another if you want to live a nice life. And, who wouldn't want to live a nice life?

3.1. Hobbesian Egoism and normativity

In Williams' terms, our Hobbesian reasons to behave morally are, therefore, internal reasons: they derive from our motivational sets, namely, our desires to live non-nasty, non-

⁵⁰ I go on to consider such an argument in detail, in Ch.6, s.4.2.2.

⁵¹ Note, however, that in Ch.5 I return to argue in *support* of Realism.

⁵² Hobbes 1651.

⁵³ This Egoistic reading of Hobbes is not entirely uncontroversial; however, I do take it to be the natural, predominant reading thereof. Moreover, Darwall also appears to read Hobbes as an Egoist, describing Hobbes in *The Second-Person Standpoint* as a Contractarian rather than a Contractualist (2006, p.308, fn.8), and elsewhere claiming that "the idea of a self-interested agreement may well be operating at a deep level in Hobbes independently of the idea of contract" (1989, p.404) and citing with approval Jean Hampton's (1986) game-theoretical account of Hobbes according to which citizens' "reasons derive from self-interest" (p.402). Besides, even if the view I describe in this section is not uncontroversially attributable to Hobbes, it is nevertheless a conceivable position, and one usefully contrasted with Darwall's Second-Personal account, as is the aim of this chapter.

brutish and long lives. Or, as Frankfurt would put it, since we care about the non-nastiness, non-brutishness and length of our lives, we thereby have reason to be moral. Note how such an account responds to the amoralist: “Why should I be moral?” “Because your life will go better that way. Because you desire a non-nasty, non-brutish and long life, and moral living is the best way to achieve that.” So, on this Hobbesian account, moral reasons come to have normative force over us by way of our first-personal reasons to avoid nasty, brutish and short lives. And, furthermore, they come to have normative force over *all* of us insofar as we all desire non-nasty, non-brutish and long lives. So, Hobbes apparently gives us an account of both features of morality: its normative force, and its universality.

3.2. Hobbesian Egoism and universality

Hobbes’ account is first-personal in virtue of grounding a person’s reasons in his/her desires – namely, the desire to avoid a nasty, brutish and short life in the state of nature. However, what if a person fails – due maybe to some pathology – to desire a non-nasty, non-brutish and long life? While it is highly probable that most psychologically normal persons will indeed desire non-nasty, non-brutish and long lives, it is possible that some won’t. And, if, on Hobbes’ account, a person is subject to moral reasons in virtue of his/her first-personal desire to avoid a nasty, brutish and short life, then it follows that a person who fails to possess that first-personal feature – that is, who fails to desire a non-nasty, non-brutish and long life – will *not* be subject to moral reasons. If the Hobbesian response to the amoralist is: “You should be moral because you desire a non-nasty, non-brutish and long life, and moral living is the best way to achieve that,” the amoralist may simply retort: “But I *don’t* want to live a non-nasty, non-brutish and long life!” That is to say, despite initial appearances, Hobbes’ Egoism is *non*-universal rather than universal. Although it does not preclude the possibility of the universality of moral reasons, it does not entail their *necessary* universality; moral reasons are only universal on this account *if* it happens that all persons desire a non-nasty, non-brutish and long life. So, although it provides a first-personal account according to which it’s clear (or, at least, clearer than it is on the world-regarding account) how moral reasons come to have normative force over us, the account does not capture the universality of moral reasons. As such, Hobbes’ Egoism fails.

So, where world-regarding Realism offers a straightforward account of the universality of moral reasons but struggles to account for their normative force, person-regarding Hobbesian Egoism offers a straightforward account of the normative force of moral reasons but struggles to account for their universality. Different as these two metaethics are, however, I suggest that both fail to account successfully for both the normativity and

universality of moral reason because they share an essentially similar, mistaken approach to morality.

4. Moral reasons of our own

I believe that Realists⁵⁴ and Hobbesians,⁵⁵ as well as Frankfurt and Williams,⁵⁶ all take a basically similar approach to morality. Furthermore, I believe that this approach is fundamentally mistaken, and that it is this mistaken approach that explains Realism's as well as Hobbes' difficulty in accounting for both morality's normative force and its universality.

This approach, that I think is fundamentally mistaken, is to treat morality as essentially *alien* to the person. This is perhaps easiest to see in the case of Realism, according to which moral reasons are world-, rather than person-, regarding, "out there", and obtaining entirely independently of our attitudes towards them. Real, world-regarding reasons' independence of persons entails that they are essentially alien to the person. Person-regarding accounts such as Hobbes' of course deny that moral reasons exist independently of persons, but I submit that accounts like Hobbes' nevertheless approach morality in effectively the same kind of way: as essentially alien to the person, and therefore as something that first has to be 'internalized', by one means or another, before it can have any normative force over us. Morality, on this conception, is regarded with suspicion, because it is essentially conceived as something like a killjoy teacher or policeman telling us what to do, restricting, controlling and oppressing us. So, Hobbes feels he has to demonstrate that morality is *good for us*, that life under a Sovereign is better than life in the State of Nature, before it could be the case that we have reason to behave as morality demands – as if morality needs first to be *vindicated*.⁵⁷ And, if morality couldn't be vindicated – if adherence to it didn't yield a non-nasty, non-brutish and long life – it simply wouldn't have any normative force over us. Similarly, on accounts like Frankfurt's, not only will a person fail to have reasons to behave morally unless he/she cares about morality, but, because a person may even "wholeheartedly love [i.e. care about] what is evaluatively nondescript, or what is bad, or what is evil",⁵⁸ he/she may even have positive reasons for behaving *immorally*. And, at the extreme end of the spectrum, we find sceptical and nihilist accounts like Williams',⁵⁹ which,

⁵⁴ See s.2, above.

⁵⁵ See s.3, above.

⁵⁶ See s.2.2, above.

⁵⁷ Korsgaard (1996, p.132) cites Gauthier (1986) as another such account.

⁵⁸ Frankfurt 2004, p.98.

⁵⁹ See Williams 1985.

(in Darwall's words) committed to the claim that moral reasons must be internal reasons, yet finding that "this presupposition is a bit of false consciousness", presents a "version of the Nietzschean critique that morality is an enslaving ideology, a form of false consciousness that shackles and sickens".⁶⁰ What all these approaches have in common, I suggest, is their suspicious view of morality as something essentially alien and therefore as something which needs to be vindicated before it can have any grip on us, failing which it has no grip on us at all, or – worse – as something which in principle cannot be vindicated and actually does us harm.

As we have seen, while the world-regarding Realist is easily able to account for the universality of moral reasons,⁶¹ he struggles to provide a satisfactory answer to the amoralist's challenge and account for moral reasons' normative force.⁶² On the other hand, while person-regarding metaethics such as Hobbes' are better able to account for the normative force of moral reasons by linking them to pre-existing reasons of ours,⁶³ such metaethics are unable to account for the universality of moral reasons, because moral reasons may simply fail to be entailed by some agents' pre-existing reasons.⁶⁴ Both of these kinds of failure by both of these kinds of metaethics, I claim, are due to their flawed conceptions of morality: as something essentially alien to the person, something out there in the world or something standing in need of vindication before we give it admission.

It is this view of morality as essentially alien to the person that makes the amoralist's challenge so difficult to meet. On this view, asking "Why should I be moral?" is like asking "Why should I vote?", when there is more than one option as to what I could do: vote, or abstain. Of course, there are better and worse answers to the question "Why should I vote?". Answers like "You should vote for the ANC" are not very helpful, since they presuppose that the question as to *why* I should vote in the first place has already been settled and then go on to inform me for *whom* I should vote.⁶⁵ On the other hand, answers like "You should vote because it increases the chances of your preferred political party being

⁶⁰ Darwall 2006, pp.93-94, referencing Williams 1995 and Nietzsche 1887/1994. Korsgaard (1996, pp.60 and 78) describes similar worries with regard to the views of Freud (1930/1961) and Nietzsche (1887/1967).

⁶¹ See s.2.1, above.

⁶² See s.2.2, above.

⁶³ See s.3.1, above.

⁶⁴ See s.3.2, above.

⁶⁵ I mean this to be a parallel of Korsgaard's critique of Realism, discussed above (s.2.2), according to which the Realist misses the point of the amoralist's challenge by insisting in response that it's true that, say, one ought not to cause undue harm, when the amoralist's challenge is really whether there's anything one *ought* to do in the first place.

the ruling party” are better answers in that they do at least address the original question, but they are contingent on certain features of my self which may or may not obtain; for instance, my non-anarchism.⁶⁶ But, I submit that the question “Why should I be moral?” isn’t like the question “Why should I vote?” at all. Rather, asking “Why should I be moral?” is like asking “Why should I breathe?” – the answer to which is: you simply have to; because of the kind of creature you are, there just is no other way for you to be. If there’s an *I* at all, it has got to be a *breathing I*. Likewise, I will argue in this project, for morality.

“Why should I be moral?” is, like “Why should I breathe?”, a question that answers itself: if there’s an *I* at all it has got to be, I argue, an *I* subject to moral reasons. This follows when morality is conceived not as essentially alien to persons, but as essentially *intrinsic* to persons. On this different understanding, morality is less like a threatening policeman and more like the rules of musical composition: although it means that there are some things we can’t do (we can’t cause undue harm to others; we can’t have three beats in a bar when the time signature is 4/4), it is the only way we have of going about our lives – there is no alternative that the imagined policeman bars us from enjoying. Without this shaping force, our lives would be unrecognizable as persons’, as our music would be no more than noise. Morality, on this intrinsic understanding, is not an infringement or limitation on our personhood; rather, it is intrinsically bound up with our being persons at all. Morality is not put upon us, but part of us. Morality is not imposed from without, but instead comes from within. Moral reasons are reasons of our own. And, as such, there is no need for morality to be vindicated before it comes to have normative force over us. Morality applies universally to all persons because part of what it is to be a person at all is to be bound by moral reasons. If there’s an *I* at all, it has got to be an *I* subject to moral reasons – not because of the way the world is, as Realism claims⁶⁷ – but because of the way *you* are – and not because of the way you are in terms of your peculiar desires, as the Egoist claims⁶⁸ – but because of the way you are in terms of the way we all are.

This is, of course, a basically Kantian idea, as I go on to describe in the following section.

But, as I subsequently go on to explain, it is an idea that Darwall’s Second-Personal account captures even better than the Neo-Kantian account.

⁶⁶ I mean this to be a parallel of my critique of Hobbes’ and similar accounts, above (s.3.2) according to which such accounts fail to capture the universality of moral reasons because an agent may always fail to possess the relevant feature (desire, etc.) to which his moral reasons are supposed to attach.

⁶⁷ See s.2, above.

⁶⁸ See s.3, above.

5. Neo-Kantianism

As above, for the purposes of this section, I am interested in a general Kantian picture, and assume the reader's basic familiarity therewith. Very briefly, then, Kant's argument is as follows.⁶⁹ We have noted that person-regarding metaethics such as Hobbes' do a better job than world-regarding metaethics of accounting for moral reasons' normative force over us.⁷⁰ But, why is this? Why is it, as I claimed above, that a person's desire to obtain her PhD gives her a reason to write a PhD dissertation? Completing one's PhD dissertation is obviously the means to the end of obtaining a PhD, but why is it a *reason*? In other words, where does the normative force come from? Kant's answer is that the normativity comes from the person herself, because of the kind of being he/she is. And, the kind of being a person is is one for whom his/her (reflectively endorsed, or rational) desires have normative force; persons are "the sources of normativity".⁷¹ I have a reason to complete my dissertation because (I desire to obtain my PhD *and* because) of the kind of being I am – just as I have a reason to breathe because of the kind of being I am. So, in recognizing that my desire to obtain my PhD gives me a reason to write a dissertation, I must recognize that this is because of the kind of being I am. That is to say, in respecting my reasons, I must also respect my own *personhood*. Now, we get from this step to the moral law by Kant's further insistence that I must therefore respect personhood generally; that is, not only my own personhood, but also yours and all others'. And, this effectively gets us to respect for others and morality.

We should note some points of comparison between the Hobbesian and the Neo-Kantian accounts. Firstly, note that both are person-regarding rather than world-regarding. The Neo-Kantian metaethic accounts for moral reasons' having a grip on us in the same manner in which my reason to complete my PhD dissertation has a grip on me: my moral reasons are a product of a feature of my self. The difference, though, is that where Hobbes' account depends on a person's desiring a non-nasty, non-brutish and long life, the Neo-Kantian account depends on no such desire or *particular* feature of a person's self, such as his/her caring about morality or his/her motivational set. Rather, the feature of a person's self upon which the Neo-Kantian account depends is, more generally, his/her mere *personhood*. And, because, on this Neo-Kantian account, you are subject to moral reasons not in virtue of a

⁶⁹ What follows is Kant's *Groundwork of the Metaphysics of Morals* argument for the 'Formula of Humanity' (1785/1959, pp.427-428) as interpreted by Korsgaard (1996). See especially Korsgaard (1996, pp.120-122).

⁷⁰ See s.3.1, above.

⁷¹ Korsgaard 1996.

particular desire or care or motivational set but just in virtue of your personhood, two things follow.

Firstly, it follows that this Neo-Kantian account of morality is one that construes moral reasons as intrinsic to persons. Part of what it is to be a person, on this Neo-Kantian account, is to be subject to moral reasons. Asking, “Why should I be moral?” is like asking, “Why should I breathe?” And, the reason why you should breathe isn’t because you desire to breathe or care about breathing, but rather just: if you desire anything at all, you ought to breathe, because of the kind of being you are (namely, a pulmonate). Likewise, if you have any reasons at all, you will have moral reasons, just because of the kind of being you are: a person. Note also how the Neo-Kantian argument works here: it works by drawing out the necessary presuppositions of an account like Hobbes’: if it’s the case that my desiring to obtain my PhD gives me a reason to complete my dissertation, then this must be because I am the kind of being for whom its (reflectively endorsed, or rational) desires have normative force – or, to put it more succinctly, because I am a person. Hobbesian Egoism is, in a sense, incomplete. When we work it out fully, when we see that my having reasons like the reason to complete my dissertation presupposes that I am a certain kind of being – the same kind of being that you are too – then we arrive at Neo-Kantianism.

The next point that follows from the Neo-Kantians’ grounding moral reasons’ normative force not in particular features of a person’s self, such as his/her desires, cares or motivational set, but in his/her mere personhood, is that moral reasons will be universal. While a person may fail to desire a non-nasty, non-brutish and long life, or fail to care about morality, or fail to have an internal reason to act morally, he/she cannot fail to be a *person*. As such, all persons, on this Neo-Kantian account, will necessarily be subject to moral reasons; that is to say, moral reasons on this Neo-Kantian account will be universal. So, the Neo-Kantian account apparently provides an account of moral reasons according to which they are intrinsic to the person, have normative force over us, and are universal.

5.1. A problem with the Neo-Kantian account

There is, however, a problem with the Neo-Kantian approach, namely, that its conclusion just doesn’t seem to follow. According to the argument, when we work out accounts like Hobbes’ to completion, we find that it is not my desires, cares or motivational set that ground(s) my reason to, say, complete my dissertation, but ultimately my mere personhood. And, since it is my personhood that gives me my reasons – and, since you too are a person –

it follows that your personhood gives me reasons too, and that I therefore ought to respect your personhood as I do my own; that is, I ought to treat you morally.

But, does this follow? If I have a reason to breathe in virtue of being a pulmonate, and you too have a reason to breathe in virtue of being a pulmonate, it doesn't thereby follow that I have a reason to help you breathe or that I ought to breathe for you, as if you have a right to my lungs equal to my own, or *vice versa*. Nothing follows from this, other than the fact that I have my reason to breathe in a similar manner to that in which you have your reason to breathe – where these are each our own reasons, and *my* having such a reason has no normative force over *you*, nor *vice versa*. Likewise, even if my having a reason to complete my dissertation entails that I value *my* personhood, why should we think that this entails that I ought thereby to value *yours* also? After all, you are one person and I am another. If the argument is to work, we have to show that “valuing humanity in your own person somehow implies, entails, or involves valuing it in that of others.”⁷²

5.2. The Generality argument

But, one may object that this misrepresents the argument. My having a reason to complete my dissertation doesn't, as I described it above, entail that I value *my* personhood in particular. Rather, it entails, more generally, that I value personhood *simpliciter*.⁷³ That is, the normative force of my reason to complete my dissertation derives from the fact that persons *simpliciter* are the sorts of creatures for whom their (rational, reflectively endorsed) desires are normative, and I happen to be one such person. My having any reasons at all commits me to the value of personhood *simpliciter*, not to the value of *my* personhood. And, since *you* too happen to be one such person, it follows that I have moral obligations to you too as I do to the one such person who is myself. That is, universality follows.

Call the inference that I value *my* personhood specifically the ‘particular conclusion’; call the inference that I value personhood *simpliciter* the ‘general conclusion’. This Generality argument claims that the problem with the Neo-Kantian account⁷⁴ is avoided when we draw the general conclusion rather than the particular conclusion. But, I argue in response, *why* should we draw the general conclusion? (Where, of course, it is not open to my opponent to answer: We should draw the general conclusion rather than the particular conclusion in order to account for the universality of moral reasons, as that would be to beg the question.) There is nothing (non-question-begging) to be said in favour of the general conclusion rather

⁷² Korsgaard 1996, p.132.

⁷³ Thanks to Professor Thaddeus Metz for bringing this response to my attention.

⁷⁴ See s.5.1, above.

than the particular conclusion here. One explains the normative force of my reason to complete my dissertation just as well as the other, so we might opt for either. But, if the Neo-Kantian account is to explain the necessary universality of moral reasons, then this is not enough: it has to be demonstrated that the general conclusion is the *correct* conclusion. Otherwise the Neo-Kantian account is no improvement upon the Hobbesian account, which is also able to account for the universality of moral reasons *if* it happens that all persons desire a non-nasty, non-brutish and long life.⁷⁵ Yet, I don't see how it could be demonstrated that the general conclusion is the only one that could be correctly drawn.

Indeed, it strikes me that there is reason for preferring the *particular* conclusion over the general conclusion. This is because the general conclusion would constitute a *hasty generalization*. To conclude on the basis of the fact that, say, I have a reason to complete my dissertation that *personhood generally* is valuable is to ground a very broad conclusion on a very narrow basis. Drawing this inference would be like concluding on the basis of the fact that Sally scraped your car that all women are bad drivers – a hasty generalization. Really, all that this evidence supports is that *Sally* is a bad driver (even though this leaves open the possibility that all women are indeed bad drivers). Likewise, the fact that I have a reason to complete my dissertation really only supports the particular conclusion that *my* personhood is valuable to me. Of course, I may have a larger evidential basis than just a singular observation, but my objection will hold. For instance, I may have a reason not only to complete my dissertation, but also to go for a run and also to prepare tomorrow's lecture and also to try to eat a healthy meal, *etc.*, but, still, all that is justifiably entailed is that *my* personhood is valuable to me, not personhood generally. (Again, to try to broaden this evidential basis even further to include, for instance, my reason not to cause you undue harm would be to beg the question – since what the argument is supposed to prove in the first place is that I have such moral duties to you.) Indeed, broadening the evidential basis in this manner seems only to offer further support for the particular conclusion: if Sally has not only scraped your car, but also scraped her mom's car and her neighbour's car and the Dean's car, then this seems like even better evidence for the conclusion that *Sally* in particular is a bad driver.

So, I argue that the correct conclusion to draw from the Neo-Kantian argument is the one that I drew originally:⁷⁶ the particular conclusion that *my* personhood is valuable to me, rather than the general conclusion the personhood *simpliciter* is valuable to me. The

⁷⁵ See s.3.2, above.

⁷⁶ See s.5.1, above.

general conclusion appears to be an unjustified, hasty generalization, and – even if it isn’t – there is no (non-question-begging) reason to prefer the general conclusion over the particular one as both explain the normative force of my reason to complete my dissertation just as well; as such, the Neo-Kantian argument is unable to explain the necessary universality of moral reasons. We are left right where we ended the previous section: having to show that “valuing humanity in your own person somehow implies, entails, or involves valuing it in that of others.”⁷⁷

5.3. The Analogical argument

Above, I argued that the value of my personhood to me does not entail the value of your personhood to me, and that therefore Neo-Kantianism is unable to account for the necessary universality of moral reasons.⁷⁸ In objection to my view, I then considered the Generality argument, which I subsequently rejected.⁷⁹ Here I consider a second way in which the Neo-Kantian might resist my argument that the value of my personhood to me does not entail the value of your personhood to me, and that therefore Neo-Kantianism is unable to account for the necessary universality of moral reasons.

One simple and natural way to show that the value of my personhood to me entails the value of your personhood to me also (and *vice versa*) is the following. If my personhood is a source of normativity to me, then, by analogy, others’ personhood is likewise a source of normativity to me. To reason otherwise, one might argue, would be inconsistent. Call this the Analogical argument. In Korsgaard’s words:

Since I regard my humanity as a source of value, I must in the name of consistency regard your humanity that way as well. So I must value the things that you value. Or, to put it another way since I think my humanity is what makes my desires into normative reasons, I must on pain of contradiction suppose that the humanity of others makes their desires into normative reasons as well.⁸⁰

There are various problems with this Analogical argument, however. Firstly, even if considerations of consistency do require that by valuing my own personhood I ought also to value that of others, there is no guarantee that a person does or ought to value *consistency*. A person may fail to desire to get a PhD; a person may even fail to desire a non-nasty, non-brutish and long life; and a person may just as well fail to value consistency. (If we valued

⁷⁷ Korsgaard 1996, p.132; as cited in s.6.1, above.

⁷⁸ See s.5.1, above.

⁷⁹ See s.5.2, above.

⁸⁰ Korsgaard 1996, p.133.

Korsgaard goes on to cite Gewirth (1978) and Nagel (1970) as propounding two such arguments for the universality of moral reasons, although Korsgaard notes that she does not recommend reading Nagel in this way.

consistency as much as this Analogical argument supposes we do, we would surely be subject to a lot less cognitive dissonance in our attitudes towards meat-eating, gender- and racial-equality, *etc.*) So, since it is possible that we fail to value consistency, it need not follow that I, on pain of contradiction, ought to value your personhood as I value my own. And, therefore, the universality of morality is not entailed.

However, it may be objected, if a theorist observes a radical difference in kind between practical and theoretical reasons, there may be some grounds for believing, *contra* the preceding argument, that a person *cannot* fail to value consistency.⁸¹ Practical reasons are the sorts of reasons with which a person exercises her autonomy, since a person may choose to desire (by reflective endorsement) some ends rather than other ends. When it comes to theoretical reasons, on the other hand, no such autonomy is exercised; persons do not get to choose what to *believe*, they are constrained in their beliefs by the world.⁸² And, consistency is *prima facie* a *theoretical* consideration rather than a practical one. So, given these differences between practical and theoretical reasons, there may be grounds for thinking that a person actually *cannot* fail to value consistency, since autonomy is not exercised in the realm of theoretical reasons as it is in the realm of practical reasons. As such, it would follow according to the Analogical argument that your humanity is valuable to me as mine is to me (and *vice versa*) on pain of inconsistency, and this would follow universally, since consistency is not something a person may autonomously choose to value or not value.

However, even if this defence works, there are further problems with the Analogical argument. To begin with, appealing to *consistency* at all is misguided. This is because the problem that the Analogical argument responds to is not a problem of *inconsistency* to begin with. The claim to which the Analogical argument responds is the claim that, as my having reasons presupposes the value of my personhood *to me*, your having reasons entails the value of your personhood *to you* – and there is nothing inconsistent about that claim. As such, the appeal to consistency cannot achieve the ends of the Analogical argument anyway. As Korsgaard puts it:

⁸¹ Thanks to Professor Thaddeus Metz for this suggestion at the PSSA conference of 2012.

⁸² But, see Fleischacker's (2009, pp.118-120) objection to Darwall (2006) that, even though [w]e don't choose whether to believe a particular proposition or not, [...] we do make certain choices about the sorts of things to count as facts – think of the rules in a court-room about admissible evidence – or about what to believe when neither position on an important question can be established definitively by reason or empirical evidence, as in the case of the religious issues Kant was addressing. (Fleischacker 2009, p.120, referencing Kant 1786/1998)

Consistency can force me to grant that your humanity is normative for you just as mine is normative for me. It can force me to acknowledge that your desires have the status of reasons for you, in exactly the same way that mine do for me. But it does not force me to share in your reasons, or make *your* humanity normative for me.⁸³

Supposing that the normativity of my personhood for me does not entail the normativity of your personhood for *me* (even though your personhood is normative for *you*) is not inconsistent, so the Analogical argument's appeal to consistency is unhelpful.

And, lastly, even if the Analogical argument *did* entail that I ought to value your personhood as I do my own, Korsgaard argues that it would fall short of generating *moral* reasons. If consistency did actually entail that your personhood is valuable to me,

[i]t would do it in the wrong way. It would show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that I have duties with respect to you, about you, but not that there are things I owe *to* you. But some duties really are owed *to* others: we may be obligated by others, I will argue, in much the same way that we may be obligated by ourselves.⁸⁴

That is, instead of providing an account of moral obligations I owe *to you*, the Analogical approach would provide an account of moral obligations that I owe to *myself* – albeit it moral obligations *concerning you* that I nevertheless owe to myself. Notice that such an account would not have progressed very far beyond its Hobbesian origins:⁸⁵ like the Hobbesian and related approaches,⁸⁶ the Analogical argument tries to account for the normative force of moral reasons by demonstrating their connection to others of our pre-existing reasons – in this case, reasons of consistency. In other words, such a person-regarding account would, like the Hobbesian account, be *first-personal*: an account of obligations owed to myself, in virtue of features of *my* self, albeit obligations about you.⁸⁷

⁸³ Korsgaard 1996, p.134, referencing Williams' (1985, Ch.IV) critique of Gewirth (1978).

⁸⁴ Korsgaard 1996, p.134.

⁸⁵ Indeed, Korsgaard (1996, pp.132-133) discusses such Analogical arguments in the same breath as Neo-Hobbesian accounts.

⁸⁶ See s.3, above.

⁸⁷ This is the way Korsgaard (1996, p.134) presents the argument, and it makes for my segue into a discussion of second-personal reasons too good to refuse. However, I am not sure that Korsgaard's argument works. *If* considerations of consistency do entail that your personhood is valuable for me as my own personhood is for me, then I do *not* think it follows, as I have stated (above), following Korsgaard, that the resultant reasons would be reasons of the same order as Hobbesian reasons, albeit reasons concerning you. This is because such reasons would, unlike Hobbesian reasons, *not* be grounded in features of the person's self, such as desires; they would, rather, be grounded in reasons of *consistency* – and consistency, I remarked above, is a *theoretical* consideration rather than a practical one. And, if, as I considered above, not possessing any autonomy in the realm of theoretical reasons, we are all compelled to be consistent, then it would seem to follow that we all possess moral reasons – and not as a matter of *our* (autonomously) valuing consistency. This strikes me as a relevant disanalogy with the Hobbesian case, *contra* Korsgaard. However, the key is: "*If* considerations of consistency do entail that your personhood is valuable for me as my own personhood is for me..." And, here I think the argument *can* be blocked, although not

What we need instead, Korsgaard suggests, is an account according to which “there are things I owe *to you*”. That is, we need a *second-personal* account.⁸⁸

6. Second-personal accounts

According to the matrix of positions in section 1.3, above, there are two possible second-personal metaethics (that I will consider): Divine Command theory, and Darwall’s Second-Personal account. Here I must for clarity’s sake spell out a convention that I have adopted: the difference between a ‘second-personal’ account (*sans* capitalization), and the ‘Second-Personal’ account (with capitalization). The label ‘Divine Command theory’ easily distinguishes it from the broader category of second-personal metaethics of which it is a member (as I explain shortly); however, the label ‘Darwall’s Second-Personal account’ is not quite as clearly distinct. For this reason, I capitalize the term ‘Second-Personal’ when referring to that particular second-personal metaethic that is Darwall’s Second-Personal account (as opposed to Divine Command theory). Divine commands, therefore, are second-personal but not Second-Personal, whereas reasons of the sort envisaged by Darwall are both second-personal and Second-Personal.

Recall that a second-personal account is one according to which reasons are given (rather than simply pointed out) by a second person to a first. Like a speech act, second-personal reasons include essential reference to two people:⁸⁹ the addressee (the first person) and the addressor (the second person). Since commands are a kind of speech act, it should be clear, therefore, how the Divine Command theory is a second-personal account: according to it, what is moral is a matter of the commands given to us (first persons) by God (the second person). Darwall’s Second-Personal account is clearly also second-personal: like the Divine

in the way described by Korsgaard (and myself), above: since, as I hope to have shown (above), considerations of consistency are irrelevant to the problem to which the Analogical argument tries to respond (since the problem is not one of *inconsistency* in the first place), the antecedent of the conditional simply doesn’t obtain.

⁸⁸ In this section I clearly set Darwall up as responding to a problem which Korsgaard recognizes and describes, so it may appear an omission on my part that I do not address Korsgaard’s own response to the problem she describes. My reasons for this are twofold. Firstly, although Darwall himself is also clearly aware of the large similarities between his account and Korsgaard’s (see, for instance, Darwall 2006, p.11), he does not attempt to explicitly argue for his divergences from her account, or for why his account is preferable to hers. As such, I will attempt (in forthcoming chapters) to tackle Darwall on his own terms, rather than arguing against him on behalf of Korsgaard. Secondly, my Master’s thesis focused on Korsgaard, and so I did not want to reproduce that material for this dissertation. I do, however, go on to draw several points of comparison between Korsgaard and Darwall throughout this project.

⁸⁹ Here I appeal to Nagel’s (1986, p.153, following Parfit 1984) distinction between the agent-relative and the agent-neutral, as cited in s.1.1, above.

Command theory, it holds that what is moral is a matter of what is (or reasonably could be) commanded – where the difference is that Darwall, unlike the Divine Command theorist, holds that we ordinary moral agents, rather than God, are the ones doing the commanding.⁹⁰

According to Darwall, the Divine Command theory is unstable.⁹¹ That is – as I will go on to explain, below – that

in holding that any rational agent God addresses is thereby accountable for obedience, the [Divine Command theorists] made their views subject to pressure from the idea of rational accountability in ways that can seem to cut against [Divine Command theory].⁹²

Furthermore, Darwall argues that when this instability is brought to light and we properly recognize those presuppositions to which the Divine Command theory commits us, what we end up with is just Darwall's Second-Personal account. In this sense Darwall's approach with regard to Divine Command theory is similar to the Neo-Kantians' approach with regard to Hobbes' and related metaethics:⁹³ demonstrating that the proper analysis of another metaethic commits us to the theorist's favoured metaethic. In Darwall's case, the argument is that the address of any second-personal reason – such as a Divine command – commits us to Darwallian Second-Personal reasons. Hence, I will arrive at Darwall's Second-Personal account by way of a discussion of Divine Command theory.

6.1. Divine Command theory

On the face of it, Divine Command theory is in a strong position. As a person-regarding (second-personal) metaethic, it is apparently, like Hobbesian Egoism⁹⁴ and Neo-Kantianism,⁹⁵ easily able to account for the normative force of moral reasons. However, unlike Hobbesian Egoism⁹⁶ and Neo-Kantianism,⁹⁷ Divine Command theory is also easily able to account for the universality of moral reasons. As Metz puts it, the universality of moral reasons on a Divine Command account is as straightforward as this: "If God were to command all persons to perform certain acts and to refrain from others, He could create

⁹⁰ See s.1.2, above.

⁹¹ This useful description as "unstable" is, however, not Darwall's term, but Watson's (2007, p.41).

⁹² Darwall 2006, p.111.

Darwall frames his discussion in terms of 'theological voluntarism' rather than in terms of 'Divine Command theory', as I have done.

⁹³ See s.5, above.

⁹⁴ See s.3.1, above.

⁹⁵ See s.5, above.

⁹⁶ See s.3.2., above.

⁹⁷ See s.5, above.

universally applicable moral norms.”⁹⁸ And, not only will Divine commands be both normative and universal, they will also be that special kind of reason that Korsgaard tells us we need, namely, a second-personal metaethic.⁹⁹ So far, it looks as if Divine Command theory ticks all the boxes.

However, Divine Command theory is not as successful as it may first appear to be. This metaethic is commonly dismissed on the basis of Euthyphro’s Dilemma; however, I do not consider that line of objection here.¹⁰⁰ Instead, I choose to discuss other, novel objections to Divine Command theory. I first raise my own, and then I go on to consider Darwall’s, which will lead us into a discussion of Darwall’s own Second-Personal account.

6.1.1. Problems with the Divine Command theory

Firstly, it’s unclear that Divine Command theory really yields universal moral reasons – or, at least, *necessarily* universal moral reasons. As we noted, for instance, with regards to Hobbesian Egoism, an agent may fail to desire a non-nasty, non-brutish and long life;¹⁰¹ likewise, it may be that God did not as a matter of fact command *all* agents, but only a select few (such as the Jews).¹⁰² As Metz was careful to note (above), “*If God were to command all persons to perform certain acts and to refrain from others, He could create universally applicable moral norms*”,¹⁰³ and it need not be the case that the antecedent of this conditional obtains. As such, Divine Command theory does no better job than Hobbesian Egoism of accounting for the universality of moral reasons.

Secondly, even though Divine commands are, as Korsgaard demands,¹⁰⁴ second-personal reasons, they are not – as I have demanded¹⁰⁵ – moral reasons of our own. It is correct that, as second-personal reasons, Divine commands include essential reference to both us and God.¹⁰⁶ However, insofar as Divine commands depend for their existence on *God* having commanded them, they are (at least partially¹⁰⁷) extrinsic to persons. Like Real, world-

⁹⁸ Metz 2013, p86.

⁹⁹ Korsgaard 1996, p.134, as cited in s.5.3, above.

¹⁰⁰ At least one Divine Command theorist, Cottingham, “has plausibly dealt with” the Euthyphro objection by appealing to God’s perfect nature (Metz 2013, pp.87, referencing Cottingham 2003 and 2005).

¹⁰¹ See s.3.2, above.

¹⁰² See Lerm (2013, pp.143-144) for a discussion of ‘Divine Command – to some but not to others – theory’.

¹⁰³ Metz 2013, p86 (emphasis added).

¹⁰⁴ Korsgaard 1996, p.134, as cited in s.5.3, above.

¹⁰⁵ See s.4, above.

¹⁰⁶ See s.6, above.

¹⁰⁷ I qualify this claim in this way because second-personal reasons are dependent as much on their addressees (in the Divine Command case: us) as they are on their addressors (God). Thus, whereas

regarding reasons, Divine commands come from outside of us – only they come from God, rather than from the world – and, as such, are not what I have called moral reasons of our own. I take it that it is for this reason that it is not uncommon for Divine Command theorists to supplement their accounts with appeals to God’s omniscience and omnibenevolence: because God needs first to be vetted as knowing and wanting what’s best for us before his commands could come to have normative force over us.

Furthermore, it’s unclear that Divine Command theory can account for the normativity of moral reasons. The amoralist can still turn around to the Divine Command theorist and retort, “So God says I must. So what?” That is, despite initial appearances, Divine Command theory accounts neither for the universality of moral reasons, nor their normativity, and (because) it does not construe moral reasons as intrinsic to the person.

One may object, however, that the Divine Command theorist *is* able to respond to the amoralist, and *is* thereby able to account for the normativity of moral reasons. When the amoralist retorts, “So God says I must. So what?”, the Divine Command theorist may respond, quite simply, “Because, if you don’t do what God says, he’ll send you to hell!” That is, you ought to do as God commands, because God will hold you accountable for it.

However, I do not think that this response to the amoralist is successful. How does the prospect of hell confer normativity on God’s commands? One answer is that it simply doesn’t. According to Darwall, this prospect of hell is not a case of genuine moral accountability, and whatever force it has over one is not *normative* force – but I reserve that discussion for the following section. Now, we should consider the somewhat more charitable suggestion that the threat of hell confers normativity on God’s commands because we naturally would wish to avoid hell. Since it is practically rational to take the means to one’s ends, we ought therefore to do as God commands.

Note, however, that claiming that one ought to do as God commands in order to avoid hell is essentially similar to the Hobbesian rationale that one ought to do as the Sovereign commands in order to avoid a nasty, brutish and short life.¹⁰⁸ And note the curious implication of such a construal: I ought not to cause you undue harm, for instance, *so that I don’t go to hell*. But, as Korsgaard observed, such a response to the amoralist would achieve

Real reasons are entirely alien to persons, Divine commands are not *entirely* so – but they are sufficiently so not to count as moral reasons of our own. This point becomes clearer in contrast to Darwall’s Second-Personal account, according to which moral reasons are not given to us by God but are given to us by *each other*.

¹⁰⁸ See s.3.1, above.

normativity “in the wrong way”, by establishing “that I have an obligation *to myself*” to treat you decently (in order to escape hell), “that I have duties with respect to you, about you, but not that there are things I owe *to you*.”¹⁰⁹ Such a response to the amoralist, that is, retreats to the *first-personal*, and, as such, forfeits all the apparent advantages of second-personality and inherits all the problems of first-personality. What if, for instance, a masochistic agent simply fails to desire to avoid hell?¹¹⁰ We should, therefore, resist construing Divine Command theory as offering this first-personal kind of response to the amoralist.¹¹¹ But, how, then, can the Divine Command theorist account for the normativity of moral reasons?

The problem with Divine Command theory, as Darwall will go on to show (below), is God – in particular, God’s greater moral authority. That is, the problem is that Divine Command theory always construes God as the addressing second person, and us as the addressed first persons, with the implication that it is only ever God who holds us accountable, and never vice versa. Such an asymmetrical set-up either (1) yields not second-personal but first-personal reasons to do as God commands in order to achieve one’s desire of avoiding hell, as I have argued in this section, or (2) is not a case of genuine moral accountability at all, but rather just a case of God manipulating us into doing his bidding by the threat of hell, as Darwall argues (below). The solution, Darwall will argue, is to opt for a symmetry, rather than an asymmetry, of moral authority: as the second person holds the first accountable, so too does the first have the authority to hold the second accountable. On the basis of this symmetry, it will follow that second-personal (non-Divine) commands *are* normative because they are made by the second person with *the same* (non-Divine) authority by which the first would give him-/herself first-personal reasons. That is, removing God from the picture will allow us to achieve normativity in the same way in which the Divine Command theorist tried to do, above – but without thereby reducing it to a first-personal account. Furthermore, by allowing that we non-Divine beings all possess the authority of second persons over one another, the account will also achieve universality.

I turn now to Darwall’s criticism of Divine Command theory.

¹⁰⁹ Korsgaard 1996, p.134, as cited in s.5.3, above.

¹¹⁰ See s.3.2, above.

¹¹¹ The same goes for an alternative, positive Divine Command theorist’s response to the amoralist, according to which one ought to obey God’s commands out of *gratitude* for all He has done for one (Pufendorf 1673/1991, p.28, as referenced by Korsgaard 1996, pp.31-32). What if an agent simply fails to be grateful? (And, note that it is not open to the Divine Command theorist to argue here that we *ought* to be grateful to God, because invoking such a moral norm would be question-begging in this context.)

6.1.2. Pufendorf's Point

Darwall¹¹² takes his cue from the 17th-century Divine Command theorist, Samuel von Pufendorf.¹¹³ As we have observed, an agent has a moral reason to perform some action, according to Divine Command theory, when that action is commanded by God. But, Divine Command theory “is often criticized on the grounds that [God] can apparently make *anything* right or wrong.”¹¹⁴ That is, if an action is morally right simply in virtue of God’s commanding it, then what’s stopping God from commanding me to, say, kill my only (legitimate) son?

Pufendorf’s solution is to draw the natural distinction between genuine moral accountability and the superficially similar “coercion”.¹¹⁵ Darwall calls this ‘Pufendorf’s Point’.¹¹⁶ Coercion works by “point[ing] out some object of terror” to the subject which “shakes the will with an external force”.¹¹⁷ ‘Coercion’ may strike the reader as too harsh a term,¹¹⁸ but the idea is basically behavioural modification by negative reinforcement: when the agent misbehaves, some kind of aversive treatment or punishment (an “object of terror”) is applied in order to deter the agent from such behaviour and thereby mould his future behaviour (“[shake] the will”) into more desirable forms.¹¹⁹ This kind of behavioural modification, according to Pufendorf, does not constitute genuine moral accountability, and is not what Divine Command theorists mean when they say we are accountable to God; God isn’t just our trainer, trying to improve our earthly behaviour. The following argument does not appear as part of (Darwall’s presentation of) Pufendorf, but it is easy to fill in these details. The behavioural modification account is clearly future-directed: bad behaviour is punished so

¹¹² Darwall 2006, mainly pp.111-112.

¹¹³ Pufendorf 1672.

¹¹⁴ Korsgaard 1996, p.23 (emphasis added).

Korsgaard makes this remark in the context of a discussion of both Hobbes and Pufendorf, so the remark is made with regards to Hobbes’ ‘Sovereign’, rather than to God (as I have edited the quote), but it is equally applicable to both.

¹¹⁵ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111.

¹¹⁶ Darwall 2006.

¹¹⁷ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111.

¹¹⁸ Other translations give the term “compulsion” (instead of “coercion”). See Pufendorf 1672/1729, p.257, and 1672/2003, p.180.

¹¹⁹ Note that this usage of ‘negative reinforcement’ is less than accurate. Strictly speaking, negative reinforcement occurs when a *present* aversive stimulus is removed when the desired outcome is achieved, e.g. your headache is relieved when you take an aspirin (so you learn to take an aspirin whenever you have a headache). As I am using the term, however, ‘negative reinforcement’ indicates the *application* of an aversive stimulus in order to decrease undesirable behaviour. Technically speaking, this is more accurately called ‘negative *punishment*’; I avoid that term, however, because ‘punishment’ is potentially confounding when it comes to drawing Pufendorf’s distinction between behavioural modification through ‘negative reinforcement’ and genuine moral accountability, both of which may take the form of punishment.

that the agent doesn't commit similar bad acts again in the future. But, since, the story usually goes, God holds us accountable at the *end* of our lives – or even at the *end* of all time – such *future*-directed considerations clearly can't be what underpins Judgement Day. When God finally gets around to holding us accountable, he's not doing so in order to modify our behaviour, since by that stage we've already done all the behaving we'll ever do. What's going on when God holds us accountable, therefore, must be something other than future-directed behavioural modification by negative reinforcement or 'coercion'.

What, then, is going on when God holds us accountable by, for instance, sending us to hell? According to Pufendorf, genuine moral accountability, as opposed to mere behavioural modification,

forces a man to acknowledge of himself that the evil, which has been pointed out to the person who deviates from an announced rule, falls upon him justly.¹²⁰

Note that some "evil" – like "some object of terror" (above) – is applied even in the case of genuine moral accountability. Thus, as I remarked, 'coercive' behavioural modification and genuine moral accountability are superficially similar. The "evil" of genuine moral accountability may even be the same as the "object of terror" of behavioural modification – they may both, for instance, be (the threat of) hell. The difference, however, is that in cases of genuine moral accountability the agent

is forced of [him]self to weigh [his] own actions, and to judge [him]self worthy of some censure, unless [he] conforms to a prescribed rule [...].¹²¹

That is, a certain inner state is present in cases of genuine moral accountability that is not present in cases of mere behavioural modification: the judgement that one is "worthy of some censure", the acknowledgement of one's deserving an "evil" for one's wrongdoing,¹²² or – as I will go on to say – the conviction that what the agent did was wrong such that he/she may justly be (proportionally) punished for it.¹²³ In Darwall's words:

¹²⁰ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, pp.111-112.

¹²¹ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111.

I must note that I have struggled to find this portion of Pufendorf's text, as well as that cited immediately above, in either the Oldfather & Oldfather (1934) translation that Darwall himself used (which is apparently reproduced in Schneewind (2003), which I used) or the earlier Kennett (1729) translation. However, since this distinction between genuine moral accountability and mere behavioural modification is such a small element of Pufendorf's (1672) larger project – the bulk of which Darwall does not address – I do not consider it damaging to Darwall's argument, even if Pufendorf has been misquoted. The distinction between genuine moral accountability and mere behavioural modification is a natural and plausible one, regardless of whether Pufendorf actually made it in these terms or not.

¹²² Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.112, as cited above.

¹²³ I expand on this in Ch.5, s.1.2.1.

To see what Pufendorf is getting at, we need to distinguish between merely external censure or blame [...] and internal blame, that is, the self-reactive attitude of guilt. Blame is purely external when it comes from outside and is not accepted. [...] When, however, one accepts and acknowledges blame or censure, one owns the blaming address of the censuring person second-personally; one credits that imputation, and thus also blames oneself.¹²⁴

Whereas behavioural modification by negative reinforcement merely “shakes the will with an external force”,¹²⁵ genuine moral accountability is marked by the internal, “self-reactive attitude of guilt”. The upshot of Darwall’s reference to the Strawsonian reactive attitudes¹²⁶ here is that a ‘self-reactive attitude’ is one by which an agent holds him-/herself accountable – “blames oneself”, as Darwall puts it, above. That is, an agent who is not merely coerced by the threat of hell but is held genuinely morally accountable by God must be one who *holds him-/herself accountable*. This is, as Pufendorf puts it, the “intrinsic impulse of the mind”¹²⁷ that is present in cases of genuine moral accountability, but absent in cases of mere behavioural modification by negative reinforcement.

And, this solves the Divine Command theorist’s worry that God could obligate me to do just *anything*, simply in virtue of commanding me to do so. Pufendorf’s Point implies that God can only hold me accountable – genuinely morally accountable – to that to which I hold *myself* accountable. As Korsgaard puts the point, “So Pufendorf [...] [was] right. [...] It is the fact that we command ourselves to do what we find it would be a good idea to do” “that obligates us to perform it.”¹²⁸

6.2. Morality as Equal Accountability

This is the point at which the instability of the Divine Command theory becomes clear.

Darwall concludes on the basis of Pufendorf’s Point that “to be accountable to God, moral agents must also be accountable to themselves.”¹²⁹ Divine Command theory’s idea that agents are accountable to God – where this is genuine moral accountability and not mere behavioural modification by negative reinforcement – actually presupposes that agents are accountable to *themselves*. And, “[o]nce we view things this way, significant pressures develop on the sort of [Divine Command theory] that Pufendorf put forward.”¹³⁰

Pufendorf divides obligations “into *mutual* and *not mutual*”, defining the latter as follows:

¹²⁴ Darwall 2006, p.112.

¹²⁵ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111, as cited above.

¹²⁶ Strawson 1968.

¹²⁷ Pufendorf 1672/2003, p.180.

¹²⁸ Korsgaard 1996, p.105 (emphasis removed).

¹²⁹ Darwall 2006, p.114.

¹³⁰ Darwall 2006, p.113.

An *Obligation not mutual* is, when one Party stands bound to perform somewhat towards another; yet so that other Party shall lie under no corresponding *Obligation*, nor be tied, to make equivalent Requitall. Such is the *Obligation* of Men towards God, by which they are engaged to pay him absolute Obedience; but he on his Part is not constrain'd by Virtue of any such external *Obligation*, to make any Return for their Obedience.¹³¹

That is, “*not mutual*” obligations are ‘one-way’ obligations, such as “the *Obligation* of Men towards God”, wherein a first person owes a duty to a second, but that second person owes the first no corresponding duty in return. Mutual obligations, by contrast, are those in which both parties are equally and reciprocally obligated to and by the other. Darwall’s argument will effectively be that Pufendorf’s Divine Command theory is unstable in that his distinction between “*mutual* and *not mutual*” obligations is inconsistent with his distinction between genuine moral accountability and behavioural modification – what Darwall identified as Pufendorf’s Point, above.¹³²

According to Pufendorf’s Point, God’s holding us genuinely morally accountable presupposes that *we too* hold ourselves accountable.¹³³ That is to say, Pufendorf’s Point reveals that a second person’s addressing a command (where this is a command to which the first person is held genuinely morally accountable) must presuppose of the first person the *same authority* as the second person to hold him-/herself accountable. Any second-personal address presupposes the equal authority of both the first and the second persons – without which the second person couldn’t be holding the first *genuinely morally accountable*. So, Pufendorf is simply mistaken that “the *Obligation* of Men towards God” is “*not mutual*”.¹³⁴ If it’s an obligation for which the agent is genuinely morally accountable (rather than liable merely to behavioural modification by negative reinforcement), then it must be a *mutual* obligation: one which not only the second person has the hierarchical, one-way authority to impose on the first person, but which the first person also has the same authority to impose on him-/herself and reciprocally on the second person. That is, the address of any second-personal reason – such as a Divine command – commits us to Darwallian Second-Personal reasons, which are mutual and reciprocal. Darwall calls this thesis Morality as Equal Accountability.¹³⁵

¹³¹ Pufendorf 1672/1729, p.258.

¹³² See s.6.1.2, above.

¹³³ See s.6.1.2, above.

¹³⁴ Pufendorf 1672/1729, p.258, as cited above.

¹³⁵ Darwall 2006, p.101 (emphasis removed).

From this point onwards, I will use ‘Morality as Equal Accountability’ as a synonym for ‘Darwall’s Second-Personal account’.

So, for instance, your giving the foot-treader a second-personal reason to get off your foot¹³⁶ presupposes not only that you hold the foot-treader accountable for compliance, but also that the foot-treader must hold *himself* accountable. That is to say, the foot-treader is equally accountable both to you and to himself. Likewise, on another occasion, the foot-treader's giving *you* a reason would presuppose not only that he holds you accountable for your subsequent action, but that you too hold *yourself* accountable – that you are equally accountable both to him and to yourself. And, it is not hard to see how this extends to every person given or giving a reason to another. That is, the *universality* of moral reasons follows on the basis of Morality as Equal Accountability. There is not only a second-personal reason for the foot-treader to remove his foot from on top of yours when you tell him to do so, but a mutual, Second-Personal reason¹³⁷ for *any* agent to remove his/her foot from on top of any other agent's, as we can see when we examine the presuppositions of second-personal address.

6.2.1. Fichte's Point

Pufendorf's Point, that there is a distinction between genuine moral accountability and behavioural modification by negative reinforcement, leads us to the conclusion that any genuinely morally accountable agent must hold him-/herself accountable.¹³⁸ And, on this basis Morality as Equality Accountability follows; that is, it follows that first and second persons share an equal authority to hold each other accountable.¹³⁹ This latter inference, however, is not made by Pufendorf (hence the instability of his Divine Command theory), but Darwall does find support for it from another philosopher: Johann Gottlieb Fichte. Darwall calls it 'Fichte's Point':¹⁴⁰

[A]ny second-personal claim or "summons" (*Aufforderung*) presupposes a common competence, authority, and, therefore, responsibility as free and rational, a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally.¹⁴¹

That is, when I give an agent a summons (to use Fichte's term) or a second-personal reason (to use Darwall's term) and subsequently hold that agent accountable for behaving in accordance with that summons/second-personal reason, I must presuppose that that agent is able to hold him-/herself accountable in precisely the same way in which I hold him/her

¹³⁶ See Darwall 2006, p.7, as cited in s.1.2, above.

¹³⁷ See the distinction between 'second-personal' and 'Second-Personal' is s.6, above.

¹³⁸ See s.6.1.2, above.

¹³⁹ See s.6.2, above.

¹⁴⁰ Darwall 2006, pp.22-23.

¹⁴¹ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000.

accountable – otherwise, as Pufendorf’s Point states,¹⁴² I’m not holding the agent genuinely morally accountable, but only trying to modify his/her behaviour by negative reinforcement.

Presupposing that that agent is able to hold him-/herself accountable in precisely the same way in which I hold him/her accountable depends on two further presuppositions, which together constitute Fichte’s Point. Firstly, I must presuppose, that the agent and I share a “common competence”,¹⁴³ or – as Darwall goes on to explain – share “a capacity to determine themselves by these reasons”¹⁴⁴ or “share a freedom to act on claims that are rooted in this authority.”¹⁴⁵ If ought implies can, then of course neither I nor the agent would have been in a position to hold the agent accountable if he/she did not possess the competence to act on the relevant reasons. This Darwall calls “second-personal competence”.¹⁴⁶ Secondly, I must also presuppose that the agent has the same authority to hold him-/herself accountable as I do, since, if he/she did not possess this authority, then he/she would not be able to hold him-/herself accountable as Pufendorf’s Point requires. This Darwall calls “second-personal authority”.¹⁴⁷ As we saw above,¹⁴⁸ it is this second-personal authority that constitutes the major shift from Divine Command theory to Darwall’s Morality as Equal Accountability, since it is no longer supposed that only God has the authority to hold us accountable, but that we also have the authority to hold ourselves – and, therefore, each other – accountable. So, when I summons you or give you a second-personal reason, what follows is, as quoted above, “a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally”.¹⁴⁹ And, to this pair of presuppositions (*i.e.* second-personal competence and second-personal authority) Darwall gives the shorthand label, *dignity*:

[T]he address of second-personal reasons of any kind carries presuppositions that, when fully worked clear, commit addresser and addressee alike to their second-personal competence and to an equal second-personal authority rooted in that, hence to the equal dignity of persons [...].¹⁵⁰

Fichte’s Point, therefore, is that any second-personal address presupposes both the first and second persons’ equal dignity, which is reciprocally recognized by both parties, and which therefore supports the universality of moral reasons.

¹⁴² See s.6.1.2, above.

¹⁴³ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited above.

¹⁴⁴ Darwall 2006, p.21.

¹⁴⁵ Darwall 2006, p.246.

¹⁴⁶ Darwall 2006, see for instance p.246.

¹⁴⁷ Darwall 2006, see for instance p.246.

¹⁴⁸ See s.6.2, above.

¹⁴⁹ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited above.

¹⁵⁰ Darwall 2006, p.81.

Darwall's invocation of Fichte helps to draw the distinction between Kant's metaethic and Darwall's Second-Personal account, whereby – as Douglas Lavin puts it – “Darwall turns against his Kantian roots”.¹⁵¹ Fichte's aim was to improve upon what he regarded as the shortcomings of Kant's philosophy, and Darwall's aim in *The Second-Person Standpoint* is to do the same. According to Darwall, his Second-Personal account provides a better account of Kantian dignity than does Kant's own account:

It turns out, I argue, that Fichte's Point is exactly what is needed to work out properly the central Kantian ideas of the dignity of persons [...].¹⁵²

I will not, however, be engaging in an extended discussion of the nature of dignity. I think we already have sufficient material to demonstrate how Darwall's Second-Personal account improves upon Kant's metaethic (or, at least, a Neo-Kantian metaethic) – as well as those other metaethics examined in this chapter – as I will go on to show in the following section. I will, however, return to a discussion of Fichte, in the final chapter.

6.2.2. Darwall's Second-Personal account

So, let us note the very favourable position that Darwall's Second-Personal account occupies. I have already remarked that Darwall employs a similar strategy to the Neo-Kantians': where the Neo-Kantians argue that an agent's having any kind of reason commits that agent to having moral reasons also,¹⁵³ Darwall argues that any second-personal reason (like a Divine command) commits one to his Second-Personal view.¹⁵⁴ And, we should note some further points of comparison too.

Firstly, notice that Darwall's Second-Personal account (or, indeed, any second-personal account) is, like the Neo-Kantian account, *person-regarding* rather than world-regarding. A Darwallian Second-Personal reason is not one that is “there anyway”¹⁵⁵ and that you have to get the other agent to recognize. Rather, Second-Personal reasons “always derive from agents' relations to one another”,¹⁵⁶ being given by a second person to a first person, and so they are person-regarding reasons. This means that, like the Neo-Kantian account, Darwall's

¹⁵¹ Lavin 2008.

¹⁵² Darwall 2006, p.22.

Towards the end of the book, however, Darwall frames this claim in terms of Kant's metaethic entailing Second-Personality:

In holding that the dignity of persons is that by which we can “demand,” “exact,” or “require” respect from others, that each thereby has a “legitimate claim” to respect, Kant is committed to the idea that the dignity of persons includes a second-personal authority to address demands for compliance with the first-order duties of respect. (Darwall 2006, p.137.)

¹⁵³ See s.5, above.

¹⁵⁴ See s.6.2, above.

¹⁵⁵ Darwall 2006, p.6, as cited in s.1.2, above.

¹⁵⁶ Darwall 2006, p.8.

Second-Personal account is able to offer a straightforward explanation of moral reasons' *normative force*: moral reasons have normative force over me in virtue of some feature of my self, namely, my standing in relationship to you. In Darwall's words:

it may be possible, consistently with the substantive normative claims of this book (that equal second-personal authority grounds [S]econd-[P]ersonal reasons), to hold a version of Williams's internal reasons thesis.¹⁵⁷

That is, deriving as they do from features of persons' selves – namely their standing in relationships to each other – Darwallian Second-Personal reasons will be what Williams calls internal reasons, and so will easily account for the normative force of moral reasons.¹⁵⁸

Indeed, Talbot Brewer reads this Second-Personal account of moral reasons' normative force as "[t]he guiding ambition of Stephen Darwall's *The Second-Person Standpoint*".¹⁵⁹ This is a "guiding ambition" that Darwall's Second-Personal account achieves, in effectively the same way in which the Neo-Kantian account achieves it.

And, not only does Darwall provide an account of moral reasons as person-regarding and therefore having normative force over us, he does so in a way that construes them as *moral reasons of our own*. Because Darwallian Second-Personal reasons "always derive from agents' relations to one another",¹⁶⁰ they are intrinsic to persons. Second-Personal reasons are not – as we saw Korsgaard complain – obligations "to myself [...] with respect to you, about you", but are, rather, genuinely obligations "owe[d] to you",¹⁶¹ because they *come from you*, rather than from myself or from God.¹⁶² Darwall may derive his Second-Personal thesis of Morality as Equal Accountability from Pufendorf's Point,¹⁶³ but it is perfectly described by Korsgaard's words: "we may be obligated by others [...] in much the same way that we may be obligated by ourselves."¹⁶⁴ Second-Personal reasons are intrinsic to persons and are moral reasons of our own because they are the functions of persons relating to other persons. As Darwall puts it:

We presuppose the equal dignity of rational beings and our ability to act on a "law" or reason – a [S]econd-[P]ersonal reason grounded in this dignity – that derives not from the value of any state of affairs or outcome that might be the object of a desire, but, ultimately, from *what it is to be one free and rational person interacting with others*.¹⁶⁵

¹⁵⁷ Darwall 2006, pp.94-95, referencing Williams 1979/1981.

¹⁵⁸ See s.4, above.

¹⁵⁹ Brewer 2008, p.163.

¹⁶⁰ Darwall 2006, p.8, as cited above.

¹⁶¹ Korsgaard 1996, p.134, as cited in s.5.3, above.

¹⁶² See s.6.1.1, above.

¹⁶³ See s.6.2, above.

¹⁶⁴ Korsgaard 1996, p.134, as cited in s.5.3, above.

¹⁶⁵ Darwall 2006, p.33 (emphasis added).

Because Darwallian Second-Personal reasons are moral reasons of our own *qua* persons, they do not stand in need of vindication before they come to have normative force over us, unlike Hobbesian Egoist reasons, for instance.¹⁶⁶ As moral reasons of our own, “a [S]econd-[P]ersonal reason [is] grounded in this dignity”, and “this dignity” is reciprocally recognized by both the first and the second person, as Fichte’s Point observes.¹⁶⁷

So, it follows, furthermore, that Darwall’s Second-Personal account enjoys the additional advantage over the Neo-Kantian account of being able to respond to the amoralist’s challenge.¹⁶⁸ If I am an amoralist foot-treader and you tell me to get off your foot, I may wonder why I should be moral. But, if Fichte’s Point is correct, then it will follow that the amoralist’s question answers itself. We need not cite the benefit of a non-nasty, non-brutish and long life, nor the value of consistency, nor the awfulness of hell – all of which the amoralist may simply fail to care about; on Darwall’s Second-Personal account, we need only cite the fact that you told me, the amoralist foot-treader, to get off your foot – that you addressed me *second-personally*, because

the address of second-personal reasons of any kind carries presuppositions that, when fully worked clear, commit addresser and addressee alike to [...] the equal dignity of persons [...].¹⁶⁹

That is, the mere fact of your addressing me second-personally by telling me to get off your foot commits me to your equal dignity, and, as such, commits me to Second-Personal reasons. The amoralist’s question answers itself. You ought to be moral just because of the kind of being you are: one who relates second-personally to others.

Furthermore, Darwall’s Second-Personal account will also be able to account for moral reasons’ *universality*.¹⁷⁰ As I have presented it, this is the major advantage of Darwall’s Second-Personal account over the Neo-Kantian account. It follows because of the basic difference between Neo-Kantian accounts and Darwall’s account – namely, the shift from first-personal to second-personal. It is the *second-personality* of Darwall’s account that enables it to evade the problem faced by Neo-Kantian accounts. That problem, recall, was

¹⁶⁶ See ss.3 and 4, above.

¹⁶⁷ See s.6.2.1, above.

¹⁶⁸ See s.2.2, above.

¹⁶⁹ Darwall 2006, p.81, as cited in s.6.2.1, above.

¹⁷⁰ Darwall, however, does not in *The Second-Person Standpoint* describe his account as “universal”, nor as (homogeneously) agent-neutral, although he does in a later paper state that it is “uncontroversial” “that fundamental moral obligations apply equally to all moral agents.” (Darwall 2010c, p.40; emphasis removed). In *The Second-Person Standpoint*, he describes Second-Personal reasons as ‘mutual’ and ‘reciprocal’, such that, if you have a reason not to stand on my foot, then it follows that I too have a reason not to stand on your foot. I take it that such mutuality will, on a large scale, amount to universality.

this: even if Kant is correct that my having a reason to, say, complete my dissertation derives its normative force from my personhood, it doesn't follow that *your* personhood is thereby also a source of normativity for me.¹⁷¹ Thus, we were faced with the problem of bridging the gap between the normativity of my personhood and that of your personhood, and we examined the Generality and Analogical arguments¹⁷² in response to this problem – as well as the various failures of those arguments. Darwall's Second-Personal account, on the other hand, doesn't face this problem and thus doesn't require anything like the Generality or the Analogical argument in order to cross the divide from me to you. This is because, on Darwall's Second-Personal account, the other person is already *built into* the account. Moral reasons, on this account, are reasons given *by a second person to a first person*, so "it takes two to make a [Second-Personal] reason."¹⁷³ There is, thus, no question of how to expand the normativity of my personhood to your personhood in order to produce an account of the universality of moral reasons; you are *there* in the normativity of my reasons right from the beginning, and *vice versa*.

Darwall's Second-Personal account, therefore, appears to tick all the boxes: as a person-regarding account, it is able to account for the normativity of moral reasons straightforwardly; as a second-personal account, it construes moral reasons as reasons of our own, intrinsic to persons; and, as Darwall's analysis of Pufendorf's Point has shown, second-personality entails universality in such a way that the amoralist's challenge is met. Darwall's Second-Personal account therefore enjoys the advantage over Realism, Hobbesian Egoism, Neo-Kantianism and Divine Command theory.

There is one final advantage of Darwall's Second-Personal account, which will not serve as a point of comparison with those other metaethics examined in this chapter, but which is worth mentioning for the sake of a more thorough examination of Darwall's philosophy. This further advantage of the Second-Personal account is that it vindicates Contractualism, which Darwall is well-known for defending.¹⁷⁴ According to Darwall's favoured Scanlonian version of Contractualism

¹⁷¹ See s.5.1, above.

¹⁷² See ss.5.2 and 5.3, above.

¹⁷³ Korsgaard 1996, p.138.

Note, however, that Korsgaard makes the remark, not with reference to Darwall's Second-Personal account, but with reference to her own account of reasons as public.

¹⁷⁴ Darwall 2006a and 2006, Ch.12, pp.300-322.

An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement.¹⁷⁵

That is, according to Contractualism, an action is wrong if it might reasonably be rejected by any member of the moral community – that’s *any* member of the moral community, not only those affected by the action, nor only the powerful, nor only the clear of conscience, etc. An action is wrong, according to Contractualism, if any one person (at least) reasonably rejects it. One might question the Contractualist, however, as to his clear presumption that any member of the moral community has the authority to veto an action in this manner. As Darwall puts this presumption,

As I understand it, therefore, the root [C]ontractualist idea is that this standing is one that you and I share. We have an equal [...] authority to make claims of one another, which we respect in seeing each other as beings to whom we should be able to justify ourselves.¹⁷⁶

But why, we might ask, should we accept this Contractualist “root [...] idea”? Why should we accept, as the Contractualist takes for granted, that we all have such equal moral authority?

Darwall’s Morality as Equal Accountability can supply the answer, and thus vindicate his favoured variety of Contractualism. As we have seen, any second person’s holding any first person genuinely morally responsible presupposes that that first person has the same authority to hold him-/herself accountable.¹⁷⁷ A one-way, hierarchical, or (as Pufendorf put it) “not mutual”¹⁷⁸ authority, therefore, is unstable, and Morality as Equal Accountability follows.¹⁷⁹ Darwall has demonstrated what “any form of [C]ontractualism is grounded on”, namely, “the equal moral status of persons”.¹⁸⁰ Morality as Equal Accountability is, as Darwall calls it in *The Second-Person Standpoint* chapter dedicated to its discussion, the “foundation for Contractualism”,¹⁸¹ or, as he says in an article published just prior to the book, “the root from which [C]ontractualism [...] grows”.¹⁸² The final advantage of Darwall’s Second-Personal account, therefore, is that it lays the groundwork for the Contractualism that Darwall wishes to defend, by establishing that all persons share equal moral authority.¹⁸³

¹⁷⁵ Scanlon 1998, p.153.

¹⁷⁶ Darwall 2006a, pp.204-205.

¹⁷⁷ See s.6.1.2, above.

¹⁷⁸ Pufendorf 1672/1729, p.258, as cited in s.6.2, above.

¹⁷⁹ See s.6.2, above.

¹⁸⁰ Ashford & Mulgan 2012, s.2 (emphasis removed).

¹⁸¹ Darwall 2006, Ch.12, pp.300-322.

¹⁸² Darwall 2006a, p.214.

¹⁸³ I return to a discussion of Darwall and Contractualism in Ch.2, s.2.3.3.2.2.

7. Conclusion

Thus, Darwall's Second-Personal account is apparently in a very strong position. It has the advantage over all the other metaethics considered in this chapter – Realism, Hobbesian Egoism, Neo-Kantianism, and Divine Command theory – in that the Second-Personal account can account for the universality and the normative force of moral reasons simultaneously, and it does so in a way that construes moral reasons as intrinsic to persons – *moral reasons of our own* – and thereby successfully addresses the amoralist's challenge.

However, I will go on to mount a detailed criticism of Darwall's Second-Personal account over the next four chapters. Despite initial appearances, I think the Second-Personal account ultimately fails. The criticism will essentially amount to the claims that, on the one hand, Darwall's Second-Personal account boils down either to a first-personal account, much like a Neo-Kantian account, or to a Realist, world-regarding account – in which case it forfeits any of its *prima facie* distinctly second-personal advantages – or, on the other hand, the argument for the Second-Personal account just doesn't get off the ground in the first place.

However, in the final chapter, I will return to attempt a defence of Darwall's Second-Personal account – or, something like it, at least – where I will return, in particular, to an examination of the Fichtean argument, presented above.¹⁸⁴ I think Fichte is the best hope for a successful Second-Personal metaethic, but I think the approach it yields is quite different to the one Darwall has in mind. But, before we get there, let us first turn to the critique of Darwall's Second-Personal account.

¹⁸⁴ See s.6.2.1, above.

2.

Morality as Equal Accountability¹

This chapter begins my critique of Darwall's Second-Personal account of the nature of moral reasons. In the preceding chapter, I argued that Darwall's Second-Personal account has the advantage over several other metaethics as it is able to account both for moral reasons' normative force and their universality, and in a way that construes moral reasons as intrinsic to persons. However, in this present chapter I go on to explain that those appearances are deceiving. As I have, in the preceding chapter, sought to give a novel presentation of Darwall's Second-Personal account, I try also to raise novel criticisms in this chapter. I will consider a range of complaints, some more compelling than others, but all of which are aimed at Darwall's argument for Morality as Equal Accountability. I group these criticisms loosely under two headings.

Firstly, I will argue that Darwall's argument for Morality as Equal Accountability² in fact does not go through. I call this the *non-sequitur* problem. I argue that, like the Neo-Kantian account, Darwall's Second-Personal account is able to explain the normative force of moral reasons but unable to explain their universality. I also question what support Pufendorf's Point can really offer Morality as Equal Accountability.

Secondly, I will present the sceptical problem, according to which we are left without any good reason for believing that there are such things as second-personal reasons in the first place. If Darwall's Second-Personal account is unable, as I will go on to show in section 1, to account for both the normative force and the universality of moral reasons, then any support by way of explanatory power is lost. There is little reason for believing that there are indeed such things as second-personal reasons, and indeed some positive reasons for believing that there are *no* such things as second-personal reasons. As such, we ought to be sceptical of the existence of second-personal reasons at all. Given the way in which Darwall sets up his account as the proper conclusion of Divine Command theory,³ Second-Personality inherits the problems of Divine Command theory: together, my criticisms effectively amount to the Euthyphro's dilemma critique.

¹ Much of this paper appeared as the second half of Lerm 2012.

² See Ch.1, s.6.2.

³ See my Ch.1, s.6.

1. The *non-sequitur* problem

Like the Divine Command theory, Darwall's Second-Personal account accepts that moral reasons have the form of commands; as such, they are *person-regarding*, rather than world-regarding.⁴ However, according to Darwall, Divine Command theory suffers a certain instability – an instability that Darwall's Second-Personal account resolves. Darwall's argument is that, as soon as we recognize Pufendorf's Point – that there is a distinction between genuine moral accountability and mere behavioural modification by negative reinforcement⁵ – the Divine Command theorist must admit that God's holding us genuinely morally accountable presupposes that we are each able to hold ourselves accountable (otherwise God would simply be attempting to modify our behaviour by the threat of hell). As such, any agent being held accountable must share the same authority to hold him-/herself accountable as he/she who holds the agent accountable. And, so, *Morality as Equal Accountability* follows from the proper analysis of Divine Command theory.⁶ Or, to put the point in Fichtean terms:

[A]ny second-personal claim or "summons" (*Aufforderung*) presupposes a common competence, authority, and, therefore, responsibility as free and rational, a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally.⁷

And, because such second-personality is "common", "share[d]", "mutual" and "appropriately recognized reciprocally" amongst us, it follows that moral reasons are *universal*. Moral reasons, therefore, according to Darwall's Second-Personal account, are commands, not that God gives us (as on the Divine Command account), but that any agent may give any other agent.

However, I do not think that this argument for *Morality as Equal Accountability* works. It simply does not follow that the second person's authority to give a first person a reason entails the first's reciprocal authority to give the second a reason. This is the *non sequitur* problem. If it sticks, Darwall's Second-Personal account will not be able to achieve the universality of moral reasons as I have presented it. The *non-sequitur* problem is really quite straightforward. I explain, below.

If it were the case that second-personality were "common", "share[d]", "mutual" and "appropriately recognized reciprocally" amongst us,⁸ it *would* follow that moral reasons are universal. The problem, however, is that Darwall does not provide sufficient reason for us to accept the antecedent of this conditional. Take Darwall's foot-treader scenario as an example.⁹ The foot-treader has a

⁴ See Ch.1, s.1.

⁵ See my Ch.1, s.6.1.2.

⁶ See Ch.1, s.6.2.

⁷ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited in my Ch.1, s.6.2.1.

⁸ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited above.

⁹ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

reason to remove his foot from on top of yours, and he has it in virtue of your giving him such a reason by commanding him to do so.¹⁰ You, therefore, have authority to hold the foot-treader accountable. And, if we accept Pufendorf's Point, it also follows that the foot-treader has the authority to hold *himself* accountable. As such, you and he share an authority – but, we must be careful about the domain of this shared authority. So far, we have established only that the two of you share an authority *over the foot-treader* (and not over anybody else). You have authority over the foot-treader, and the foot-treader has authority over the foot-treader. It does not follow on this basis, however, that the foot-treader has reciprocal authority over *you*. Thus, this falls short of being a *mutual* accountability.¹¹ We can illustrate this point with a (horrible) joke my partner likes to make. He says, “Jess, I love you”, to which I reply, with all the sweetness I have within me: “Me too, darling.” His response: “What, you mean you love *yourself*?” Of course, what I mean by “me too” is that I love *him*, as he loves me – that his love is *reciprocated* – but, taken literally, all that is established is that I share the same love for *myself* that he has for me. This is insufficient to establish universality.

Darwall's argument for Morality as Equal Accountability fails in the same way as my attempts at romance. All that it establishes is that the second and first person share authority over the first; it does not on this basis alone follow that the first person has reciprocal authority over the second. The earlier analogy of second-personal, “bipolar” reasons with electrical current¹² actually points to this problem: current doesn't flow *between* poles; it flows *from* one pole *to* another. Current is unidirectional. And, it would seem that this is all Darwall's analysis of second-personal reasons is able to establish about accountability: unidirectional accountability, as opposed to the mutual, reciprocal accountability necessary for Morality as Equal Accountability. The possibility is left open that a second-person amoralist may enjoy authority over a first person – whereby a first person will owe duties to the amoralist – and, yet, the amoralist may fail to owe any reciprocal duties to the first person, if that first person does not have reciprocal authority over the amoralist. Darwall's Second-Personal account does not preclude the possibility of the amoralist after all. *Contra* Fichte's Point, reciprocal dignity is not entailed: second-personality doesn't entail Second-Personality. And, without this, Darwall's Second-Personal account cannot establish the universality of moral reasons.

This outcome is reminiscent of the problem with the Neo-Kantian account, as discussed in the previous chapter. Although the Neo-Kantians are able to demonstrate that my having any reasons

¹⁰ Here I refer, of course, to Darwall's (2006, p.7) second-personal response to the foot-treader (not the world-regarding one). See my Ch.1, s.1.2.

¹¹ See Watson 2007 for much the same critique.

¹² See Ch.1, s.1.2.

at all entails my valuing *my* personhood, it does not follow that I ought thereby to value *your* personhood too, and as such the Neo-Kantian account struggles to account for the universality of moral reasons.¹³ Likewise, although Darwall is able to show with the help of Pufendorf that your holding me genuinely morally accountable entails my authority over *myself*, it does not follow that I thereby have reciprocal authority over *you*; all that follows is that both you and I have authority over *me* (only). As such, even though these person-regarding accounts are able to explain the normative force of moral reasons, they are unable to explain their universality. And, as I go on to argue, below, the similarities between Darwall's Second-Personal account and the Neo-Kantian account go even further than this.

1.1. The Generality argument

In response to the *non-sequitur* problem, Darwall might argue that what Pufendorf's Point demonstrates is not that, say, both you and the foot-treader have authority over *the foot-treader* (only), but that both you and the foot-treader have authority over *persons*, generally speaking. And, if it's the case that both you and the foot-treader have authority over persons in general, then that *will* entail that the foot-treader has reciprocal authority over you, since you are a person. This appears to be part of Fichte's Point (quoted above), where Darwall asserts that equal dignity follows simply in virtue of being "free and rational".¹⁴ Later on, this move is made explicitly:

Thus, a request or a demand to another that he move his foot from on top of yours might most naturally be presented and interpreted as grounded in normative relations presumed to hold between free and rational persons as such. In interpreting your request as an address of such a (second-personal) reason, your addressee must presuppose that you are claiming a second-personal authority simply as a rational person with second-personal competence, which status you must assume he has also.¹⁵

When we think in these more general terms – of persons as "simply" free and rational – then it *does* follow that the foot-treader has a reciprocal, mutual authority over you; he has it in virtue of being, like you, free and rational. And, insofar as *all* persons are free and rational, all persons will be subject to Second-Personal moral reasons, and so universality follows.

If this is the route that Darwall takes, then the resemblance of his Second-Personal account to Neo-Kantianism is even more so. As we have seen,¹⁶ Neo-Kantians may take a similar line of defence in arguing that what my reasons commit me to is not the value of *my* particular personhood to me but the value of personhood *simpliciter*, or generally, such that it follows that your personhood – in virtue of being an instance of personhood generally – is also valuable to me. However, we have

¹³ See Ch.1, s.5.1.

¹⁴ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited in s.1, above.

¹⁵ Darwall 2006, p.258.

¹⁶ See Ch.1, s.5.2.

already seen the problems with this Generality argument, and they apply as well to the Darwallian case.

Firstly, there is no more reason to conclude that my being held accountable by you presupposes your authority over free and rational persons *generally* than there is that it presupposes your authority over *me* (only). Either explains the fact that you have the authority to hold me accountable just as well as the other: if you have a wide-ranging authority over all free and rational persons, then, in virtue of my being one such free and rational person, it will follow that you have authority over me; *or*, if you have circumscribed authority over just me, then it will still follow (tautologically) that you have authority over me. For all that has been said so far, either the general conclusion or the particular conclusion explains the fact that you have authority over me, just as well as the other. So, we have no reason to prefer the one over the other, and thus no reasons to conclude that second-personal reasons are necessarily reciprocal, and, thus, no reason to conclude that they are necessarily universal.

To this Darwall might respond, however, (as quoted, above) that when a second-personal reason is given it “might *most naturally* be presented and interpreted as grounded in normative relations presumed to hold between free and rational persons as such”¹⁷ rather than as holding between a second person and one particular first person over whom that second person has circumscribed authority. That is, Darwall might respond that the general conclusion is *more natural* than the particular one, and that we thereby *do* have reason to prefer the former to the latter, and therefore to conclude that second-personal reasons are indeed necessarily reciprocal and, therefore, universal. However, this strikes me as a weak argument.

Firstly, when it comes to authority, I doubt that the general conclusion is indeed more natural than the particular one (which may be a relevant difference between the Generality argument in this Darwallian context and that in the Neo-Kantian context¹⁸). This is because it seems to be part of our concept of authority – our *natural* use of the term – that authority is circumscribed, not general or wide-ranging.¹⁹ I have authority over the students enrolled in *my* courses (not all students); a mother has authority over *her* children (not all children); a sergeant has authority over *his* platoon (not all soldiers); a government has authority over *its* citizens (not all citizens); *etc.* That is, if the argument hinges on what’s ‘most natural’, then I doubt in the first place that it’s most natural to think that you have authority over, say, the foot-treader in virtue of his being simply a free and rational person; rather, it seems that you have authority to demand that he remove his foot from on

¹⁷ Darwall 2006, p.258 (emphasis added), as cited above.

¹⁸ See Ch.2, s.5.2.

¹⁹ Consider, for instance, the shared etymologies of “dominion” and “domain”.

top of yours in virtue of your being the victim of his big, careless feet²⁰ – that is, in virtue of his causing you harm *in particular*.

Secondly, even if it is more natural to draw the general conclusion than the particular one, it is, of course, well established that our ‘most natural’ inferences are not always the logically best ones. We drastically overestimate the likelihood of being eaten by a shark or dying in a plane crash, we are generally hopeless at the Wason selection task, and we are subject to all sorts of biases and other effects. Even if the general conclusion is the more natural one to draw, this isn’t a good guide as to whether it’s the *better* one to draw. And, I have argued that it *isn’t*, that there is some consideration *against* drawing the general conclusion rather than the particular one, for fear of committing a hasty generalization. To repeat the example from the previous chapter,²¹ if Sally scrapes your car, you should on that basis conclude that *Sally*, in particular, is a bad driver – not that all women, in general, are bad drivers. Likewise, if you give the foot-treader a second-personal reason, you should infer that you have authority over *the foot-treader* (only), not over all free and rational persons generally. As such, Darwall’s argument for Morality as Equal Accountability entails only that, say, you and the foot-treader share an authority over *the foot-treader* (only), which need not be mutual and reciprocal, and so second-personal moral reasons need not be universal.²²

1.2. The Analogical argument

Or, again in parallel to the Neo-Kantian case, Darwall may try a version of the Analogical argument in response to this problem.²³ That is, he may try to argue that consistency demands that, because you and I are alike (in being free and rational persons,²⁴ we might add), the fact that you and I have authority over me entails that you and I also have authority over *you*, and as you hold me

²⁰ Indeed, see Darwall’s own description of the second-personal response to the foot-treader: “The [second-personal] reason would not be addressed to him as someone who is simply in a position to alter the regrettable state of someone’s pain or of someone’s causing another pain [...]. It would be addressed to him, rather, *as the person causing gratuitous pain to another person* [...].” (2006, p.7; emphasis added). Part of my complaint against Darwall, which will germinate in this present chapter and be developed over the subsequent one, is that Darwall appears to abandon this *specificity* of second-personality in order to account for the mutuality (universality) of Second-Personal reasons. The result of this, as I will go on to argue, is that the sense in which such “Second-Personal” reasons are genuinely second-personal will become so watered down that it is ultimately lost.

²¹ See Ch.2, s.5.2.

²² As before (see Ch.2, s.5.2), note that this objection only claims that the universality of moral reasons isn’t *necessarily* entailed by Darwall’s argument for Morality as Equal Accountability; it leaves open the possibility that, *if* the general conclusion is employed, rather than the particular one, then the universality of moral reasons will follow. However, if Darwall’s argument for Morality as Equal Accountability is unable to account for the *necessary* universality of moral reasons, then it is no serious improvement upon the Hobbesian account, which is also able to account for the universality of moral reasons *if* it happens that all persons desire a non-nasty, non-brutish and long life (see Ch.2, s.3.2).

²³ See Ch.2, s.5.3.

²⁴ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited in ss.1 and 1.1, above.

accountable, I too may hold you accountable. And, again, if we extend this line of argument to all persons then universality will follow.

We should note that the Analogical argument in the context of Darwallian, second-personal reasons is actually more plausible than it was in the context of Neo-Kantian, first-personal reasons. This is because, in the Neo-Kantian context, there was no asymmetry or inconsistency to be resolved by the Analogical argument in the first place: I had my reasons because I valued my personhood, and you had your reasons because you valued your personhood.²⁵ In the Darwallian context, on the other hand, there does seem to be a genuine asymmetry to which the Analogical argument responds: I have authority over myself (by Pufendorf's Point), you have authority over me (because you lay a second-personal claim against me) – and, presumably, and plausibly, you have authority over yourself (for the same reasons I have authority over myself). So, as it stands, two people have authority over me – me, and you – yet only one person has authority over you: you. So, consistency would indeed seem to demand that I, likewise, have authority over you. As such, the Analogical argument is immediately more plausible in this present Darwallian, second-personal context than it was in the Neo-Kantian, first-personal context examined in the previous chapter.

This means that one key objection to the Analogical argument examined in the previous chapter will not be applicable here,²⁶ since the force of that objection was that there is no inconsistency to which the Analogical appeal to consistency responds. If I am correct that there *is* a genuine inconsistency involved in the present, second-personal case, however, then the Analogical appeal to consistency is indeed appropriate. I will argue, however, that the final objection examined in that previous chapter – namely, Korsgaard's complaint – will still find purchase in this present Darwallian context.

Korsgaard's complaint, recall, is that if the Analogical argument were to entail the universality of moral reasons,

[i]t would do it in the wrong way. It would show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that I have duties with respect to you, about you, but not that there are things I owe *to you*. But some duties really are owed *to others*: we may be obligated by others, I will argue, in much the same way that we may be obligated by ourselves.²⁷

Of course, this objection is addressed specifically to first-personal accounts, and effectively criticizes them for not being second-personal accounts, for not positing obligations owed "*to others*". So, it may appear, on the face of it, impossible that I could mount the same objection to Darwall's argument for Morality as Equal Accountability – which is, of course, a *second-personal* account. But,

²⁵ See Ch.2, s.5.3.

²⁶ See Ch.2, s.5.3.

²⁷ Korsgaard 1996, p.134, as cited in my Ch.2, s.6.3.

that is precisely what I will try to do. The objection will be that, if the Analogical argument works in this Darwallian context, then the account actually *isn't* a second-personal one. (And, of course, if the Analogical argument *doesn't* work, then we are left without an account of the universality of moral reasons.)

Korsgaard's complaint is that, if I am obligated, say, not to tread on your feet because I have a reason to be consistent (as the Analogical argument posits), then I am not really obligated *to you*; I am obligated only to *myself* – albeit it *concerning* you – in virtue of my independent reason to be consistent. So, the way around this problem is to suppose that my reason to be consistent is itself a *second-personal* reason, an obligation *to you*. That is: suppose I am obligated, say, not to tread on your feet because I have a reason to be consistent (as the Analogical argument posits) – and, suppose, I have this reason to be consistent because *you gave me a (second-personal) reason to be consistent*, you, say, commanded me to be consistent. If this is the scenario, then my obligations to you will not bottom out as obligations *to myself*, as Korsgaard complained. Instead, it would be the case that I am obligated *to you* to be consistent, and, since consistency demands that I have reciprocal authority over you, we would therefore share a mutual, reciprocal authority over each other.

In the Darwallian context, the Analogical contention is that consistency entails that I have the authority to hold you accountable as you have the authority to hold me accountable (and as I have the authority to hold myself accountable). And, as we have established, above, if this is to evade Korsgaard's complaint, it must be that my reason to be consistent is a *second-personal* reason – that is, that *you* commanded me, second-personally, to be consistent. That is, you command me, say, to remove my foot from on top of yours, thereby presupposing (as Pufendorf pointed out) that I share the same authority over myself as you do; you also command me to be *consistent*, thereby entailing that I have the reciprocal authority to hold you accountable, which (as Pufendorf pointed out) presupposes that you share the same authority over yourself as I now do. Thus, we have mutual authority, and if we extend the same line of argument to all persons, universality will follow. However, I will show that this proposal cannot hold for the Darwallian Second-Personal account.

As a first, but small, objection to this Analogical proposal, it is surely a little strange to think of having a reason to be consistent *because a second person commanded you to do so*. But, then again, this is apparently what, for instance, animal rights activists are doing when they point out to omnivores the extreme affection they bear for their pets and yet do not extend to other (equally intelligent) animals such as cows and pigs: the charge is to *be consistent* and realize that pigs are not relevantly different from dogs. So, this Analogical argument can perhaps cope with a little strangeness. But, I

have a deeper objection in mind. Despite appearances, this Analogical argument actually will not yield second-personal reasons but *first-personal* ones, and as such falls foul of Korsgaard's complaint. I explain this tricky point, below.

The Analogical argument has proposed this much: you command me, say, to remove my foot from on top of yours, thereby presupposing (as Pufendorf pointed out) that I share the same authority over myself as you do; you also command me to be consistent, thereby entailing that I have the reciprocal authority to hold you accountable, which (as Pufendorf pointed out) presupposes that you share the same authority over yourself as I now do. Now, notice that my resultant authority to hold you accountable (by way of your commanding me to be consistent) entails that *you* have obligations to *me*. Just as your having authority to hold me accountable means that I have the obligation to you not to, for instance, stand on your foot, my having the authority to hold you accountable means that you have an obligation to me (for instance, not to stand on my foot). So far, so good: I have an obligation not to stand on your foot (for which you have the authority to hold me accountable) and you have an obligation not to stand on my foot (for which I now have the authority to hold you accountable); that is, we have mutual Second-Personal obligations. However, notice how this latter obligation – your obligation to me – has come about. According to this Analogical argument, it has come about because *you* commanded me to be consistent. (So, since you had the authority over me to command that I not step on your foot, consistency entails that I now have the reciprocal authority over you.) That is to say, I have the authority to hold you accountable – meaning that you are obligated to me – because *you* commanded so. *You* have an obligation because *you* commanded so.

The upshot of this, of course, is that your resultant obligation not to stand on my foot is just a *first-personal* obligation – a reason you have given yourself – and not, as it first appeared, a second-personal obligation. If you have an obligation to me because I have the authority to hold you accountable, and if I have the authority to hold you accountable because you have commanded me to be consistent, then, although your obligation has taken a detour through me, it is nevertheless a *first-personal* obligation – something like asking a friend not to let you have more than two drinks. Just as Korsgaard complained, it turns out that you have an obligation “with respect to” me, “about” me, but not *to* me.²⁸

So, this Analogical argument, despite initial appearances, actually falls foul of Korsgaard's complaint. An argument such as this one can only achieve universality by ultimately appealing to first-personal

²⁸ Korsgaard 1996, p.134, as cited above.

reasons, rather than the requisite second-personal ones.²⁹ And, if this Analogical argument fails, and if we have no (non-question-begging) reason to accept the Generality argument (or, indeed, positive reasons against accepting it),³⁰ then we have no reason to think that second-personal reasons are indeed mutual and ultimately universal. As it stands, the Darwallian Second-Personal account offers no advantage over the Neo-Kantian account.³¹

1.3. Pufendorf's Point

I have argued, above, that Darwall's argument for Morality as Equal Accountability does not work. Even if Pufendorf's Point entails that your having authority over me and holding me accountable entails that I also have the authority to hold myself account, it does not on that basis follow that I also have the reciprocal authority to hold *you* accountable.³² That is, I have argued that Darwall's argument for Morality as Equal Accountability fails to go through *even if* we accept Pufendorf's Point. However, I would now question whether Darwall can accept Pufendorf's Point at all. While Pufendorf's Point *does*, independently, strike me as correct, I am not sure that it is consistent with *second-personality*.

Recall that Pufendorf's Point is the claim that genuine moral accountability – as opposed to mere behavioural modification by negative reinforcement – presupposes that the agent must hold *him-/herself* accountable also.³³ An agent who is held genuinely morally accountable “judge[s] [him]self worthy of some censure, unless [he] conforms to a prescribed rule”³⁴ and “acknowledge[s] of himself that the evil, which has been pointed out to [him], falls upon him justly”.³⁵ But, we should ask, why is it that the status of *your* holding an agent accountable as either a case of genuine moral accountability or not is determined by *the agent's* judging the punishment to “[fall] upon him justly” or otherwise? That the *agent* is the judge here suggests to me that any reason to which a second person might hold an agent accountable must first be endorsed by the agent; if the agent fails to endorse such a reason, fails to judge that the subsequent punishment would fall upon him justly, then any subsequent punishment is not genuine moral accountability but merely behavioural

²⁹ This may be objectionable. I may indeed have a second-personal reason to hold you accountable – but not a reason given to me by *you*; I may be commanded by some third party to hold you accountable, and thereby have a second-personal reason to hold you accountable, which does not boil down to a first-personal reason. However, note that this would not be the result of any *analogy*: the Analogical argument claims that, as you hold me accountable, I too have the authority to hold you accountable. Without the force of this analogy between me and you, there are no grounds for supposing that some third party gives me a (second-personal) reason to hold you accountable.

³⁰ See s.1.1, above.

³¹ See Ch.1, s.5.

³² See ss.1.1 and 1.2, above.

³³ See Ch.1, s.6.1.2.

³⁴ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111, as cited in my Ch.1, s.6.1.2.

³⁵ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, pp.111-112, as cited in my Ch.1, s.6.1.2.

modification by negative reinforcement. And, that the *agent* is the final arbiter in this way implies that the reasons involved are not second-personal reasons at all, but rather *first-personal* ones. I, the first person, decide what I may or may not be held accountable for doing; any authority you, as the second person, subsequently have to hold me accountable is due to my, the first person's, having made it so.

Indeed, this is precisely the way Korsgaard interprets Pufendorf's Point. As she was (partially) quoted, in the preceding chapter: "So Pufendorf and Hobbes were right. [...] It is the fact that we *command ourselves* to do what we find it would be a good idea to do" "that obligates us to perform it."³⁶ That is, Korsgaard reads Pufendorf's distinction between genuine moral accountability and mere behavioural modification through negative reinforcement as a distinction between (respectively), on the one hand, cases in which agents 'command themselves' – that is, give themselves *first-personal* reasons, to which you too may hold them genuinely morally accountable – and, on the other hand, cases in which agents do *not* 'command themselves' as you do and in which your subsequent punishment is therefore experienced as imposed from outside, by an illegitimate authority, and thus as coercive. Moreover, notice that Korsgaard makes this remark of Pufendorf in the same breath as Hobbes.³⁷ Clearly, she reads Pufendorf's Point as much closer to a Hobbesian, first-personal account than Darwall does. Indeed, recall that this is how Pufendorf was able to cope with the Divine Command theorist's problem that "[God] can apparently make *anything* right or wrong"³⁸ simply in virtue of commanding it: as the Hobbesian Sovereign only has legitimate authority over his/her subjects by way of those agents' *submitting* their wills to the Sovereign, so the Pufendorfian agent is protected from God's potentially making illegitimate demands of one: if I don't consent to it (first-personally), then God can't legitimately demand it of me. So, Darwall is correct that Pufendorf's Divine Command theory is unstable³⁹ – but, not because it resolves itself into Darwall's Morality as Equal Accountability. Rather, Pufendorf's Divine Command theory is unstable because it devolves into a *first-personal* account.⁴⁰

This leads Korsgaard to claim, in her review of *The Second-Person Standpoint*, that

³⁶ Korsgaard 1996, p.105, as cited in my Ch.1, s.6.1.2.

³⁷ I omitted this from my previous citation of this quote. See Ch.1, s.6.1.2.

³⁸ Korsgaard 1996, p.23 (emphasis added), as cited in my Ch.1, s.6.1.2.

³⁹ See my Ch.1, s.6.

⁴⁰ I should point out that the complicated scenario I presented, above (s.1.2), in support of the Korsgaardian (1996, p.134) complaint that reasons resulting from an Analogical argument – even in the Darwallian case – are "about you, but not [...] owe[d] to you" is just a Pufendorfian scenario: it says that I only have the authority to hold you accountable when *you* give me a reason to hold you accountable (by commanding me to be consistent). As above, any authority I have over you derives *from you*. And, as I complained in that section, this is a *first-personal* matter rather than a second-personal one.

Since in order to address a second-personal reason to you, and so to obligate you, I have to assume that you have the capacity to obligate yourself, it would be natural to suppose that the capacity to obligate oneself is the prior notion.⁴¹

That is to say, “it would be natural to suppose” on the basis of Pufendorf’s Point that Darwall’s Second-Personal (“Second-Personal”) account is grounded not in the second-personal but in the *first-personal*. Likewise, Jay Wallace notes in his review of *The Second-Person Standpoint* that Darwall’s appeal to Pufendorf’s Point “change[s] the subject” from the second-personal to the first-personal:

If I have reason to do something in virtue of your authority to demand it of me, then what I need to be able freely to acknowledge is precisely *your* authority in this matter. To the extent I see the demand as one I have reason to comply with because I have the authority to impose it on myself, it seems I have changed the subject; the normativity of the demand is no longer based in your distinctive authority over me but in my own reflexive authority to make laws for myself.⁴²

Thus, since Pufendorf’s Point apparently grounds authority in the *first person*, rather than in the second person, it appears to be fundamentally inconsistent with Darwall’s Second-Personal approach. Above,⁴³ I argued that, even if we accept Pufendorf’s Point, Morality as Equal Accountability does not follow. Here, I have argued, along with Korsgaard and Wallace, that Pufendorf’s Point actually can’t, in the first place, be employed to support Morality as Equal Accountability, since it is fundamentally inconsistent with Darwall’s Second-Personal approach.

In response, it might be objected that this conclusion is harsh, and that a more charitable reading will construe Darwall as trying to develop his Second-Personal account on the basis of a first-personal (Pufendorfian) one. This would be to continue a trend we have seen emerge in the preceding chapter: Hobbes derives his moral account on the basis of a non-moral one;⁴⁴ Neo-Kantians derive a universal (categorical) account on the basis of a non-universal (hypothetical) one;⁴⁵ so, likewise, one might think that Darwall attempts to derive his Second-Personal account on the basis of a first-personal one. Indeed, this seems to be Fleischacker’s reading of Darwall’s account, according to which “I [...] grant you second-personal authority over me and oblige myself, in the course of feeling obliged by you, to remove my foot.”⁴⁶

However, a reading like Fleischacker’s is mistaken. With regard to the Neo-Kantian strategy of deriving the universal Categorical Imperative from the non-universal Hypothetical, Darwall

⁴¹ Korsgaard 2007, p.11.

⁴² Wallace 2007, p.35.

⁴³ See ss.1.1 and 1.2, above.

⁴⁴ See Ch.1, s.3.

⁴⁵ See Ch.1, s.5.

⁴⁶ Fleischacker 2009, p.120.

believes [the Neo-Kantian claim] that autonomy and the Categorical Imperative in its various formulations are transcendental presuppositions of the practical [i.e. hypothetical, first-personal] standpoint [...] to be mistaken. In my view, there is no way of deriving these ideas outside of the [S]econd-[P]erson[al] standpoint [...].⁴⁷

That is, Darwall would not be amenable to this proposed charitable strategy of reading his argument as forming part of the trend mentioned above, and trying to derive his Second-Personal account on the basis of Pufendorf's first-personal one. Fleischacker's first-personal reading, according to which "I [...] grant you second-personal authority over me and oblige myself, in the course of feeling obliged by you, to remove my foot",⁴⁸ tries to derive Second-Personality outside of the Second-Personal standpoint, and as such would be rejected by Darwall. It is correct that Darwall does present his account as the evolution of another account, namely, Divine Command theory⁴⁹ – but note, of course, that Divine Command theory is itself a *second-personal* account,⁵⁰ not a first-personal account, so Darwall is true to his claim that "there is no way of deriving these ideas outside of the [S]econd-[P]erson[al] standpoint". If this is the case, however, then Pufendorf's Point, with its first-personal foundation, is a real problem for Darwall. (Indeed, as I will go on to show as I return numerous times to Darwall's insistence that there is no way of deriving Second-Personality outside of the Second-Personal standpoint, it is not only Pufendorf's Point that is a problem for his account.)

There is one further way in which Darwall might respond to my charge that his employment of Pufendorf's Point is inconsistent with his Second-Personal account. He may retort that what does or does not fall upon an agent justly (to paraphrase Pufendorf⁵¹) is not – as I have suggested in this section – a matter of a first-person's "grant[ing]" second-personal authority to a second,⁵² and that a second person's authority to hold a first person accountable does not derive from that first person's endorsement of it. Nor, he may continue, does the second person's authority arise from within him-/herself. This would be a misreading of Pufendorf's Point. Rather, the first person holds him-/herself accountable from the *same, shared* point of view from which the second person also holds the first person accountable: the *moral point of view*. Indeed, this is precisely the solution that Darwall⁵³ offers in his response to *The Second-Person Standpoint* reviewers Korsgaard,⁵⁴ Wallace⁵⁵ and Watson⁵⁶ – although I take it to be already present in that original book. If the agent holds him-

⁴⁷ Darwall 2006, p.42, fn.12.

⁴⁸ Fleischacker 2009, p.120; as cited above.

⁴⁹ See Ch.1, s.6.

⁵⁰ See Ch.1, s.6 and the matrix of positions in s.1.3.

⁵¹ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, pp.111-112.

⁵² Fleischacker 2009, p.120, as cited above.

⁵³ Darwall 2007.

⁵⁴ Korsgaard 2007, as cited above.

⁵⁵ Wallace 2007, as cited above.

⁵⁶ Watson 2007.

/herself accountable from the moral point of view, then the agent isn't – as I describe it, above – the 'final arbiter' of what he/she can and can't be held accountable for; rather, the agent is constrained in what he/she can and can't hold herself accountable for – by the moral point of view. Likewise, we avoid the problem that threatened Divine Command theory: what's stopping God (the second person) from demanding *anything at all* from me?⁵⁷ The moral point of view is. So, when we understand Pufendorf's Point in this way, it doesn't boil down to a first-personal account, and thus isn't inconsistent with Darwall's Second-Personal account. However, I set this possible response aside for now, to pick it back up again, below, when I turn to the question of validity.⁵⁸

In the following section, I aim to deal Darwall's employment of Pufendorf's Point one more blow. Where I have argued, in this section, that Darwall's Second-Personal account is incompatible with the first-personal basis of Pufendorf's Point, I argue in the next section that the Divine Command theory *is* compatible with Pufendorf's Point. If this is the case, then there is no instability inherent in Pufendorfian Divine Command theory, and thus no instability that Darwall's Morality as Equal Accountability resolves.

1.4. The Divine Command theory advantage

Darwall argues for his thesis of Morality as Equal Accountability by way of uncovering an inconsistency between Divine Command theory and Pufendorf's Point, that any second person's (such as God's) holding a first person genuinely morally accountable presupposes that that first person holds him-/herself accountable too.⁵⁹ In this section, I argue, *contra* Darwall, that Divine Command theory is indeed compatible with Pufendorf's Point.

My argument depends on the idea that God has the authority to hold me accountable – whereas, contrary to Morality as Equal Accountability,⁶⁰ you don't – because God is omniscient and knows everything about me. You, on the other hand, have only partial insight into my motivations, psychological history, *etc.* So, you are not in a position to judge me, whereas God is. I think this is one very common understanding of Divine Command theory – so much so that it even makes its way into pop songs. It's a very common theme of rap songs in particular, and (taking my cue from Darwall's delightful Bob Dylan⁶¹ and Alanis Morissette⁶² references) it appears very explicitly as the chorus of this house music song by Armand van Helden:

⁵⁷ See Ch.1, s.6.

⁵⁸ See s.2.3.3, below.

⁵⁹ See my Ch.1, s.6.

⁶⁰ See Ch.1, s.6.2.

⁶¹ See Darwall 2006, p.74, fn.19, and p.135, fn.30.

⁶² See Darwall 2006, p.73, fn.19.

You don't even know me,
you say that I'm not livin' right
You don't understand me,
so why do you judge my life?⁶³

On the other hand, since God – being omniscient – does “know me”, he *does* have the authority to “judge my life”.

Furthermore, the Divine Command theorist might argue, I too have the authority to “judge my life”, since I too possess the requisite knowledge of my life – again, unlike you. You are barred from judging me because you do not have adequate knowledge of me – whereas I, being me, have all this knowledge available to me (or, at least, have a lot more of it available to me than you do). Even if you don't, *I* know why, for instance, I really dived into the pool to save the drowning child – as does God. So, both God and I have the authority to hold me/myself accountable, whereas you do not, since you lack the requisite knowledge.⁶⁴ This is compatible with Pufendorf's claim that a second person's holding a first person accountable presupposes that that first person also holds him-/herself accountable.

If it is the case that, *contra* Darwall, the Divine Command theory is compatible with Pufendorf's Point, then there is no purported instability between a Divine Command account and Pufendorf's Point to provide impetus towards Darwall's thesis of Morality as Equal Accountability. Not only does this account leave us without cause to embrace Morality as Equal Accountability, it is, moreover, incompatible with Morality as Equal Accountability: you do *not* have equal authority to hold me accountable as I have and as God has, since you lack the requisite knowledge of me and my circumstances. Far from entailing Morality as Equal Accountability, Pufendorf's Point seems actually to be a better fit for Divine Command theory than it is for Darwall's Second-Personal account.

Darwall, however, would flatly reject the suggestion that authority could be “epistemic”⁶⁵ – that is, grounded in the knowledge of facts – as I have assumed, above. According to Darwall,

the authority to address practical reasons can take forms that are quite different from the epistemic authority that is presupposed either by theoretical reason-giving or by other forms of practical reason-giving, like advice, where the reasons are not second-personal.⁶⁶

⁶³ Van Helden 1999.

⁶⁴ This may be objectionable. It may be that, by some combination of brain-imaging and Truman-show-type surveillance, a non-omniscient second person *does* possess all the knowledge about me in order to have the authority to hold me accountable. But, even if this is possible, it would still fall short of the *necessary* universality of authority, since presumably not all second persons would have access to this knowledge. As such, the account would fall short of establishing the necessary universality of moral reasons as second-personal reasons.

⁶⁵ Darwall 2006. The claim is made several times throughout the book. See, for instance, pp.12, 57, 123-125 and 233-234.

⁶⁶ Darwall 2006, p.59.

Darwall claims that a second person's giving a first person an instruction, and subsequently holding that first person accountable, *because the second person knows best (better)* is not – to borrow a distinction from Hobbes – for the second person to *command* the first, but merely for the second person to *counsel*, or give advice to, the first.⁶⁷ As Darwall cites Nomy Arpaly as saying, “Just because I know more than you doesn't make me boss.”⁶⁸ On Darwall's account, the authority to counsel is epistemic, but the authority to genuinely issue a command is grounded not in epistemic authority, but in one's *dignity*. You have the authority to give the foot-treader a reason to remove his foot from on top of yours not in virtue of your extensive training as a podiatrist nor your accompanying knowledge of the physical injuries that foot-treading can cause, but simply in virtue of your dignity as a person.

I wonder, however, whether Darwall's account can cope with his rejection of epistemic authority. If it is the case that an individual possesses the authority to hold an agent accountable merely in virtue of that individual's *dignity*, then it follows that *any* (dignity-possessing) individual will have the authority to hold any agent accountable. But, this surely cannot be right: if you don't know that I was obligated to ϕ , and you don't know that I failed to ϕ – that is, if you don't possess the relevant *knowledge* – then of course you don't have the authority to hold me accountable for my failure to ϕ .⁶⁹ I don't see how Darwall's account of accountability – or, indeed, any account of accountability – could get by without this basic epistemic criterion. Fleischacker, likewise, urges us to “question whether practical reason can be as wholly independent of the world as Darwall suggests”,⁷⁰ writing:

If you ask me not to step on your foot, it is only because I know that that *hurts* that I see the reasonableness in your request of me not to do it – and that I, therefore, [...] [have a reason] to remove my foot. If you were instead to issue the command ‘Move your foot!’ quite arbitrarily, when my foot was nowhere near yours and doing nothing that I could reasonably see as disturbing you, I would ignore you, or accuse you of arrogance. So the facts of the world are going to have to play a role in morality [...].⁷¹

If Fleischacker is correct, as seems plausible, that “the facts of the world are going to have to play a role in morality”, then it follows that individuals' *knowledge* of those facts of the world is “going to have to play a role in” their authority to give, and hold agents accountable to, moral reasons – that is, their authority must be (at least partially) *epistemic*.

I return to the question of epistemic authority several times over the remainder of this chapter and subsequent ones. For now, however, I conclude that it is at least *prima facie* plausible that, on the

⁶⁷ Hobbes 1642/1983, Ch.XIV, s.1, as cited by Darwall 2006, pp.12-13, fn.25. See also Hobbes 1651, pt.2, Ch.25.

⁶⁸ Arpaly (personal communication), as cited by Darwall 2006, p.15, fn.29.

⁶⁹ Indeed, I go on to develop an account of accountability along precisely these lines in Ch.5, s.1.2.1.

⁷⁰ Fleischacker 2009, p.120.

⁷¹ Fleischacker 2009, p.120.

Divine Command theory, God has the authority to give me reasons and hold me accountable for compliance because of his epistemic status – namely, his omniscience. And, likewise, it follows that I too have the authority to hold myself accountable, since I also have epistemic authority with regards to my own life. So far, this Divine Command approach is all consistent with Pufendorf’s Point: that an agent’s being held genuinely morally accountable (by God) presupposes that that agent holds him-/herself accountable. On the other hand, because “[y]ou don’t even know me” and “don’t understand me,” you are in no position to “say that I’m not livin’ right” and to “judge my life”.⁷² That is to say, *contra* Morality as Equal Accountability, you do *not* possess the authority to hold me accountable, since you do not have the epistemic standing to do so: you don’t know enough about me, whereas God and I do. Divine Command theory, therefore, is compatible with Pufendorf’s Point, whereas, as I have argued (above⁷³), Morality as Equal Accountability is not. It is far from clear to me, therefore, that there is any instability inherent in Pufendorf’s Divine Command theory, nor that Darwall’s Second-Personal account is capable of resolving it.

In preceding sections, I have argued that *even if* Pufendorf’s Point is correct, Morality as Equal Accountability does not follow on the basis of, say, your giving the foot-treader a second-personal command. I examined both the Generality⁷⁴ and the Analogical⁷⁵ arguments in response to this objection, and countered them both. I then moved on to the deeper critique that Darwall actually can’t employ Pufendorf’s Point in arguing for his Second-Personal account, as Pufendorf’s Point turns out to be fundamentally inconsistent with Darwall’s Second-Personal account,⁷⁶ even though the Divine Command theory is able to accommodate it. As such, we are left without an argument for the mutual, reciprocal – and, ultimately, universal – nature of second-personal reasons. Therefore, contrary to initial appearances,⁷⁷ Darwall’s Second-Personal account fails to account for the universality of moral reasons.

I turn now to consider whether Darwall’s argument for Morality as Equal Accountability can even get off the ground. Even if we accept Pufendorf’s Point that *if* a second person holds a first person accountable, then that first person must also hold him-/herself accountable – why should we accept that the second person has the authority to hold the first person accountable in the first place? Again, when that second person is – as it is in the case of Divine Command theory – *God*, who is omniscient, it is easy (or, at least, easier) to see why such a second person might have the authority

⁷² Van Helden 1999, as cited above.

⁷³ See s.1.3, above.

⁷⁴ See s.1.1, above.

⁷⁵ See s.1.2, above.

⁷⁶ See s.1.3, above.

⁷⁷ See my Ch.1, s.6.2.

to hold me accountable. But, when that second person is – as it is on Darwall’s Second-Personal account – just *you*, then (with all due respect) it is far from clear why or how you might have such authority. Who are you to tell me what to do? I consider this line of objection, below.

2. The sceptical problem

The first general problem with Darwall’s account, the *non sequitur* problem, questioned how a first and second person could be mutually accountable – how, for instance, given that you are able to give the foot-treader a reason to get off your foot by demanding that he do so, the foot-treader would likewise be able to give you a reason.⁷⁸ The second general problem with Darwall’s account – the sceptical problem – takes a step back and questions whether you could even give the foot-treader a reason to begin with. Is this really what’s going on when you say, “Get off my foot!” – that you’re literally *giving* the foot-treader a reason that wasn’t there before? Or, are you – as seems more plausible – simply drawing his attention to a reason that was there all along, the reason not to cause others undue harm?

Recall that this is precisely how Darwall opens his argument for his Second-Personal account, by presenting two different ways in which we might respond to the foot-treader: drawing his attention to the independent, pre-existing reason not to cause you pain, or literally giving him a reason to remove his foot from on top of yours by commanding him to do so.⁷⁹ Interestingly, note that this *isn’t* presented as a mere thought experiment – Darwall does think that we face a real choice as to which kind of reason we might invoke in our response to the foot-treader: the second-personal reason, or the (equally real) world-regarding reason. That Darwall concedes the existence of non-second-personal reasons is made clear on a number of occasions, for instance:

It is entirely possible to give people reasons for acting that are not second-personal [...].⁸⁰

There may of course be non-second personal reasons for others to do what we also have a right to demand their doing [...].⁸¹

Of course, many of the reasons that ground claims of wrong and obligation are not themselves second-personal.⁸²

Since Darwall allows that there are non-second-personal reasons,⁸³ matters are made all the more confusing as to why we should believe that there are such things as second-personal reasons over

⁷⁸ See s.1, above.

⁷⁹ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁸⁰ Darwall 2006, p.24.

⁸¹ Darwall 2006, p.19.

⁸² Darwall 2006, p.99.

This third quote raises an interesting question about the relationship between second-personal and non-second-personal reasons. I pick this up in s.2.3.1, below.

and above non-second-personal reasons. Can't non-second-personal reasons just do all the work that second-personal reasons would be doing anyway?⁸⁴

I am sure that Darwall's answer to this question would be "no". Non-second-personal reasons can't just replace second-personal reasons; second-personal reasons fulfill a role that non-second-personal reasons cannot. In particular, the claim seems to be that only second-personal reasons can account for the distinctive character of *moral* reasons. For instance:

[M]orality exists only through the possibility of a second-personal address on which one rational agent can give another a distinctive kind of reason he wouldn't otherwise have had [...].⁸⁵

So, even though there exist non-second-personal reasons, the claim is apparently that such reasons are *non-moral* reasons.⁸⁶ Only second-personal reasons – in particular, only Darwallian Second-Personal reasons – can account for the distinctive character of moral reasons.⁸⁷

But, if moral reasons have normative force and are universal,⁸⁸ and, if I am correct that Darwall's argument for Morality as Equal Accountability fails to establish the universality of moral reasons,⁸⁹ then Second-Personality cannot account for the distinctive character of moral reasons (in particular, their universality), and Darwall's Second-Personal account occupies no privileged position over non-second-personal accounts. As such, we are left, as before, without any reason to believe that Darwallian Second-Personal reasons actually exist. I develop this argument, below, claiming that we

⁸³ Darwall does, however, appear in a more recent paper (2010a) to distance himself from this claim, although his precise position remains less than explicit.

⁸⁴ Indeed, in Ch.5 my argument will be precisely that Real, world-regarding reasons can do the same job as Second-Personal reasons – and do it better.

⁸⁵ Darwall 2006, p.107.

⁸⁶ If non-second-personal reasons are non-moral reasons, however, it is an interesting question as to what kind of non-moral reasons all of the non-second-personal reasons in the preceding three quotations, above (Darwall 2006, pp.24, 19 and 99), might be. Darwall even describes such a non-second-personal reason in the third of these quotations as an "obligation" – and what is this (non-second-personal) obligation if not a *moral* obligation? A prudential obligation?

⁸⁷ Interestingly, the existence of non-second-personal reasons is what Darwall identifies as the main point of contrast between his account and Korsgaard's. For instance, he writes:

Korsgaard [...] seems to think that all reasons involve the making of claims (whereas I restrict this to second-personal reasons): "To say that you have a reason is to say something *relational*, something which implies the existence of another, at least another self. It announces that you have a claim on that other, or acknowledges her claim on you" [...]. In my view, this conflates the general category of normative reasons with the distinctively second-personal reasons that figure in justifying ourselves to one another. (Darwall 2006, p.102, fn.23, citing Korsgaard 1993/1996, p.301, and referencing pp.275-276)

And, early on in the book:

It is a fundamental point of agreement between this claim of Korsgaard's and the outlook I defend in this book that moral obligations are irreducibly second-personal in this way (although unlike Korsgaard, I fail to see why all reasons for acting must be shareable [i.e. second-personal] in this sense). (Darwall 2006, p.11)

⁸⁸ See Ch.1.

⁸⁹ See s.1, above.

ought to be sceptical of the existence of second-personal reasons at all. As such, Darwall's argument for Morality as Equal Accountability doesn't even get off the ground.

2.1. Unfounded assumptions

Recall Fichte's Point: that the address of any second-personal reason presupposes the common dignity (i.e. second-personal competence and second-personal authority) of the first and second persons.⁹⁰ And, the argument goes, since the first and second person must share an equal dignity (if second-personal address is to be possible), then Morality as Equal Accountability is entailed. In section 1 of this chapter, I argued that this actually doesn't follow, that second-personal address actually doesn't entail the mutuality of Darwall's Second-Personal account. In this section, I will question whether that argument for Morality as Equal Accountability can even get off the ground. Fichte's Point is that *if* we address any second-personal reason, then we are committed to equal dignity. But, why suppose that we address any second-personal reasons in the first place?

I have already drawn several analogies between Darwall's argument for Morality as Equal Accountability and the Neo-Kantian argument. The Neo-Kantian argument is a transcendental deduction.⁹¹ A transcendental deduction

begins, namely, with some conception of possible experience that is so basic that it is very difficult to subject it to s[c]eptical doubt without undermining the interest of the doubts themselves. Starting with a minimal and fundamental conception of experience, the transcendental philosopher argues for our entitlement to employ other concepts by showing them, and even their instantiation within experience, as a condition for the possibility of even this minimal starting point.⁹²

Although Darwall does not describe his argument in this way, he appears to be mounting a transcendental deduction of his own.⁹³ As the Neo-Kantian argues that my having any reasons at all commits me to the value of my personhood and therefore the value of your personhood,⁹⁴ Darwall argues that your addressing any second-personal reason to me at all commits us to my authority over me and therefore to my reciprocal authority over you too. The difference, however, is that, where the Kantian begins with "some conception of possible experience that is so basic that it is very difficult to subject it to s[c]eptical doubt", Darwall does not. Where the Kantian begins with the "minimal starting point" of the fact of my having reasons (such as, for instance, my reason to complete my dissertation⁹⁵), Darwall begins his transcendental deduction, rather, with the

⁹⁰ See Ch.1, s.6.2.1.

⁹¹ See Kant 1785, Ch.III.

⁹² Wood 2006, p.67.

⁹³ Brewer (2008, p.164) also reads Darwall's argument as a transcendental deduction.

⁹⁴ Or, according to the Generality argument: commits me to the value of personhood *simpliciter*, and therefore to your personhood, since you are a person. See Ch.1, s.5.2.

⁹⁵ See Ch.1, s.5.

assumption that *you give me a second-personal reason*, and then attempts to show that such a giving of a second-personal reason presupposes our equal dignity. But, the fact that you could give me a second-personal reason is far from obvious or beyond sceptical doubt to me. As Metz puts the point, “it is hard to see how the bare fact of being commanded to do something could be normative; that is, provide an overriding, categorical reason for action.”⁹⁶ Recall that giving a second-personal reason is not merely a case of drawing the first person’s attention to a reason that already exists; rather, it is the case of a first person coming to have a reason in virtue of the second person’s saying so.⁹⁷ But, who are you to tell me what to do?

Darwall’s argument for Morality as Equal Accountability starts off assuming that half the work has already been done for it, that it has already been demonstrated that there *are* such things as second-personal reasons, and that now all that remains to be done is for it to be shown that such reasons are mutual (universal). The shaky foundation of this argument is, I think, masked by the way that Morality as Equal Accountability is set up as being entailed by a proper analysis of Divine Command theory.⁹⁸ When God occupies the role of the second person, as he does on the Divine Command account, there is nothing presumptuous about the suggestion that a second person – when that second person is *God* – could give a first person (you or me) reasons. That seems to be entailed just by the fact that he is God. (Consider how the most common objection to Divine Command theory is not: “But who does God think he is to tell me what to do like that?”, but rather: “Yes, but God doesn’t really exist.”) But, then, when we move away from the Divine Command account and embrace instead Morality as Equal Accountability, there does seem to be a real question as to whether you – just ordinary, mortal you – could obligate me just in virtue of your saying so. There are, to put it mildly, some relevant differences between you and God, ones that make it less than obvious that you could possess the same authority as God to obligate others. Who are you to tell me what to do? So, why should we accept that there are such things as second-personal reasons? Until we have an answer to this question – and not just any answer, but one that “is very difficult to subject [...] to s[c]eptical doubt”⁹⁹ – Darwall’s transcendental deduction of Morality as Equal Accountability can’t even get off the ground.

2.2. Second-personal intuitions

Are there such things as second-personal reasons? As a useful starting point, let us consider our intuitions on this score.

⁹⁶ Metz 2013, p.87.

⁹⁷ See Ch.1, s.1.2.

⁹⁸ See Ch.1, s.6.

⁹⁹ Wood 2006, p.67, as cited above.

There is, of course, a very common practice of (for want of a more neutral term) exchanging reasons: you ask me to get off your foot; I tell you to get out of the way; you complain that you were here first; I tell you not to act like a child; you curse me for being so rude; *etc.* But, this isn't enough to conclude that there are such things as *second-personal* reasons. Second-personal reasons, recall, are reasons *given* by the second person to the first person, reasons that did not exist independently of that second-personal relationship between addressor and addressee. And, the fact that we *talk about* reasons doesn't entail that we're *giving* each other reasons we "wouldn't otherwise have had"¹⁰⁰ in so talking about them. So, is it true that, but for your *telling me* to get off your foot, I wouldn't have had a reason to do so?

I think not. I think I would still have had a reason to remove my foot from on top of yours, even if you had stoically said nothing and even if you had not even noticed that I was standing on your foot. If I had become aware by means other than your telling me to get off your foot that my foot was on yours (perhaps I gradually become aware of the uneven surface beneath my foot), I would still have felt the same guilt I would have felt had you addressed me instead. I would, for instance, probably still have spoken an apology, although perhaps more murmured in this case of non-address by you. So, it seems that I *would* still have had a reason to remove my foot from on top of yours even if you had not second-personally told me to do so. And, of course, Darwall will admit as much, since he concedes that there are non-second-personal reasons – and even concedes in particular that there are non-second-personal reasons not to step on people's feet.¹⁰¹ Although this doesn't establish an intuitive case *against* second-personal reasons, it doesn't do anything *for* them either. When it comes to foot-treading, at least, my intuitions are inconclusive.

I do, however, have intuitions in favour of the second-personal in at least one other domain: namely, what we do with our bodies, and, in particular, sexual consent.¹⁰² Here, it *does* seem intuitively to be the case that my say-so is the ultimate determinant of what you are or are not allowed to do to me (and of what I am and am not allowed to do to myself). If I say you are not allowed to kiss me or that you are not even allowed to touch me, you are absolutely *not* allowed to do so – regardless of whether I'm smaller than you, employed by you, or even married to you. "No" means: *no*. Likewise, if I say you *are* allowed to douse me with glue and cover me in feathers, well, then, that's cool too. Or, less facetiously, if I say you're allowed to inflict (a degree of) physical pain on me – something we

¹⁰⁰ Darwall 2006, p.107, as cited in s.2, above.

¹⁰¹ See s.2, above, and see Darwall's (2006, pp.5-7) description of the foot-treader scenario, as cited in my Ch.1, s.1.2.

¹⁰² Interestingly, in a more recent paper by Darwall (2010a), where he again presents his foot-treader scenario, it is this time framed in terms of consent: "you would have also to think that you have as right-holder some standing or authority to claim or demand that people not step on your feet without your consent" (p.262).

would ordinarily think you are *not* allowed to do – then, in virtue of my consent, you *are* allowed to do so. Wallace gives a less exciting medical example:

For example, though you would ordinarily wrong me by cutting into my gouty toe with a scalpel, I can render this form of treatment permissible by choosing to authorize it.¹⁰³

That is, in this domain, my say-so appears to have greater weight; I as a second person have the authority to *override* those (non-second-personal) reasons to which you as a first person are normally subject. That is, I can give you a *second-personal* reason.

However, I think we need to reflect on these second-personal appearances a little bit more.

Although I may be able to consent to your dousing me in glue and covering me in feathers, it appears that there are some actions I actually *can't* consent to, even when those actions concern my own body – like, for instance, sharing a meal of my sautéed penis before you kill me. Consider, for instance, the 2001 case of Armin Meiwes and Bernd Jürgen Armando Brandes, who together consumed Brandes' penis before Meiwes went on to kill Brandes. Meiwes was originally convicted of manslaughter (rather than murder), since “Brandes had voluntarily and knowingly participated in the act”;¹⁰⁴ that is, the wrongness of Meiwes' actions were deemed to be mitigated by the fact that Brandes had consented (both to the consumption of his penis and his subsequent murder). Notice, however, that even though Meiwes' actions were deemed *less wrong* than they otherwise would have been, they were not deemed *permissible* – he was still convicted of manslaughter. Presumably, this is because, even if the victim consents, it is still not permissible to kill him/her. Even if consent can push back against other reasons, it doesn't appear to be a definitive consideration. It does not appear to be true, therefore, that your moral reason to not kill me is had in virtue of my second-personally commanding so – because, had I, like Brandes, said otherwise, you would *still* have had a reason not to kill me (even if that reason is somewhat weakened by my consent).

Furthermore, note that three years after the manslaughter conviction Meiwes was retried and convicted of *murder*, where “[a]mong the questions courts answered is whether Brandes agreed to his killing, and whether he was legally capable of doing so at the moment, taking into account his apparent mental problems as well as his significant intake of alcohol.”¹⁰⁵ Likewise, with reference to sexual consent again, notice *who* is allowed to decide what kind of sex they participate in: mentally unimpaired adults. Like mentally impaired adults, children also are absolutely not allowed to partake in such activities, *regardless of whether or not they consent*. The guiding presumption here appears to be that, as Brandes' purported consent was (later) deemed to be nullified by his impaired

¹⁰³ Wallace 2007, p.29.

¹⁰⁴ Wikipedia 2014.

¹⁰⁵ Wikipedia 2014.

mental state, children don't yet have the cognitive capacity to make such (potentially life-changing) decisions; that is, children don't yet *know enough* to make an informed decision. And, so, we are back to the idea that the authority to give commands/consent is predicated on *knowledge*. Children and the adult mentally impaired lack *epistemic* authority, and thus they lack the authority to be the final arbiters of what can and cannot be done to their bodies. (Likewise, such individuals also lack the authority to vote in a democratic society.¹⁰⁶) We have already encountered a similar idea in discussion of Divine Command theory,¹⁰⁷ where I argued that both God and I share the authority to hold me/myself accountable because we – unlike you – both possess the requisite knowledge of me/myself and my circumstances. Darwall, however, as we saw, rejects epistemic authority.¹⁰⁸ However, the rejection of the epistemic basis of authority again appears to be out of keeping with our intuitions and common practice with regard to matters of consent, since the authority to give consent appears to depend on epistemic status. Consent, therefore, seems to be constrained by pre-existing non-second-personal reasons (such as the reason to not murder), and the authority to give consent appears to depend on the agent's epistemic status. That is, consent seems to be relevantly dissimilar to the sort of second-personal reason that Darwall thinks constitutes our moral reasons.

As such, what appeared to be the leading intuition in favour of second-personal reasons – matters of consent – is exposed as otherwise. What intuitive support there is for the existence of second-personal reasons is limited. I turn now to consider positive intuitions *against* the existence of second-personal reasons.

Generally, to do something (solely) *because another person told you to* is a poor reason for action.

Lavin, I think, is correct to ask of Darwall

how it can belong to a sound condition of the faculty of practical reason that I, who after all am not the same agent as you, should nevertheless not merely take your reactions into account, but should submit myself to the standard of what *you* could deem to be reasonable.¹⁰⁹

How could it be that / “who after all am not the same agent as you” could have a reason to perform an action because *you* told me to? Children who use the excuse “But Sally told me to!” are corrected by their parents: “If Sally told you to jump into the fire, would you?” The answer, of course, is supposed to be: No, unless I had independent reasons either for jumping into the fire or for doing what Sally tells me to do, I wouldn't have a good reason for jumping into the fire, simply in

¹⁰⁶ See Estlund (2009), who argues convincingly that “democratic authority” (i.e. the right to vote) is grounded not (solely) in a person's dignity, but in the “epistemic value” of his/her decisions.

¹⁰⁷ See s.1.4, above.

¹⁰⁸ See s.1.4, above.

¹⁰⁹ Lavin 2008.

virtue of Sally's having told me to do so. The fact that somebody addresses a command to you doesn't, in and of itself, give you a reason to comply. As Fleischacker was quoted, above, "If you were [...] to issue the command 'Move your foot!' quite arbitrarily, when my foot was nowhere near yours and doing nothing that I could reasonably see as disturbing you, I would" not have a reason to move my foot.¹¹⁰ Like Sally's commanding me to jump into the fire, I wouldn't have a reason to move my foot simply in virtue of your telling me to do so. Rather, it seems that the second person must him-/herself have good reasons for telling the first person to behave as he/she commands.

Now, Darwall does seem to be aware of this. On its own, Sally's telling you to jump into the fire *isn't* a good reason to do so. A second-personal command isn't a magic spell that can be cast over a first person, whether he/she likes it or not. And, this is why Darwall lists *validity* as a condition of second-personal address.¹¹¹ Jumping into the fire isn't (*ceteris paribus*) a valid demand to make of another person, and that is why I do not have a reason to do so simply in virtue of Sally's telling me to.

But, what makes a second-personal command valid? I think that the intuitive status of second-personal reasons hinges on this question. I can intuitively accept that there are such things as second-personal reasons – as long as they are valid. Sally's telling me to jump into the fire is invalid, and thus I do not have a reason to do so – on the other hand, my telling you not to touch me like that *is* valid, and thus you *do* have a reason not to treat me like that. I can intuitively accept, therefore, that a second person can give a first person a reason simply in virtue of the second person's *valid* say-so. So, the pertinent question is: what makes a second-personal command valid? As I will go on to argue, however, Darwall is unable to provide a satisfactory account of validity. If I am correct, then I think whatever intuitions I may have in favour of (apparent) second-personal reasons are not of reasons literally given by a second person to a first, but merely of cases in which one individual informs or reminds an agent of the reasons that that agent already, independently possesses.

2.3. Validity

A second-personal command isn't a magic spell that can be cast over a first person by the second's mere say-so, because the command has to be a *valid* one. Invalid "commands" have no normative force over their addressees. So, for instance, in Darwall's characterization of the second-personal response to the foot-treader, he carefully stipulates that "you might lay a claim or address a purportedly *valid* demand".¹¹² But, as Lavin asks:

¹¹⁰ Fleischacker 2009, p.120, as cited in s.1.4, above.

¹¹¹ Darwall 2006. The claim is present throughout the book, but see in particular pp.11-13.

¹¹² Darwall 2006, p.7 (emphasis added), as cited in my Ch.1, s.1.2.

How can the fact that others expect certain things from me, hold certain attitudes toward me, etc. – how can any of this constitute a valid claim on *my* thinking about what to do, one that I must accept if I am to think rationally?¹¹³

Unfortunately, Lavin only poses the question and does not take it up, as I will do. What are the conditions of validity? I can envisage three possible answers.

2.3.1. Validity as grounding in independent reasons

One natural suggestion is that a command is valid if there is an independent reason in support of it – “something [...] which lies beneath”,¹¹⁴ as Brewer puts it – and invalid if there is not. Thus, Sally’s command that I jump into the fire is invalid because there is no independent reason for me to do such a thing. On the other hand, say, your command that the foot-treader get off your foot is a valid one since there’s an independent reason for it: your pain. To quote Darwall:

That an action would cause severe harm, or even pain to your bunions, is a reason for someone not to do it, whether or not anyone has any standing to demand that he not, and it *supports*, moreover, a relevant demand.¹¹⁵

If, as I think is reasonable to presume, what Darwall means by a “relevant demand” here is: a second-personal reason, then Darwall appears to be claiming that your second-personal command that the foot-treader remove his foot from on top of yours (for instance) is valid when it is “support[ed]” by independent facts – namely, that the “action would cause severe harm”. When it’s not so supported, it’s not valid.

The problem with this approach, however, is that it is no longer *second-personal*, since it grounds the validity of an individual’s authority in (his/her knowledge of) *non-second-personal* reasons, such as the fact of your pain. As discussed, above,¹¹⁶ just that a person tells you that you have a reason to, say, get off his/her foot is not enough to make this a second-personal reason. Second-personal reasons, recall, “give another a distinctive kind of reason he wouldn’t otherwise have had”;¹¹⁷ they don’t just draw the agent’s attention to independent reasons he already has – this is precisely what makes the difference between Darwall’s two possible responses to the foot-treader.¹¹⁸ So, if Darwall attempts to define validity as grounding in independent reasons, then the second person’s role is reduced to one of simply pointing out to the first person what reasons he/she is already subject to, and the resultant reasons are not fully second-personal. As such, Darwall cannot take this approach.

¹¹³ Lavin 2008.

¹¹⁴ Brewer 2008, p.165.

¹¹⁵ Darwall 2006, pp.99-100 (emphasis added).

See also: “Of course, many of the reasons that ground claims of wrong and obligation are not themselves second-personal.” (Darwall 2006, p.99, as cited in s.2, above).

¹¹⁶ See s.2.2, above.

¹¹⁷ Darwall 2006, p.107.

¹¹⁸ Darwall 2006, pp.5-6, as cited in my Ch.1, s.1.2.

I read this as the Darwallian analogue of Euthyphro's dilemma. Euthyphro's dilemma is an objection to Divine Command theory; it asks: "Is ϕ ing good because God commands it, or does God command ϕ ing because it's good?" The problem for Divine Command theory is that, if we take the former horn and accept that ϕ ing is good because God commands it, then it turns out that God's commands are arbitrary; had he commanded, say, that we jump into the fire, then we would indeed have been required to jump into the fire. But, that can't be right.¹¹⁹ On the other hand, if we take the latter horn and accept instead that God commands ϕ ing because it's good, then we have abandoned Divine Command theory. Likewise, if we accept that a second-personal command has normative force just because Sally said so, then Sally's commands are like a magic spell and their normative force over me is arbitrary; on the other hand, if we accept that Sally commanded me to remove my foot from on top of hers because of the fact that that is the best way to relieve her pain, then we have abandoned second-personality.

If this is the case then Darwall's Second-Personal account, like Divine Command theory, falls prey to Euthyphro's dilemma. Of course, given that Darwall's presents his Second-Personal account as the descendant of Divine Command theory,¹²⁰ we shouldn't be surprised to see that it inherits the sins of the father.

2.3.2. Validity as authority

To be fair, although I think it is the most natural one, the above understanding of validity¹²¹ isn't well supported by the rest of *The Second-Person Standpoint*. Instead, Darwall explicitly and repeatedly explains validity as follows:

someone can sensibly accept this [...] reason for moving his foot, one embodied in your claim or demand, only if he also accepts your authority to demand this of him (second-personally). That is just what it is to accept something as a valid claim or demand.¹²²

We asked what it is that gives a second person the authority to lay a second-personal claim against a first person. Darwall rejects the notion that it's the second person's epistemic authority,¹²³ so we followed up on the suggestion that it's the validity of the second person's claim that lends it its authority. However, notice that Darwall defines 'validity', above, *in terms of authority*. A second person has the authority to lay a second-personal claim against a first person when that claim is valid, but a claim is valid when the second person has the authority to lay a second-personal claim

¹¹⁹ We have already encountered this worry in Ch.1, s.6.1.2, as the motivation for Pufendorf's distinction between genuine moral accountability and mere behavioural modification by negative reinforcement.

¹²⁰ See Ch.1, s.6.

¹²¹ See s.2.3.1, above.

¹²² Darwall 2006, p.8. Similar claims are made throughout the book.

¹²³ See ss.1.4 and 2.2, above.

against that first person. Of course, this is just a circular definition. Brewer observes similarly that Darwall's

position seems at certain junctures to boil down to the unhelpful insistence that moral obligations are required because people can legitimately require one to perform them, and that one must act morally because other people can legitimately demand that one do so.¹²⁴

To his credit, Darwall is aware of the circularity:

These notions – second-personal authority, valid claim or demand, second-personal reason, and responsibility to – [...] comprise an interdefinable circle; each implies all the rest.¹²⁵

I will reserve comment as to whether or not we can accept circular definitions like this; it's not obvious to me that all circles are necessarily vicious circles.¹²⁶ There is a larger problem anyway.

The quote continues:

Moreover, I contend, there is no way to break into this circle from outside it. Propositions formulated only with normative and evaluative concepts that are not already implicitly second-personal cannot adequately ground propositions formulated with concepts within the circle.¹²⁷

We have already seen a similar assertion, above,¹²⁸ where Darwall rejects the Neo-Kantian strategy of deriving the Categorical from the Hypothetical Imperative for precisely the same reason as above: there is no way to break into the Second-Personal circle from outside it. The problem, however, is that breaking into the Second-Personal circle from outside it is precisely what Darwall does (tries to do) in his appeal to Pufendorf's Point, as I argued, above.¹²⁹ If Pufendorf's Point is that a genuinely morally accountable agent is one who holds him-/herself accountable, then the agent has the prerogative to determine whether or not he/she is held genuinely morally accountable, which is to say that *the agent determines his/her own reasons*; or, in terms of validity, Pufendorf's Point is that your authority over me is valid when I grant you that authority. As Fleischacker puts it "I [...] grant you second-personal authority over me and oblige myself, in the course of feeling obliged by you, to remove my foot."¹³⁰ Yet, as Wallace notes, that would be to "[change] the subject",¹³¹ from a second-personal account to a *first-personal* one. Pufendorf's Point tries to break into the Second-Personal circle by way of the first-personal. If Darwall's argument for Morality as Equal Accountability depends on Pufendorf's Point, then Darwall is already on the outside of the circle looking in.

¹²⁴ Brewer 2008, p.167.

¹²⁵ Darwall 2006, p.12.

¹²⁶ Euclidean geometry, for instance, appears to be founded on a pair of circularly defined axioms: a point is a position on a line, and a line is the shortest distance between two points.

¹²⁷ Darwall 2006, p.12.

¹²⁸ See Darwall 2006, p.42, fn.12, as cited in s.1.3, above.

¹²⁹ See s.1.3, above.

¹³⁰ Fleischacker 2009, p.120, as cited in s.1.4, above.

¹³¹ Wallace 2007, p.35, as cited in s.1.3, above.

There is, however, an alternative way to read Pufendorf's Point, according to which it is not obviously grounded in the first-personal. I alluded to it, above,¹³² and we are now in a position to pick it back up again. It is the suggestion that the first person holds him-/herself accountable from the *same, shared* point of view from which the second person holds the first person accountable: the *moral point of view*. If the agent holds him-/herself accountable from the moral point of view, then the agent is under some constraint as to what he/she can and can't hold herself accountable for, and so it is no longer clear that what he/she can and can't be held accountable for is a matter of his/her first-personal authority. Likewise, the second person will be constrained by the moral point of view in what he/she can and can't legitimately demand of a first person. That is, valid demands will be those made from the moral point of view, and invalid demands will be those which are not. And, since Darwall construes validity and authority as "interdefinable",¹³³ likewise, a second person will have the authority to make a demand of a first person when it is made from the moral point of view.

Now, this may sound a lot like a suggestion we have already considered, above: the suggestion that a valid command is one grounded in independent reasons.¹³⁴ And, of course, this was a suggestion we went on to *reject*, on the grounds that embracing it would be to abandon second-personality. However, the 'moral point of view' suggestion is relevantly different to the 'independent reasons' suggestion, the difference being that the moral point of view is *not* independent of the first and second persons. Rather, the moral point of view is *made up of* the points of view of first and second persons – and, indeed, of *all* beings with dignity. Thus, grounding validity or the second person's authority to make a demand of a first in the moral point of view does not try to break into the Second-Personal circle from outside. The moral point of view is *already* second-personal.

This 'moral point of view' solution is the one Darwall¹³⁵ offers in his response to *The Second-Person Standpoint* reviewers Korsgaard,¹³⁶ Wallace¹³⁷ and Watson.¹³⁸ However, it seems to me that this solution is already present in that original book – only under the banner of "the [S]econd-[P]erson[al] standpoint"¹³⁹ (although Darwall does in *The Second-Person Standpoint* also employ the

¹³² See s.1.3, above.

¹³³ Darwall 2006, p.12, as cited above.

¹³⁴ See s.2.3.1, above.

¹³⁵ Darwall 2007.

¹³⁶ Korsgaard 2007.

¹³⁷ Wallace 2007.

¹³⁸ Watson 2007.

¹³⁹ Darwall (2006) actually calls it the "second-person standpoint". (See, for instance, the title of the book.) 'Second-Personal standpoint', however, strikes me as a better grammatical construction. This small difference isn't intended to indicate a meaningful deviation from Darwall. (And, recall that I've chosen to capitalize the term in order to distinguish it from other second-personal accounts, such as Divine Command Theory.)

“moral point of view” a handful of times). I examine this suggestion, below, under its original label, from which the book takes its title. I do this for two reasons. Firstly, because this project is primarily an examination of *The Second-Person Standpoint*, it makes sense to stick to its original terminology. Secondly, I think that ‘the Second-Personal standpoint’ is a better label. ‘The moral point of view’ strikes me as distinctly Nagelian terminology, and seems to connote that there is a thing, called morality, to which we refer when we make demands of one another – that is, it seems to suggest the *non*-second-personal, ‘independent reasons’ reading I considered in the preceding paragraph.¹⁴⁰ ‘The Second-Personal standpoint’, on the other, bears no such connotations of non-second-personality. Although, as I will go on to argue, this may belie the reality.

2.3.3. Validity as a grounding in the Second-Personal standpoint

We began this section 2.3 with the observation that doing something purely because a second person told you to do so seems a poor reason. If such a command is to issue a good reason, it has to be *valid*. But, we asked, what does this mean? So far, we have examined the suggestions that a valid command is one supported by independent reasons¹⁴¹ and that a valid command is one made by an authoritative second person,¹⁴² both of which were found to bottom out as non-second-personal. Now, I examine a third suggestion as to what it might mean to say that a Second-Personal command is ‘valid’. It is the suggestion that a Second-Personal command is valid when it is made *from the Second-Personal standpoint*.

To be fair, this has been part of Darwall’s account all along, although I have played it down until this point. So, for instance, Darwall writes that

To intelligibly hold someone responsible, we must assume that she can hold herself responsible in her own reasoning and thought. And to do that, she must be able to *take up a [S]econd-[P]erson[al] standpoint on herself and make and acknowledge demands of herself from that point of view*.¹⁴³

Of course, what Darwall is describing in the quote above is Pufendorf’s Point. Thus, if it’s the case that an agent must hold him-/herself accountable not from *his/her own* point of view but from *the Second-Personal standpoint*, then it follows that it is *not* (entirely¹⁴⁴) the agent’s prerogative what he/she can and can’t be held accountable for – as I argued it was, above¹⁴⁵ – since the agent is subject to some kind of constraint in this regard. That is, it is *not*, after all, at bottom a first-personal

¹⁴⁰ I go on to pursue an idea like this in Ch.5.

¹⁴¹ See s.2.3.1, above.

¹⁴² See s.2.3.2, above.

¹⁴³ Darwall 2006, p.23 (emphasis added).

¹⁴⁴ Assuming that the Second-Personal standpoint includes the agent’s own point of view as well as others’, then it follows that the prerogative isn’t *entirely* the agent’s own – although it is partially, due to the agent’s point of view constituting part of the Second-Personal standpoint. Nevertheless, this is sufficient as some kind of constraint on what the agent can and can’t hold him-/herself accountable for.

¹⁴⁵ See s.1.3, above.

account, as I complained.¹⁴⁶ Rather, a Second-Personal command is valid when it is made by the second person from this Second-Personal standpoint – and this is the *same* standpoint from which the agent holds him-/herself accountable.

So, it all comes together: as the first person is constrained in what he/she must hold him-/herself accountable for, so the second person is constrained in what he/she can demand of the first person; what the agent holds him-/herself accountable for will coincide with what the second person holds the agent accountable for, as both parties do this from the shared Second-Personal standpoint. In Darwall's own words, second-personal competence¹⁴⁷ "must consist in something like the capacity to make demands of oneself from a [S]econd-[P]erson[al] standpoint" – at which point Darwall goes on to consider the objection: "But isn't acting on demands that others can make of one heteronomy rather than autonomy, being governed by them rather than by oneself?" or, as I have put it: isn't acting on demands that others can make of one to have that other *invalidly* impose his/her will on you? To which Darwall replies:

The response to this objection is that when one decides to reject the prince's [or any other second person's] demand because this is what the moral community authoritatively demands, the [S]econd-[P]ersonal perspective of a member of the moral community is as much one's own as it is anyone else's. One demands the conduct of oneself from a point of view one shares [with the second person].¹⁴⁸

So, grounding validity in the Second-Personal standpoint can explain second-personal authority in a way that is consistent with Pufendorf's Point. Furthermore, notice that on this account it *will* follow that the first and second persons share the same authority. That is, the *non-sequitur* charge I laid in section 1 of this chapter will fall away. It *will* follow that first and second persons share mutual authority when their authorities are both grounded *in the Second-Personal standpoint*. When we finally acknowledge the point I've been neglecting for almost the full duration of this chapter – namely, that authority and validity are grounded in the Second-Personal standpoint – Darwall's Second-Personal account is vindicated.

However, this victory is short-lived. The reason that I avoided this element of Darwall's Second-Personal account until now was not to present a skewed picture of the account, but, rather, because it's so unclear what it *means* to "take up a [S]econd-[P]erson[al] standpoint on [one]self."¹⁴⁹

Darwall, disappointingly, is less than forthcoming on this score, leading Wallace, for one, to complain that

¹⁴⁶ See s.2.3.2, above, for the most recent version of this persistent complaint.

¹⁴⁷ See Ch.1, s.6.2.1, for an explanation of second-personal competence.

¹⁴⁸ Darwall 2006, p.35, referring to Kant's (1788/1996, p.30) example of the prince who demands of an agent "on pain of the same immediate execution, that he give false testimony against an honourable man whom the prince would like to destroy under a plausible pretext."

¹⁴⁹ Darwall 2006, p.23, as cited above.

there is a fundamental unclarity about how Darwall understands the standpoint of second-personal address, which his wide-ranging and erudite elucidations do not succeed in dispelling [...].¹⁵⁰

On the face of it, taking up the Second-Personal standpoint looks like it means that one puts oneself in the shoes, so to speak, of the second-person addressor, and holds oneself accountable from the point of view of *that second person*.¹⁵¹ This reading seems to be supported by Darwall's discussion of "The Psychology of the Second Person"¹⁵² and his numerous appeals to a kind of Simulation Theory; for example: "Second-personality thus requires empathy or the capacity to *put oneself in another's shoes*",¹⁵³ and: "Second-personal interaction requires empathy in the sense of simulation or imaginative projection into *the other's point of view*",¹⁵⁴ where the point of view one simulates is clearly a *particular* second person's. But, this doesn't help the argument at all. What, we must ask, makes it the case that that particular second person's point of view is a *valid* one? This is precisely the question we are trying to settle. We are presently considering the suggestion that a Second-Personal command is a valid one when it is made from the Second-Personal standpoint – and if the Second-Personal standpoint is simply whatever the second person commands, then this can't be accepted as an account of *validity* at all.

In a more recent paper, however, Darwall makes clear that

In the case of moral demands, the requisite standing is not any *individual authority* anyone has to exercise at his or her discretion. It is rather a *representative authority* we all, all second-personally competent persons, have to hold one another and ourselves to moral demands.¹⁵⁵

That is, it is not that case – as I worried, above – that taking up the Second-Personal standpoint is a matter of simulating the point of view of a particular, "*individual*" second person, whose point of view may or may not be valid; rather, it is a matter of simulating the "*representative*" point of view.

As Darwall goes on to clarify, with reference to second-personal competence:

What, then, does second-personal competence involve? [...] I take it to follow [...] that [...] it includes the capacity for reactive attitudes, in addition to imaginative and empathetic abilities necessary to put oneself in others' shoes, *including those of a representative person*.¹⁵⁶

¹⁵⁰ Wallace 2007, p.31.

¹⁵¹ Darwall's term "the second-person standpoint" (see fn.139, above) leads even more directly to this (mis)understanding of the term. The "second-person standpoint" seems straightforwardly to mean: the standpoint of the second person. The term I have chosen to use instead, the 'Second-Personal standpoint', hopefully leaves room for the possibility of there being a new kind of standpoint that comes into being when (at least) two persons, the first and the second person, interact *Second-Personally* – which is not reducible to either the first or the second person's standpoints.

¹⁵² Darwall 2006, Ch.7.

¹⁵³ Darwall 2006, p.44 (emphasis added).

¹⁵⁴ Darwall 2006, p.45 (emphasis added).

¹⁵⁵ Darwall 2010c, p.31.

¹⁵⁶ Darwall 2010c, p.39-40 (emphasis added).

The shoes into which we second-personally put ourselves, therefore, are “*those of a representative person*”, not those of a particular second person. The trouble with this suggestion, however, is that, unless we know what the shoes of the ‘representative person’ look like, so to speak, simulating this point of view will not enable us to make judgements of validity.¹⁵⁷ Who is the representative person? Indeed, it seems that we encounter the same problem with this suggestion as we did with that of the preceding section:¹⁵⁸ presumably, we will only know what the representative point of view is when we know what claims can validly be made of persons – but, according to the present suggestion, we can only know what claims can validly be made of persons when we put ourselves into the shoes of the ‘representative person’.

Similarly, in the very first sentence of *The Second-Person Standpoint*, Darwall provides this apparent definition:

Call the [S]econd-[P]erson[al] standpoint the perspective you and I take up when we make and acknowledge claims on one another’s conduct and will.¹⁵⁹

But, again, this unfortunately isn’t helpful, since our present interest is in taking up the Second-Personal standpoint as the condition under which “we make and acknowledge claims on one another’s conduct and will” *validly*. We need to know more about the Second-Personal standpoint than this opening line reveals – we need to know what is it about this standpoint that makes it the case that demands made from it are valid demands. I hope by this stage to have done enough to show why we *have to* have an (acceptable) account of the Second-Personal standpoint, and why leaving it unanalysed won’t work. Fichte’s Point – that any second-personal reason commits us to mutual, reciprocal second-personal authority – only supports Darwall’s Second-Personal account *if* there are such things as second-personal reasons.¹⁶⁰ There does seem to be some limited intuitive support for the existence of second-personal reasons – but only if they are *valid*.¹⁶¹ A second-personal command isn’t a magic spell; you can’t obligate me to do just anything simply by telling me that I have to do it. The problem, however, is to cash out validity in a way that doesn’t try to break into the Second-Personal circle from outside it; that is, in a way that doesn’t boil down to the non-second-personal.¹⁶² The current suggestion is that a valid second-personal command is one made from the Second-Personal standpoint – but, unless Darwall tells us what that ‘Second-Personal

¹⁵⁷ Furthermore, it’s unclear to me that the point of view of a representative person could be *simulated* at all. Since the representative person is not an actual person, it strikes me that the process of working out what would be the representative person’s point of view would have to be a theory-driven, rather than a simulation-driven, one (see, for instance, Davies & Stone 1998), and thus would not be an exercise of “imaginative and empathetic abilities” (Darwall 2010c, p.39-40, as cited above) at all.

¹⁵⁸ See s.2.3.2, above.

¹⁵⁹ Darwall 2006, p.3.

¹⁶⁰ See s.2.1, above.

¹⁶¹ See s.2.2, above.

¹⁶² See s.2.3.2, above.

standpoint' actually *is*, we are no closer to being persuaded that there actually are such things as second-personal reasons, and thus no closer to accepting Darwall's Second-Personal account. I consider one last suggestion, below.

2.3.3.1. *The Second-Personal standpoint as contract*

One *prima facie* plausible suggestion as to what the Second-Personal standpoint might be, then, is the standpoint of the *contract* between (amongst) second persons. This suggestion will strike those readers familiar with Darwall's earlier work on Contractualism as natural. As Darwall himself observes in *The Second-Person Standpoint*, "the arguments I have been advancing in this book lead naturally to a [C]ontractualist approach".¹⁶³ That is, the Second-Personal standpoint appears to lend itself to cashing out in Contractualist terms. On this reading, the Second-Personal standpoint is the standpoint of a contracting party, from which that contracting party agrees that another party is (not) allowed to expect certain actions from him/her – provided that he/she may (not) in turn expect those same actions from that other party. The resulting obligations will be second-personal in the sense that they are a new kind of reason that didn't exist prior to the contract, and which now come into being in the relationship between contracting parties; they are "a distinctive kind of reason [the agents] wouldn't otherwise have had"¹⁶⁴ but for entering into contract. They are second-personal in the sense that they depend as much on the second person as they do on the first; if there were not two parties, there could be no *contract* between them.¹⁶⁵

This Contractualist reading also opens up Pufendorf's Point (that, when a second person holds a first person accountable, that first person must also hold him-/herself accountable):¹⁶⁶ when you, as a second person, hold me, a first person, accountable from the Second-Personal standpoint, you effectively say to me, "Look here – this is what *you yourself* would agree to!" That is, you hold me accountable to that to which I *already* hold myself accountable, as a contracting party. Pufendorf's Point is already built into such a Contractualist account. And, second-personal commands – that is, commands made from the Second-Personal standpoint – will thus be *valid* too: when you lay a second-personal claim against me from the Second-Personal standpoint, you don't attempt to cast a magic spell over me, according to which I am obligated simply in virtue of your having said so; rather, you hold me to that to which I would myself consent, and therefore to that which you have the authority to validly demand. Understanding the Second-Personal standpoint in this way – that

¹⁶³ Darwall 2006, p.300.

¹⁶⁴ Darwall 2006, p.107.

¹⁶⁵ Furthermore, this Contractualist approach to the Second-Personal standpoint makes good sense of Darwall's (2006, p.9) otherwise curious remark that "the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint". I raise this in Ch.3, s.1.3.

¹⁶⁶ See Ch.1, s.6.1.2, as well as s.1.3, above.

is, as the standpoint from which we would all consent to a contract stipulating what we may and may not expect from one another – accounts for second-personal reasons’ validity in a way that is consistent with Pufendorf’s Point, and for second persons’ mutual authority such that Darwall’s argument for Morality as Equal Accountability is not a *non-sequitur*, as I complained, above.¹⁶⁷

Notice that this Contractualist account of the Second-Personal standpoint promises to provide that sought-after account of *both* the normative force and the universality of moral reasons.¹⁶⁸ Moral reasons will have normative force over us because they will be person-regarding: they are not states of the world, out there, but are constructed on the basis of features of ourselves, namely, the contract into which we would enter with second persons. And, because of the nature of that feature of ourselves to which they attach – namely, the contract into which we would enter *with second persons*, rather than, say, our individual and idiosyncratic desires – it will also follow that moral reasons are universal. As I put it in the previous chapter, the other is *built into* the Second-Personal account, so universality follows straightforwardly.¹⁶⁹ The Second-Personal account does not need to mount a Generality¹⁷⁰ or Analogical¹⁷¹ argument in order to reach its universal conclusion, unlike the Neo-Kantian account. Finally, the Contractualist account of the Second-Personal standpoint allows us to see why this is so.

2.3.3.2. *Problems with the Second-Personal standpoint as contract*

Given all the apparent advantages of understanding the Second-Personal standpoint in Contractualist terms, it may be surprising that Darwall does not explicitly make this case. Although he claims that “the arguments I have been advancing in this book lead naturally to a [C]ontractualist approach”,¹⁷² he falls short of making the claim the arguments of *The Second-Person Standpoint* commit us to Contractualism. And, although Darwall does devote a chapter of the book to Contractualism,¹⁷³ at no point in this chapter does he cite Contractualism as furnishing his Second-Personal account with the requisite theory of validity, as is our concern in this present section.

However, Darwall’s reluctance to appeal to Contractualism as providing the requisite account of validity is well-founded when one examines the real nature of Contractualist obligation. There are broadly two versions of Contractualism: one – more properly called *Contractarianism* – which has its

¹⁶⁷ See s.1, above.

¹⁶⁸ See Ch.1.

¹⁶⁹ See Ch.2, s.6.2.2.

¹⁷⁰ See s.1.1, above.

¹⁷¹ See s.1.2, above.

¹⁷² Darwall 2006, p.300, as cited in s.2.3.3.1, above.

¹⁷³ Darwall 2006, Ch.12, pp.300-320.

roots in Hobbes, and another which has its roots in Kant. Both versions of Contractualism, I will argue, are unacceptable as accounts of the Second-Personal standpoint.

2.3.3.2.1. Contractarianism

Contractarianism is that brand of Social Contract theory rooted in the Hobbesian tradition, according to which “moral principles of right derive from a mutually advantageous contract”.¹⁷⁴ On this account, I bind myself by the terms of the contract to, say, refrain from stepping on your foot, on condition that *you* in turn will refrain from stepping on *my* foot – and the same goes for you: since you want to avoid having me step on your foot, you agree to not step on mine as long as I don’t step on yours. If, therefore, you go ahead and step on my foot, violating the terms of our contract, I no longer have a reason to refrain from stepping on your foot (and *vice versa*):

And this happens, when I devolve into a *State of War* with another. For the *Obligation* being *reciprocal*, in case one party break it, and return Hostilities for the good Offices of Humanity, the other party is no longer bound to deal with him in friendly manner; but may maintain his own safety and his own Rights, although with the hurt of the unjust Opposer.¹⁷⁵

Since you want to avoid suffering “the hurt of the unjust Opposer”, you have a reason to refrain from stepping on my foot. Now, it may appear that this is a reason I have *given* you to refrain from stepping on my foot – that is a *second-personal* reason – since it is a reason that only comes into being when we enter into contract. However, this is not the case.

Of course, there can only be such a thing as a contract when there are (at least) two contracting parties, two persons, but it doesn’t follow on this basis alone that the subsequent obligations are thereby *second-personal* in the relevant sense of the term. On this Contractarian account, the sense in which I ‘give’ you a reason not to step on my foot by entering into contract with you is only a very thin one, not relevant to the Second-Personal account. It is not a command issued from my second-personal authority; you don’t have this reason in virtue of my dignity. The only sense in which I ‘give’ you a reason not to step on my foot, on this Contractarian account, is *so that I don’t step on yours*. But, this is not a second-personal reason. It is, rather, a self-interested reason – a *first-personal* reason. You enter into contract with me in order to protect *yourself*, not out of respect for *me* but because respecting (‘respecting’) me is the means to getting me to respect (‘respect’) *you*. And this, recall, was just Korsgaard’s complaint against the Analogical argument:

¹⁷⁴ Darwall 2006, p.304, fn.8.

As I noted in Ch.1, fn.53, this reading of Hobbes is not entirely uncontroversial; however, I do take it to be the natural, predominant reading thereof. Moreover, Darwall clearly also reads Hobbes as belonging to the Contractarian tradition (2006, p.308, fn.8, as cited above). (See also Darwall 1989.) Readers who prefer an alternative reading of Hobbes are directed to s.2.3.3.2.2, below.

¹⁷⁵ Pufendorf 1672/1729, p.258.

The quote could easily have been from Hobbes. However, it is actually from Pufendorf – another reason for questioning whether Darwall’s account is really compatible with Pufendorf’s (see s.1.3, above), as will become clear, below.

It would show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that I have duties with respect to you, about you, but not that there are things I owe *to you*.¹⁷⁶

As I remarked, above,¹⁷⁷ it appears that the Second-Personal account really has not made much headway over the first-personal Neo-Kantian account. If we understand the Second-Personal standpoint in these Contractarian terms then it appears that the resultant reasons are not second-personal at all, but rather first-personal. (And, if Hobbesian Egoism is a first-personal metaethic,¹⁷⁸ it is hardly surprising that such Contractarian obligations – rooted as Contractarianism is in the Hobbesian tradition – will be first-personal.)

Finally, notice that this Contractarian reading of the Second-Personal standpoint is actually in conflict with Pufendorf's Point. Above, I claimed that the suggestion under consideration "opens up" Pufendorf's Point,¹⁷⁹ since, if you hold me accountable from the Second-Personal standpoint – that is, from the standpoint of the contract between us – and if I hold myself accountable from that same standpoint, then it will follow that you and I share the same authority, namely, the authority of the Second-Personal standpoint/contract. However, notice *how* you hold me accountable to this contract: if I don't comply with what we have agreed to and I violate the contract, then *you* no longer have any reason to curb your behaviour in the same way. If I don't refrain from, say, stepping on your foot, then I can't be assured that you will not step on *mine*. That is, I refrain from stepping on your foot because of the *threat* of your stepping on my foot. Your stomping foot is the "object of terror" that "shakes [my] will" into compliance.¹⁸⁰ But, this is not genuine moral accountability.¹⁸¹ Because Contractarian reasons come about as a reaction to the threat of what you might do to me (if I don't refrain from doing the same thing to you), they are not compatible with Pufendorf's Point about genuine moral accountability.

Since, on this Contractarian reading of the Second-Personal standpoint, the resultant reasons are first-personal, as well as incompatible with the Pufendorfin view of genuine moral accountability, this Contractarian reading of the Second-Personal standpoint cannot ground valid second-personal reasons. So, let us try, instead, a Contractualist reading.

2.3.3.2.2. Contractualism

According to Elizabeth Ashford and Tim Mulgan, Contractarianism differs from Contractualism primarily in that: "By contrast, any form of [C]ontractualism is grounded on the equal moral status of

¹⁷⁶ Korsgaard 1996, p.134, as cited in Ch.1, s.5.3, and in s.1.2, above.

¹⁷⁷ See s.1.1, above.

¹⁷⁸ See Ch.1, s.3.

¹⁷⁹ See s.2.3.3.1, above.

¹⁸⁰ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111, as cited in my Ch.1, s.6.

¹⁸¹ See Ch.1, s.6.

persons.”¹⁸² The contract involved in this version of Contractualism, therefore, is the result not of self-interested bargaining but of regarding the other as an equal moral agent to whom one is required to justify oneself. As Darwall puts it:

the root [C]ontractualist idea is that [...] [w]e have an equal (second-personal) authority to make claims of one another, which we respect in seeing each other as beings to whom we should be able to justify ourselves.¹⁸³

The version of Contractualism Darwall has primarily in mind is that advanced by Scanlon.¹⁸⁴ On this Scanlonian Contractualist account, a moral agent will have justified him-/herself to those sharing his/her equal moral authority when his/her action passes the “‘no-reasonable-rejection’ test”.¹⁸⁵

Having at its heart, as such a Scanlonian Contractualism does, the equal moral authority of all contracting parties, note that this version of Contractualism will avoid the problem faced by Contractarianism, above.¹⁸⁶ Whereas, on the Contractarian account, I abstain from treating you in certain ways in order self-interestedly to ensure that you too abstain from treating me in those ways, on this Scanlonian Contractualist account, I abstain from treating you in certain ways out of respect for your dignity, a dignity which I share with you, the second person. That is, unlike Contractarianism, a Scanlonian Contractualist account will yield genuinely Second-Personal reasons.¹⁸⁷ It would not, as Korsgaard has required, “show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that [...] there are things I owe *to you*.”¹⁸⁸

We have, in fact, encountered this version of Contractualism in this project already. As we noted in Chapter 1, one advantage of Darwall’s Second-Personal account is that its derivation of Morality as

¹⁸² Ashford & Mulgan 2012, s.2 (emphasis removed).

¹⁸³ Darwall 2006a, pp.204-205.

¹⁸⁴ Scanlon 1998, as referenced throughout Darwall 2006, Ch.12, pp.300-320. See also Darwall 2006a.

¹⁸⁵ Darwall 2006, p.301, referencing Scanlon 1998.

¹⁸⁶ See s.2.3.3.2.1, above.

¹⁸⁷ This is actually a simplification of Darwall’s position. In a paper published just before *The Second-Person Standpoint*, Darwall (2006a, pp.210-211) argues that

Scanlon’s claim is that the wrongness of breaking promises is to be explained by the wrongness of violating *Principle F*. Unless, however, we interpret the idea of “providing assurance” in a ([S]econd-[P]ersonal) way [...], I doubt that this is so.

That is, Darwall argues that Scanlon’s account needs to be “*interpret[ed]*” (emphasis added) Second-Personally – which implies that there is a *non*-Second-Personal reading of Scanlon available. If that is the case, then it’s not a straightforward matter that Scanlonian Contractualist reasons will be Second-Personal reasons, as I have suggested, above.

Darwall does, however, gloss over this complication in his discussion of Scanlonian Contractualism and Second-Personality in *The Second-Person Standpoint* (2006, Ch.12, pp.300-320), so I trust that my presentation above is acceptable. Besides, my argument of this section will effectively be that it is irrelevant whether or not Scanlonian Contractualist reasons are Second-Personal, since Darwall is not permitted to appeal to Scanlonian Contractualism in defence of Morality as Equal Accountability anyway.

¹⁸⁸ Korsgaard 1996, p.134, as cited in Ch.1, s.5.3, and in ss.1.2 and 2.3.3.2.1, above.

Equal Accountability vindicates the Contractualism that Darwall is well-known for having defended elsewhere.¹⁸⁹ Scanlonian Contractualism presupposes that all persons share the equal moral authority to disallow a principle of action by reasonably rejecting it,¹⁹⁰ or, as Darwall was quoted above, “the root [C]ontractualist idea is that [...] [w]e have an equal (second-personal) authority [...]”¹⁹¹ And, Morality as Equal Accountability is supposed to demonstrate that we do indeed all share this equal moral authority. (This argument for Morality as Equal Accountability, recall, was that a second person’s holding a first genuinely morally accountable presupposes that that first also holds him-/herself morally accountable, and that the first and the second person therefore share equal moral authority.¹⁹²)

There are, however, two problems here. The first is that I have already rejected Darwall’s argument for Morality as Equal Accountability as a *non-sequitur*. A second person’s holding a first genuinely morally accountable does not demonstrate that that first and second person share equal (mutual) moral authority; it only demonstrates that the first and second person share an authority *over the first person*, and not that the first has reciprocal authority over the second.¹⁹³ If I am correct that Darwall’s argument for Morality as Equal Accountability does not go through, then Contractualism is not vindicated; we are left without good reason for thinking that we do, as Scanlonian Contractualism presupposes, share equal morality authority. Why should I have to justify my actions to the likes of *you*?

Secondly, this reminder of our previous encounter with Scanlonian Contractualism brings to light a further problem with the present consideration of it. We must bear in mind that our present consideration of Scanlonian Contractualism is *as a candidate account of validity*. To recap the argument of this chapter, Darwall’s argument for Morality as Equal Accountability argues from the fact of a second person’s giving another a reason, to mutual and reciprocal second-personal authority.¹⁹⁴ But, in order for this argument to get off the ground, we must first establish that there are indeed such things as second-personal reasons,¹⁹⁵ and that was the task of this present section 2. There seems to be some intuitive support for the existence of second-personal reasons – but only when the commands a second person might make of an agent are *valid* commands.¹⁹⁶ The suggestion presently under consideration is that a second-personal command is valid when it is

¹⁸⁹ See Ch.1, s.6.2.2.

¹⁹⁰ Scanlon 1998, p.153, as cited in Ch.1, s.6.2.2.

¹⁹¹ Darwall 2006a, pp.204-205.

¹⁹² See Ch.1, s.6.2.

¹⁹³ See s.1, above.

¹⁹⁴ See s.1, above.

¹⁹⁵ See s.2.1, above.

¹⁹⁶ See s.2.2, above.

made from the Second-Personal standpoint,¹⁹⁷ where this is understood in the Scanlonian Contractualist terms described above. However, as we have noted several times now, Scanlonian Contractualism presupposes that we share equal moral authority,¹⁹⁸ or, as Darwall captures it, that “we are mutually accountable for regulating ourselves by principles that are acceptable [...] to each as *equal moral persons*”.¹⁹⁹ That is, Scanlonian Contractualism *presupposes* what Morality as Equal Accountability sets out to prove, namely, that we share equal moral authority to hold each other accountable. If, therefore, Morality as Equal Accountability is supposed to vindicate a Scanlon-style Contractualism, Contractualism of course cannot be employed to support Darwall’s argument for Morality as Equal Accountability by providing an account of validity. Such an argument would be circular.

As a point of clarification, note that the preceding argument is not the objection, as it may be liable to be misread, that Darwall’s Second-Personal account is vacuous or that it falls short of providing the substantive ethical injunctions that should guide our behaviour.²⁰⁰ That is not the task of a *metaethical*, which is to provide an account of the nature of moral reasons (rather than the content of those moral reasons). Reading my complaint as one of vacuity may be understandable, however, on the basis of claims by Darwall such as the following:

the most promising way to work out the content of moral principles we are committed to by the equal dignity we must presuppose in the second-person perspective is through some version of [C]ontractualism.²⁰¹

But, note that my argument has *not* been that Contractualism is unable to supply us with “the content of moral principles”.²⁰² Rather, what I have questioned is Darwall’s claim to “the equal

¹⁹⁷ See s.2.3.3, above.

¹⁹⁸ See Ch.1, s.6.2.

¹⁹⁹ Darwall 2006, p.304 (emphasis added).

²⁰⁰ This is, however, a task that Darwall takes up in a subsequent paper, ‘Moral obligation: form and substance’, (2010c) in which he explores whether any “substance” might be derived from moral obligations’ second-personal “form”. Note, however, that this paper clearly “[b]egin[s] from an analysis of moral obligation’s *form* that [Darwall] defend[ed] in *The Second-Person Standpoint*” (p.31), as Darwall notes in that paper’s opening line. That is to say, the second-personal “substance” that Darwall aims to derive in that paper can be of no help to our present call for an account of validity, since the former takes for granted the Second-Personal account – yet we are currently seeking an account of validity in order to vindicate precisely that Second-Personal account.

²⁰¹ Darwall 2006, p.304.

Note that by the time of publication of ‘Moral obligation: form and substance’ (2010c; see fn.200, above), Darwall appears no longer to believe that “the most promising way to work out the content of moral principles [...] is through some version of [C]ontractualism” (2006, p.304, as cited above). Instead, in that later paper, Darwall aims to derive the “substance” of moral principles through a Hobbesian version of “the golden rule” (Darwall 2010c, pp.41-42, referencing Hobbes 1651/1994, Ch.15, para.35).

²⁰² For what it’s worth, I *do* in fact have doubts that Contractualism will be able to furnish us with such moral principles. If “An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement” (Scanlon 1998, p.153), we will of course first need an account of *reasonable*

dignity we must presuppose in the second-person perspective”. This is a claim to which we cannot help ourselves in the absence of Morality as Equal Accountability – and we cannot assume Morality as Equal Accountability until we have an account of validity to vindicate our second-personal intuitions. Since I have questioned whether we are “committed to [...] the equal dignity we must presuppose in the second-person perspective”, it does not follow that “the most promising way to work out the content of moral principles” “is through some version of [C]ontractualism”, and hence I have *not* argued that Contractualism is unable to supply us with “the content of moral principles”. To put the point another way, I have here not mounted any objection *to Contractualism* at all. Instead, I have objected to Darwall’s possible *employment of Contractualism* as providing us with an account of validity. And, I have objected to such a possible employment of Contractualism not because of any fault with Contractualism *per se*, but because Darwall’s appealing to it would render his account circular. Darwall’s Second-Personal account, if it is able to establish that we share equal moral authority,²⁰³ vindicates Contractualism.²⁰⁴ Contractualism, therefore, cannot be employed as an account of validity, in order to vindicate the intuition that there are second-personal reasons.²⁰⁵ Such an argument would be circular.

To be fair, I must note that Darwall does not himself advance such a viciously circular argument. He does not explicitly suggest Contractualism as an account of the validity of second-personal commands. I have, however, explored the possibility of Contractualism’s grounding Second-Personality in this section – not to falsely accuse Darwall of a circularity he doesn’t commit – but because it occurs to me that readers familiar with Darwall’s earlier work will likely be inclined to read his Second-Personal account in this manner. However, I think this is to read too much into *The Second-Person Standpoint*. It is not, as I read it, a book about Contractualism. It does include a chapter on Contractualism,²⁰⁶ but that chapter is the very last chapter of the book, and thus not indispensable to the important ideas set out in its first 11 chapters. More to the point, the subject of that chapter is *Second-Personality’s grounding Contractualism* (as is very clearly indicated by its title, ‘A foundation for Contractualism’), not Contractualism’s grounding Second-Personality (the possibility of which has been the topic of this present section 2.3.3.2.2). So, although Darwall is clear that “In this final chapter, [he] briefly develop[s] the idea that the [S]econd-[P]erson[al] standpoint

rejection, and one might be sceptical that such an account is available. (Presumably, this explains Darwall’s (2010c) abandonment of Contractualism. See fn.201, above.) However, my present point is that I do *not* here take up this objection to Contractualism. In fact, I do not here take up any objection *to Contractualism* at all; instead, I object to Darwall’s possible *employment of Contractualism* as providing an account of validity.

²⁰³ Which, of course, I have denied in s.1, above.

²⁰⁴ See Ch.1, s.6.2.2.

²⁰⁵ See s.2.3, above.

²⁰⁶ Darwall 2006, Ch.12, pp.300-320.

can ground a [C]ontractualist normative moral theory”,²⁰⁷ I find very little evidence for the converse claim that Contractualism might ground Darwall’s Second-Personal account, by supplying the requisite account of validity. Indeed, there are positive reasons for thinking that Darwall would *not* make the claim that Contractualism could provide the requisite account of validity, since, as I have argued, such a claim would constitute a circular argument, by presupposing to the equal moral authority that Morality as Equal Accountability is supposed to prove. Darwall claims only that Second-Personality can ground Contractualism, not that Contractualism grounds Second-Personality. I expand on this in the course of rebutting a possible response to my argument, below.

One possible response worth considering is that the role of Contractualism in Darwall’s Second-Personal account points to another circle,²⁰⁸ in the same manner in which we saw the relationship between validity and authority to be circular, above²⁰⁹ – that is, non-viciously circular. According to this possible response, although I have shown that Darwall takes Contractualism to presuppose Second-Personality, might it not *also* be the case that Darwall takes Second-Personality to reciprocally entail Contractualism? Positing such a Second-Personality/Contractualism circle would, admittedly, be in keeping with the trend of *The Second-Person Standpoint*, and at least one Darwallian does indeed read the Second-Personal account in this way; consider the following remark by Robin Bradley Kar: “we must [...] proceed from [C]ontractualist standards that we are implicitly committed to when we take up the [S]econd-[P]erson[al] standpoint and hold people responsible [...]”²¹⁰ That is, Kar appears to be of the opinion that Second-Personality “commit[s]” us to Contractualism (over and above Contractualism’s presupposing Second-Personality).²¹¹ If this is the case, then my pointing out, above, that Darwall takes Second-Personality to vindicate Contractualism does not block the possibility that Darwall might also, circularly but non-viciously so, take Contractualism to “vindicate” his Second-Personal account by supplying the requisite account of validity.

However, I reject this suggestion of a Second-Personality/Contractualism circle. While there is plenty of evidence that Darwall thinks that Contractualism presupposes Second-Personality,²¹² there is little evidence to suggest that he thinks the circle is completed by Second-Personality’s also

²⁰⁷ Darwall 2006, pp.300-301.

²⁰⁸ Thanks to an anonymous reviewer for this suggestion.

²⁰⁹ See s.2.3.2, above.

²¹⁰ Kar 2007, p.884; see also pp.885 and 890.

²¹¹ Kar (2007) does, however, make this remark in discussion of legal and political authority, rather than *moral* authority, so it is possible that Kar does not hold that Second-Personal *moral* authority commits us to Contractualism.

²¹² This is especially clear in Darwall’s (2006a) review of Scanlon’s work.

entailing Contractualism. Rather than claiming that Second-Personality “commit[s]”²¹³ us to Contractualism, Darwall writes quite clearly: “*although they do not require it*, the arguments I have been advancing in this book [i.e. Second-Personality] lead naturally to a [C]ontractualist approach”²¹⁴ – that is, although Contractualism may, so to speak, be a good fit for the Second-Personal approach, the former is not entailed by the latter. We should compare Darwall’s modest claim here of Second-Personality’s “lead[ing] naturally to a [C]ontractualist approach” with the much stronger claims he makes with reference to the relationship in the other direction (*i.e.* from Contractualism to Second-Personality), according to which Contractualism has as its “foundation”²¹⁵ or “root”,²¹⁶ and is “grounded in”,²¹⁷ Second-Personality. Darwall makes no claim that Contractualism is, reciprocally, the “foundation”, “root” or “ground” of Second-Personality. Furthermore, although *The Second-Person Standpoint* does indeed leave the reader with the clear sense that Darwall remains committed to the Contractualism he has defended elsewhere, the book includes a short section explaining how Second-Personality might be compatible with even *Rule-Consequentialism*, à la Parfit.²¹⁸ Holding, as he does, Consequentialism to be “[C]ontractualism’s foil”,²¹⁹ I take Darwall’s admission of Rule-Consequentialism’s compatibility with Second-Personality to be a decisive consideration in support of my view that, although Darwall holds Contractualism to presuppose Second-Personality he does not also hold Second-Personality to reciprocally entail Contractualism, and does not posit a circular dependency between the two. Second-Personality is, according to Darwall, the “root”²²⁰ or the “foundation”²²¹ of Contractualism, and not *vice versa*.

I conclude, therefore, that Scanlonian Contractualism cannot serve as an account of validity, since it would presuppose the equal moral authority that Morality as Equal accountability is supposed to prove. Nor can Hobbesian Contractarianism supply the requisite account of validity, since such reasons would be first-personal rather than second-personal.²²² As such, we are left without an account of what the Second-Personal standpoint actually *is*. We examined in this section 2.3 two other suggestions as to the validity of second-personal commands – firstly, the suggestion that valid commands are ones supported by independent reasons,²²³ and secondly, the suggestion that a valid

²¹³ Kar 2007, p.884, as cited above.

²¹⁴ Darwall 2006, p.300 (emphasis added), as partially cited in ss.2.3.3.1 and 2.3.3.2, above.

²¹⁵ Darwall 2006, Ch.12, pp.300-320.

²¹⁶ Darwall 2006a.

²¹⁷ Darwall 2006a, p.214.

²¹⁸ Darwall 2006, pp.310-313, referencing Parfit 2004.

²¹⁹ Darwall 2006a, p.203.

²²⁰ Darwall 2006a, as cited above.

²²¹ Darwall 2006, Ch.12, pp.300-320, as cited above.

²²² See s. 2.3.3.2.1, above.

²²³ See s.2.3.1, above.

command is one made with second-personal authority²²⁴ – but both were rejected on the same grounds on which the two contract readings were rejected: they either bottom out as non-second-personal or wind up as circular. So, in conclusion of this section 2.3, we are left without any useful account of the validity of second-personal reasons.

3. Conclusion

In section 1 of this chapter, I set out to show that Darwall’s argument for Morality as Equal accountability is a *non-sequitur*. Even if we accept Pufendorf’s Point, it does not follow on the basis of your giving me a second-personal reason that we thereby have mutual and reciprocal authority; all that follows is that you and I share an authority over *me* (only).²²⁵ I also questioned whether Pufendorf’s Point is compatible with a Second-Personal account at all,²²⁶ whereas I argued that it is compatible with the Divine Command theory.²²⁷

In section 2 I went on to question whether Darwall’s argument for Morality as Equal Accountability can even get off the ground. That argument takes for granted that we give each other second-personal reasons, and tries to demonstrate, on that basis, that we all have the authority to hold each other accountable – but why should we think that there are such things as second-personal reasons in the first place? This is hardly an uncontroversial foundation on which to ground a transcendental deduction for Morality as Equal Accountability.²²⁸ There does appear to be some intuitive support for the existence of second-personal reasons, but that support is limited by the accompanying intuition that you can’t obligate me to do just anything simply in virtue of commanding me to do so; your command only has normative force over me if it’s valid.²²⁹ However, we have been unable to construct a useful account of validity that is consistent with Darwall’s claims of Second-Personality.²³⁰ To this end, we examined the suggestions that validity is grounding in independent reasons,²³¹ that validity is authority,²³² and that validity is grounding in the Second-Personal standpoint, understood as the standpoint of a contract.²³³ However, all these candidates for validity were objectionable, either on the grounds that they abandoned Second-Personality, or on the grounds that they were ultimately circular. I am at a loss as to what else might constitute the

²²⁴ See s.2.3.2, above.

²²⁵ See ss.1.1 and 1.2, above.

²²⁶ See s.1.3, above.

²²⁷ See s.1.4, above.

²²⁸ See s.2.1, above.

²²⁹ See s.2.2, above.

²³⁰ See s.2.3, above.

²³¹ See s.2.3.1, above.

²³² See s.2.3.2, above.

²³³ See s.2.3.3, above.

validity of a second-personal command. As such, I conclude that whatever intuitive appearances there may be of second-personal reasons, they are explained away as cases of the second person merely drawing the first person's attention to independently existing reasons (either world-regarding or first-personal), rather than as genuinely second-personal cases of the second person literally *giving* the first person reasons. We ought to be sceptical of the existence of second-personal reasons at all. As such, Darwall's argument for Morality as Equal Accountability doesn't even get off the ground.

However, there is another possible argument for the existence of second-personal reasons, arguing from our practice of holding agents accountable to the nature of those reasons to which we hold agents accountable. If this new argument works, then we will have a non-question-begging basis for Darwall's argument for Morality as Equal Accountability. I turn to this Mill's-Point argument in the following chapter.

3.

Mill's Point

In the previous chapter, I claimed that Darwall's argument for Morality as Equal Accountability is a *non-sequitur*, demonstrating only that a first and second person share authority over the first, and not that the first and second person share equal, reciprocal authority over each other, as Darwall wishes to claim.¹ I then went on to argue that, not only is Darwall's argument for Morality as Equal Accountability a *non-sequitur*, the argument itself doesn't even get started: the argument has the form of a transcendental deduction, which takes as its starting point the presumption that there exist second-personal reasons – but it is far from clear that such second-personal reasons exist at all. While there does appear to be some intuitive support for the existence of second-personal reasons, this support is limited by the concomitant intuition that a second person can only give a first person a reason when that reason is *valid*, yet Darwall does not provide an acceptable account of validity.² As such, not only is Darwall's argument for Morality as Equal Accountability a *non-sequitur*, it doesn't even get off the ground because we have no good reason to believe that second-personal reasons exist in the first place.

The argument of the preceding chapter sought intuitive support for the existence of second-personal reasons, and found that support to be inconclusive. However, there is another argument available for the existence of second-personal reasons as Darwall conceives of them, one that does not rely on our direct intuitions about their existence. *The Second-Person Standpoint* does not foreground this argument in the same way that it does the argument for Morality as Equal Accountability,³ but it is there nevertheless. The argument is this:

Mill famously remarked that “we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.” We may think there are strong moral reasons for people to do something and “dislike or despise them for not doing” it; but unless we think “blame” or some other form of “punishment” is warranted, perhaps just the feeling of guilt, we do not think it “a case of moral obligation” [...]. Many philosophers today broadly concur with these Millian sentiments [...].⁴ Matters of moral

¹ See my Ch.2, s.1.

² See my Ch.2, s.2.

³ See my Ch.2.

⁴ Here Darwall references: Adams 1999, p.238; Baier 1966; Brandt 1979, pp.163–176; Gibbard 1990, p.42; and Skorupski 1999, p.142.

obligation are the moral community's "business," as Kurt Baier puts it [...]. When we are morally obligated, we are not morally free to act otherwise; members of the moral community have the authority to hold us responsible if we do. Moral obligations are thus *to* others in a more robust way than those of logic are.⁵

In this passage, Darwall's objective is to establish that the distinctive feature of moral obligation, or "what is felt to be special about moral requirements",⁶ is: not that moral obligations are categorical, since norms of etiquette are also categorical;⁷ nor that they have overriding force, since logical requirements also have overriding force;⁸ but that when we are morally obligated we are liable to being *held accountable* for failing to act as we are obligated.⁹ As Darwall goes on to say, there is a "conceptual connection between moral obligation and responsibility".¹⁰ Or, as he puts it later on:

⁵ Darwall 2006, p.27, citing Mill 1863/1998, Ch.5, and Baier 1966, p.223.

⁶ Darwall 2006, p.26.

⁷ Darwall 2006, p.26, referencing Foot 1972.

⁸ Darwall 2006, pp.26-27.

⁹ In *The Second-Person Standpoint*, it's not quite clear whether Darwall thinks that it's accountability alone that is the distinctive feature of moral obligation, or whether it's the combination of categoricity, overriding force and accountability that is distinctive. His presentation of the argument inclines me towards the latter interpretation: he seems to be arguing not that moral obligations *aren't* categorical or *aren't* overriding, but that these two features are insufficient to account for the distinctiveness of moral obligation, and thus, that a third feature is required: accountability. A later article supports this reading, where Darwall asserts (although unfortunately without explanation): "While both features [i.e. categoricity and overriding force] seem necessary, we can fairly easily see that both are, even jointly, insufficient." (2010c, p.35). However, this isn't quite so "easily" seen by me. Moral requirements seem already to be distinctive in that they are *both* categorical and overriding, and neither requirements of etiquette nor of logic are both. Although requirements of etiquette are, as Foot (1972) points out, categorical, they are clearly not overriding: queue-barging is rude, but if you can save somebody's life by jumping a queue, that's exactly what you ought to do. And, although logical requirements are, as Darwall (2006, pp.26-27) argues, overriding, it's not clear to me that they are categorical. Moral requirements are categorical in the sense that you ought to, say, refrain from causing undue harm regardless of your desires on the matter (or on causally related matters); I ought to refrain from causing undue harm, *full-stop*. Logical reasons are, of course, theoretical reasons rather than practical reasons, so the analogue of categoricity in the logical realm would apparently be: you ought to believe *p* regardless of your beliefs on the matter (or on conceptually (?) related matters). But this definition of logical categoricity is incoherent. Logic of course doesn't demand that I believe *p*, *full-stop* – it demands that I believe *p* if, for instance, I also believe *If a then p* as well as *a*. That is to say, the logical requirements to which one is subject are a matter of the rest of one's beliefs – in other words, precisely the opposite of categoricity (or its theoretical analogue). So, even though logical requirements may be overriding, they don't appear to me to be categorical.

As such, moral requirements appear already to be distinctive in their being both categorical (like requirements of etiquette, but unlike requirements of logic) and overriding (like requirements of logic, but unlike requirements of etiquette). If this is the case, then it's unclear to me why Darwall thinks we need some third special feature of moral reasons (i.e. accountability) in order to distinguish them.

Nevertheless, it strikes me as entirely plausible that moral reasons indeed *are* reasons in accordance with which we are accountable for acting. As such, I will simply overlook what strikes me as the problematic contextualization of the argument, and will focus my attention instead on Darwall's "Mill's Point": the claim that moral obligation is conceptually connected to moral accountability (whether it is this feature that alone distinguishes moral requirements from other requirements or not).

¹⁰ Darwall 2006, pp.27-28.

“In a slogan: The moral sense of ‘responsible for’ is conceptually tied to ‘responsible to’ (whether to individuals or to one another as members of the moral community).”¹¹ Call this ‘Mill’s Point’.¹²

On the basis of Mill’s Point, Darwall argues, it follows that, “[l]ike moral responsibility, moral obligation is an irreducibly second-personal concept.”¹³ That is, because others “have the authority to hold us responsible [...] [m]oral obligations are thus *to others*”,¹⁴ as I quoted Darwall, above. If Mill’s Point is correct, if the fact that we are held accountable *by* others entails that our moral obligations are *to others*, then our moral obligations are second-personal. That is, if Mill’s Point is correct, then Darwall will have demonstrated on the basis of the nature of moral accountability that there are indeed such things as second-personal reasons. Call this Darwall’s Mill’s-Point argument.

I read the Mill’s-Point argument as an inference to the best explanation; that is, the fact that our moral obligations are *to others* (i.e. second-personal) is the best explanation of the fact that we are rightly held accountable *by* others. In Darwall’s words: “Part of the argument for the claim [...] is the claim’s role in an overall picture of second-personal address and reasons that I will hope will seem compelling and able to explain significant ethical phenomena”,¹⁵ where (some of) those “significant ethical phenomena” are the instances of our being held accountable by others. If Darwall is able to infer the existence of second-personal reasons on the basis of our practice of holding others accountable, then the conclusion of my preceding chapter must be (partially¹⁶) revised.

Before I proceed to an examination of Darwall’s Mill’s Point argument, let me first make a quick note on reasons and obligations. Readers will have noticed that discussion has gradually shifted from talk of ‘reasons’ to that of ‘obligations’. Darwall does not draw a sharp division between the two in *The Second-Person Standpoint*, and seems to use the two terms interchangeably.¹⁷ I take it that

¹¹ Darwall 2006, p.68.

¹² Note that ‘Mill’s Point’ is not Darwall’s (2006) own term. Unlike all his other “Points” (Pufendorf’s, Fichte’s, Strawson’s, etc.), this ‘point’ does not get its own name. As I remarked, above, the Millian argument that will be the subject of this chapter is not foregrounded in Darwall’s (2006) book in the same way that others are. I do, however, nevertheless think that it plays an important role. For ease of reference, therefore, and following the suit of all the rest of Darwall’s “Points”, I dub it ‘Mill’s Point’, since Darwall makes the point mainly with reference to Mill, as quoted, above.

¹³ Darwall 2006, p.91.

¹⁴ Darwall 2006, p.27, as cited above.

¹⁵ Darwall 2006, p.76, fn.26.

Darwall makes the inference-to-the-best-explanation form of his argument even more explicit in a more recent paper, writing: “Because accountability to someone entails the latter’s *authority* to demand and *hold one* to conduct (‘second-personally’), I claim that accountability necessarily involves a *second-personal* relation and authority.” (Darwall 2010c, p.32.)

¹⁶ The conclusion of s.2 of my Ch.2 (namely, that we have no good reason to believe that second-personal reasons exist at all) will have to be revised – however, the conclusion of s.1 of my Ch.2 is unaffected: Darwall’s argument for Morality as Equal Accountability remains a *non-sequitur*.

¹⁷ In a more recent paper, however, Darwall (2010c, p.31) draws the following distinction:

Darwall's discussion of Mill's Point is framed in terms of 'obligations' rather than in terms of 'reasons', since the point of that discussion is to establish that our reasons/obligations are a relational matter (i.e. a second-personal matter), and are had *to* others. And, the language of 'obligation' makes this point more succinctly and naturally than does that of 'reasons': where 'obligation' can cover both the non-second-personal case of one's having an obligation (*full-stop*) or the second-personal case of one's having an obligation-to-one's-children (for instance), the vocabulary of 'reasons' is less versatile in that we wouldn't normally speak of one's having a 'reason-to-one's-children'. Following Darwall, therefore, the discussion of this chapter will be framed largely in the terms of 'obligation', where this is supposed to bear the same implications for the nature of moral reasons, and is not intended to signal any shift from the preceding discussion of 'reasons'.¹⁸ If, therefore, Mill's Point is that there is a conceptual connection between obligation and accountability, and if accountability is a second-personal matter, then it will follow that obligation is also a second-personal matter – that is, that there exist such things as second-personal *reasons*.

The objective of this chapter, as well as that of the following two, will be to dispute the Mill's-Point argument that the fact of our liability to being held accountable¹⁹ by others entails that we were obligated *to* those others, second-personally. Note that I do *not* dispute that element of Mill's Point

By 'moral obligation' I mean something narrower than what moral reasons recommend, even conclusively, namely, what morality *requires* or *demands*—in other words, what it would be morally *wrong* not to do.

This draws a distinction between moral obligation and the *content* of moral reasons or "what moral reasons recommend" (although it does not elucidate the relationship between moral obligations and moral reasons *per se*, but only the content of the latter). It's unclear to me, however, that this distinction is consistent with Darwall's claim in *The Second-Person Standpoint* that

morality exists only through the possibility of a second-personal address on which one rational agent can give another a distinctive kind of reason he wouldn't otherwise have had [...]. (2006, p.107, as cited in my Ch.2, s.2)

This claim, I have remarked previously (see Ch.2, s.2), suggests that all moral reasons are second-personal, and that any non-second-personal reasons must therefore be non-moral. As such, there apparently can be no such thing as "what *moral* reasons recommend" (emphasis added) in the absence of those reasons being "*require[d] or demand[ed]*" by a second person, since "morality exists only through the possibility of a second-personal address". Therefore Darwall's distinction, above, between moral obligation and the (non-required or non-demanded) content of moral reasons is spurious.

Furthermore, since we have accepted that the correct account of moral *reasons* must be one according to which they have normative force (see Ch.1, s.1.3) – that is, one according to which "it would be morally *wrong* not to do" what one has moral reason to do, it follows that all "moral reasons" will just be moral obligations, according to Darwall's (2010c, p.31) definition, above.

¹⁸ In Ch.7, fn.174, however, I go on to suggest another possible motivation for Darwall's employment of *obligation*.

¹⁹ I use the complicated phrase, "liability to being held accountable", for good reason. The argument is not that because we *are* actually, as a matter of fact, held accountable by individuals, we are therefore obligated to those individuals. The conclusion still follows even if we aren't actually held accountable; all it requires is that we may justly have been held accountable, that we were *liable* to being held accountable, even if (for instance), as a matter of fact, we got lucky and managed to escape the punishment. For the same reason, I will also speak of individuals' *authority* to hold agents accountable.

which is that there is a “conceptual connection between moral obligation and responsibility”;²⁰ this much strikes me as correct. What I do dispute, however, is that we can on the basis of Mill’s Point establish that moral reasons are second-personal in nature. The objections of this present chapter and the next will concern internal problems with the argument, and that of Chapter 5 will demonstrate that the Mill’s-Point argument as an inference to the best explanation fails, since Realism can accommodate Mill’s Point just as well as – if not better than – Darwall’s own Second-Personal account. The argument is split over three chapters only because it is long; the contents of this chapter and the next two should be read as one continuous series of objections to Darwall’s Mill’s-Point argument. As far as I am aware, this is the first thorough examination Darwall’s Mill’s-Point argument has received in the literature.

Let us turn, first, to the internal problems with the Mill’s-Point argument. Not only does Darwall (following Mill) allow that an agent may be held accountable by the *victim* of his wrongdoing, but he also makes what I will call the ‘broader points’ that an agent may hold himself accountable by the “reproaches of his own conscience”²¹ and that an agent may be held accountable by members of the general moral community, who are not (direct) victims of that agent’s misdeeds. The basic objection of this chapter will be that, if an agent may be held accountable by individuals *not* the victims of his wrongdoing, then it is unclear why it should follow that the agent’s obligation was *to a second person*. Lavin captures this line of objection concisely:

In the early chapters, Darwall talks us into the idea of the [S]econd-[P]erson[al] standpoint [...] by meditating on interactions between two distinct individual agents, [...] for example, [...] I, who am standing on your foot, and you, who are telling me to get off. Yet, as the book develops, the scope for second-personal relations expands, leaving room for, indeed, requiring room for, second-person[al] relations between me and myself, and ultimately, between me and the moral community. But does the idea of a solitary second-personal demand make sense? Does the idea of "the moral community" as a maker of demands and an occupier of the [S]econd-[P]erson[al] standpoint [make sense]?²²

That is, if the idea of a second-personal reason is plausible at all, it seems to be most plausible in cases in which a first person has a reason to treat a second person a certain way, because that second person has commanded it, and is therefore liable to being held accountable by that second person in particular.

This is a fitting description of the kinds of cases that Darwall initially uses to, as Lavin puts it, “[talk] us into the idea of the [S]econd-[P]ersonal standpoint”;²³ for instance, the foot-treader purportedly has a second-personal reason to remove his foot from on top of yours because you have told him to

²⁰ Darwall 2006, pp.27-28, as cited above.

²¹ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited above.

²² Lavin 2008.

²³ Lavin 2008, as cited above.

do so, and subsequently hold him accountable for his action.²⁴ This foot-treader scenario involves “two distinct individual agents”,²⁵ one first person (the foot-treader) and one second person (you), who gives the first a reason and holds the first accountable for acting in compliance with it. Likewise, for Darwall’s other main working example of a second-personal reason in *The Second-Person Standpoint*, according to which a private – one individual – is given a reason to do ten push-ups by the command of a sergeant²⁶ – a second individual.²⁷ Darwall’s allowance (following Mill) of the broader points that an agent might hold him-/herself accountable, or even that a member of the general moral community might do so, deviates from this original kind of case involving only a first and a second person. This casts doubt over the claim that the obligation to which the moral agent him-/herself or the member of the general moral community holds the agent accountable is a *second-personal* obligation, as Darwall claims on the basis of Mill’s Point. Thus, Lavin is prompted to ask, as above, “[D]oes the idea of a solitary second-personal demand make sense? Does the idea of “the moral community” as a maker of demands and an occupier of the [S]econd-[P]erson[al] standpoint [make sense]?”²⁸ Whereas Lavin merely raises the questions, these are questions I aim to answer over the course of the next few chapters. As will become clear, many of the objections to Darwall’s Mill’s-Point argument raised in this and the following two chapters will play into the same themes of objection already raised in the preceding chapter.

²⁴ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

²⁵ Lavin 2008, as cited above.

²⁶ Note, for the record, that this second main example of a second-personal reason is surely not even an example of a *moral* reason. A private does nothing *morally* wrong if he refuses to do ten push-ups. Indeed, it is an all too familiar truth that a private’s following his commanding officer’s commands may indeed be the morally *wrong* thing to do.

²⁷ Darwall 2006, p.259-273; see also p.60.

In fairness to Darwall, I must note that some of his other references to the sergeant are ones according to which she commands not just one individual (the private, as above) but a whole *platoon* (pp.12-13 and 123). That is, some of Darwall’s other references to the sergeant do not place her in a one-on-one relationship with another single agent, but rather with a whole group of agents. Note, however, that this does not undermine my present point. My present point is that, if a whole group of individuals (the moral community) may hold an agent accountable, then it does not follow that that agent was second-personally obligated to his/her victim; that is, it does not follow that the victim gave the agent a reason not to treat him/her as the agent went on to do. This is not analogous to the sergeant/platoon example, which is one according to which one individual (the sergeant) purportedly gives a group of agents (the platoon) a reason, and each member of that group of agents is subsequently purportedly accountable to the sergeant for compliance. That is, the sergeant/platoon example is one in which a group of agents are accountable to one individual; the kind of case I am presently concerned with is one in which a single agent is purportedly accountable to a group of individuals. I examine in depth the idea of accountability to a whole group of individuals (i.e. the moral community) is s.2, below, as well as in Ch.4.

²⁸ Lavin 2008, as cited above.

1. The “reproaches of his own conscience” objection

Mill’s Point is that there is a “conceptual connection between moral obligation and responsibility.”²⁹ That is, (part of³⁰) what it means to say that I am morally obligated to, say, keep my promise to Sally is that, if I were to break my promise to Sally, Sally would have the authority to hold me morally accountable for my bad behaviour. I will for the duration of this chapter and the next employ this case of promising as the working example. As I have noted, above, Darwall himself provides the reader of *The Second-Person Standpoint* with just two working examples (one of which isn’t even plausibly a case of a second-personal *moral* reason³¹). As a matter of charity, I have opted not to employ Darwall’s foot-treader example in this present chapter. As I have already noted, it is open to reasonable doubt that in telling the foot-treader to get off your foot you are *giving* the foot-treader a reason to get off your foot; rather, it is quite likely that all you are doing is drawing the foot-treader’s attention to a perfectly ordinary reason that was there all along: that we ought not to cause others undue harm.³² Promising, on the other hand, seems – at least *prima facie* – to be less susceptible to the alternative, Realist interpretation. If second-personal reasons have the form of address,³³ and if making a promise is a form of address,³³ then it is at least plausible that making a promise is a case of giving a second-personal reason. Moreover, promises don’t exist until they are made, and, if second-personal reasons are reasons given by the second to the first, then this makes an even better case for interpreting promising as the giving of a second-personal reason. As such, this will be my working example of choice. According to Darwall’s Mill’s-Point argument, my being held accountable by Sally for keeping the promise I made to her reveals that my obligation was *to* Sally – that is, that it was second-personal.

However, Darwall, following Mill,³⁴ allows that an agent’s being held accountable may take various forms – including that of the agent holding him-/herself accountable “by the reproaches of his own conscience”.³⁵ My objection in this section will be that this kind of accountability doesn’t support second-personality, and therefore that Mill’s Point does not entail Darwall’s desired conclusion. As Darwall, quoting Mill, was quoted at the beginning of this essay,

we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.³⁶

²⁹ Darwall 2006, pp.27-28, as cited above.

³⁰ See fn.9, above.

³¹ See fn.26, above.

³² See Ch.2, s.2.

³³ Darwall 2006, p.3. See also my Ch.1, s.1.2.

³⁴ Darwall 2006, p.27, referencing Mill 1863/1998, Ch.5, as cited in the introduction to this chapter.

³⁵ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited in the introduction to this chapter.

³⁶ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited in the introduction to this chapter.

The difficulty here is fairly obvious. By way of endorsing this quote, Darwall allows, as Mill does, that the kind of moral accountability to which a morally obligated agent is subject may include “the reproaches of his [i.e. the agent’s] own conscience” – “perhaps just the feeling of guilt”,³⁷ as Darwall goes on to add. If Darwall wishes to use Mill’s Point to establish his Second-Personal account, then allowing for “the reproaches of [the agent’s] own conscience” as a form of moral accountability won’t get him very far. Darwall’s conclusion that we are obligated “to others” depends on the premise that we are accountable to *others*. If, however, we are (sometimes) accountable only to *ourselves* – as Mill declares, and Darwall apparently endorses – then it doesn’t follow that we are obligated to others; it follows only that we are obligated to *ourselves*. Being obligated and accountable to oneself (not to others) is not a *second*-personal matter but a *first*-personal one. Thus, if we are liable to being held accountable only by ourselves, then the existence of second-personal reasons does not follow on the basis of Mill’s Point. This is my “reproaches of his own conscience” objection.

This objection is, of course, reminiscent of my objection in the previous chapter to Darwall’s employment of Pufendorf’s Point. Pufendorf’s Point, recall, is that if an individual holds an agent genuinely morally accountable (as opposed to merely modifying that agent’s behaviour by negative reinforcement), then that agent must hold him-/herself accountable³⁸ – which, I argued, is incompatible with Darwall’s claims to second-personality.³⁹ But, notice that this is not quite the same as my present “reproaches of his own conscience” objection. The present objection is that Mill’s Point seems to allow that sometimes an agent is liable to being held accountable by *nobody other than him-/herself*. Consider, for instance, Mill’s claim that “we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.”⁴⁰ This expression seems to suggest that ‘first prize’ cases of accountability are legal, ‘second prize’ cases consist in social disapproval – and, if all else fails, we can still establish the conceptual connection between obligation and accountability (i.e. Mill’s Point), because an obligated agent will hold himself accountable “by the reproaches of his own conscience”. Thus, if Mill’s Point allows – relatedly to, but distinctly from Pufendorf’s Point – that sometimes an agent is liable to being held accountable by nobody other than him-/herself, then I do not see how Darwall can on

³⁷ Darwall 2006, p.27, as cited in the introduction to this chapter.

³⁸ See Ch.1, s.6.1.2.

³⁹ See Ch.2, s.1.3.

⁴⁰ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited in the introduction to this chapter.

this basis infer second-personality. Like Pufendorf's Point, it seems to support, rather, *first-* personality.⁴¹

1.1. The by-and-large objection

One may wish to object to my argument on the grounds that we hold ourselves accountable, first-personally, only *some* of the time; by and large, this work is indeed done by others, second-personally, and thus I exaggerate the import of Mill's comment that moral accountability can consist in the "reproaches of [the agent's] own conscience".

In response to this possible objection, I quote Darwall:

It is perhaps obvious that reasons that depend on orders or requests are second-personal in this sense, but I argue that moral obligations and demands are quite generally second-personal also.⁴²

As the quote makes clear, Darwall does *not* wish to argue for the modest claim that *some* moral obligations (those that depend on orders or requests from others, or those to which we actually are held accountable by others) are second-personal in nature – nor even for the less modest claim that *most* moral obligations are second-personal in nature – but for the bold claim that *all* of them are.⁴³ Recall the context of the opening quote of this chapter: Darwall argues that the distinctive feature of moral reasons is not their categoricity, not their overriding nature, but their second-personality (given their conceptual ties to accountability).⁴⁴ Darwall aims for the conclusion that second-personality is the essence of moral obligation; therefore, if it turns out that there are some moral obligations that are not second-personal in nature – those moral obligations for which we are accountable only to ourselves – then this is a problem for Darwall's argument. Darwall cannot be defended on the grounds that most of the time we are held accountable by others and not only by ourselves. If we are morally accountable to ourselves even only *some* of the time, this is a problem for Darwall's strong Second-Personal conclusion.

And, intuitively, it *does* seem to be the case that we are accountable only to ourselves, at least some of the time. I can think of two such kinds of case, which I mention very briefly. The first kind of case involves the so-called 'duties to oneself', such as, for instance, the duty to provide a minimal level of

⁴¹ Some of the following objections that I go on to consider to my "reproaches of his own conscience" objection (to Darwall's Mill's-Point argument), may (with suitable adjustment) also be raised against my objection to Darwall's employment of Pufendorf's Point (in his argument for Morality as Equal Accountability) (see my Ch.2, s.1.3). In this regard, s.1.3, below, seems especially relevant. However, as I respond to these (possible) objections in the context of discussion Mill's Point, so too would I respond to them in the context of discussion of Pufendorf's Point.

⁴² Darwall 2006, p.8.

⁴³ See fn.17, above.

⁴⁴ Darwall 2006, p.27, citing Mill 1863/1998, Ch.5, and Baier 1966, p.223, as cited in the introduction of this present chapter.

care for one's own body. Imagine I develop a secret habit of self-mutilation, or eat so much junk food that I put myself at serious risk of heart disease. Childless, friendless, and contributing nothing to society, I harm nobody but myself. Nevertheless, it seems that I do myself a moral wrong by treating my body as I do – but it is a moral wrong for which I am accountable only to myself. The McDonald's employee surely has no authority to tell me off for eating what I do. And, though it may appear that my doctor – knowing what he does about the grave risk my treatment of my body presents – has the authority to hold me accountable for treating my body as I do, Darwall would of course object that such authority is merely *epistemic*: the doctor possesses only the authority to *counsel* me to treat my body better; he possesses no authority to *command* me to do so.⁴⁵ But, if I am the only one with the authority to hold myself accountable for what I do, how is this *second-personal*?

The second kind of case in which it seems that only I have the authority to hold myself accountable is one in which nobody besides me knows that I have violated my duty. If I am obligated not to ϕ , but I go ahead and ϕ anyway – unbeknownst to anybody else – then I am surely the only individual with the authority to hold myself accountable. How could you, or anybody else, have the authority to hold me accountable for ϕ ing if you don't even know that I ϕ ed?⁴⁶ And, notice that this appears to hold no matter what ' ϕ ' stands for – even if ϕ stands for treating *you* in a certain way.⁴⁷ Again, if I am liable to being held accountable only by myself in such cases, accountability does not appear to be a second-personal matter. Rather, if Darwall allows that we are sometimes liable to being held accountable by the “reproaches of our own conscience” only, then accountability appears to be a *first-personal* matter.

1.2. Restricting the range of moral obligations

One further attempt to defend Darwall may be made, but I will consider it only very briefly. This defence would be to accept that, when one is accountable only to oneself, one has a first-personal (non-second-personal) obligation – while denying that such first-personal obligations are *moral* obligations. That is, one might respond on Darwall's behalf by restricting the range of genuinely moral obligations. Properly moral obligations are had, according to this defence, only when the agent is obligated to another, that is, second-personally. This would resolve the tension in Darwall's appeal to Mill's Point. So, for instance, if I make a promise to Sally and then – unbeknownst to

⁴⁵ See Ch.2, ss.1.4 and 2.2, for discussions of epistemic authority.

⁴⁶ See Ch.2, ss.1.4 and 2.2, for my suggestion that Darwall can't dismiss the epistemic component of authority. I go on to develop an account of accountability along these lines in Ch.5, s.1.2.1.

⁴⁷ Here, ' ϕ ' can stand for anything, provided that I could have failed to ϕ without your having realized it. Presumably this won't hold for the obligation not to step on your foot (unless you suffer some kind of nerve disorder), but it could hold for the obligation not to scratch your car, or not to talk about you behind your back, etc.

anybody (including Sally) besides myself – go on to break that promise, my subsequently feeling guilty about (i.e. holding myself accountable for) breaking my promise isn't evidence of some *moral* failure but only of some *personal* failure: I feel guilty because I let myself down by violating my own, personal code of conduct to be a woman of my word. I feel guilty in just the same way when I let myself down by not making it to gym (again), yet this isn't a *moral* failure. In these case, I hold myself accountable to a first-personal, *non-moral* reason.

However, I dismiss this response. Firstly, I find no textual support for it in Darwall's work. I don't think Darwall advocates a revisionary approach to what we normally consider to be moral obligations; rather, he aims to provide an account of the distinctive feature(s) of what we normally consider to be moral obligations – and this surely includes keeping promises, even when nobody but me will know that I have broken it. Secondly, of course it's a *moral* failure that I break my promise to Sally – whether Sally knows it or not. Were Sally to learn that I had broken my promise to her, she would surely have the authority to hold me accountable for it. My feeling guilty over breaking my promise is a response to the same wrong – the same *moral* wrong – as Sally's holding me accountable would be, were she to learn that I had broken my promise. Breaking promises, like not making it to gym (again), may also be a personal failure – but breaking promises is surely a moral failure too.

Furthermore, there is something very strange about the implication that the moral status of an action can change depending on how many people know about it. If, as I have argued, Sally lacks the authority to hold me accountable for breaking my promise if she lacks the knowledge that I have indeed broken my promise, then I am accountable only to myself, first-personally, so – on the suggestion of the present response – my failure was not a *moral* one. But, if Sally were to learn that I had broken my promise, and were subsequently to hold me accountable for it, second-personally, then I would indeed be accountable to another, so – on the suggestion of the present response – my breaking my promise *would* now constitute a moral failure. But, this can't be right. I surely behave immorally by breaking my promise to Sally whether she knows it or not, and whether she therefore has the authority to hold me accountable for it or not.

Thus, I don't think that Darwall does or would advocate this revisionary response. And, so, the allowance that an agent may be liable to being held accountable only by the "reproaches of his own conscience" remains a problem for Darwall's argument that accountability – and, therefore, obligation – is a second-personal matter.

1.3. Second-personality as a version of first-personality

I have argued that, if Darwall allows that we may be liable to being held accountable by ourselves only, then it does not follow that accountability is a second-personal matter. Indeed, it seems instead to support the claim that accountability is a *first-personal* matter. However, in fairness to Darwall, it must be pointed out that he does state that “the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint”.⁴⁸ And, if “the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint”, then first- and second-personality are not mutually exclusive, as my objection, above,⁴⁹ presumed. Thus, my pointing out that Mill’s Point entails that some moral obligations are *first-personal* is no strike against Darwall, since, on this suggestion, first-personality needn’t entail *non-second-personality*.

Before I deal with this possible response to my “reproaches of his own conscience” objection, we should note that there is something to be said for the suggestion that the Second-Personal standpoint “is a version of” the first-personal standpoint. Firstly, if the Second-Personal standpoint is a version of the first-personal standpoint, then the purported incompatibility of Darwall’s Second-Personal account with Pufendorf’s Point will dissipate.⁵⁰ Secondly, if the Second-Personal standpoint is “a version of” the first-personal standpoint, Contractarianism will be able to provide us with a neat, non-circular account of validity: I have a first-personal reason to take the means to prevent your stepping on my feet, so I enter into contract with you, which validates my second-personal (where this is a version of the first-personal) reason to refrain from stepping on your feet so that you, in turn, will refrain from stepping on mine.⁵¹ And, this Contractarian account of the Second-Personal standpoint will resolve at least one of the difficulties of the preceding chapter: by yielding an account of validity,⁵² it will vindicate the intuition that there are indeed such things as second-personal reasons,⁵³ which will deliver the starting point of Fichte’s transcendental argument,⁵⁴ thereby entailing Morality as Equal Accountability.⁵⁵

However, a reason had in virtue of my first-personal reason to take the means to prevent your stepping on my feet would *not* constitute a *second-personal* reason. It may *look* a lot like a second-personal reason, in that it would dictate how I am (not) allowed to treat you, but it would not be a

⁴⁸ Darwall 2006, p.9.

⁴⁹ See s.1.1, above.

⁵⁰ See Ch.2, s.1.3.

⁵¹ See Ch.2, s.2.3.3.2.1.

⁵² See Ch.2, s.2.3.

⁵³ See Ch.2, s.2.2.

⁵⁴ See Ch.2, s.2.1.

⁵⁵ Of course, I have also argued that, even if we are able to establish that there are indeed such things as second-personal reasons (i.e. vindicate the starting point of Fichte’s transcendental argument), Darwall’s argument for Morality as Equal Accountability is still a *non sequitur* (Ch.2, s.1).

second-personal reason – at least not insofar as I, following Darwall, have defined second-personal reasons⁵⁶ – in that it would not be a reason *addressed* or *given* to me *by you*. Such a Contractarian reason to refrain from stepping on your feet would be “an obligation *to myself*” (to prevent you from stepping on my feet), that is, an obligation “with respect to you, about you, but not [...] owe[d] to you”,⁵⁷ but owed to *myself*. It would not be one of the second-personal “[m]oral obligations [...] to others”⁵⁸ that Darwall tries to demonstrate on the basis of Mill’s Point. Although reading “the [S]econd-[P]erson[al] stance [as] a version of the first-person[al] standpoint” appears *prima facie* to make good sense of Darwall’s account, it is simply inconsistent with the rest of his claims about second-personality. This is clear even on the basis of Darwall’s own description of the first-/second-personal relationship; as Darwall writes:

What the [S]econd-[P]erson[al] stance excludes is [...] first-personal thought that lacks an addressing, second-personal aspect.

Thus, although second-person[al] address is always also first-personal, it is never merely first-personal. One can occupy a first-person[al] perspective [...] without explicitly addressing anyone.⁵⁹

What’s special about Darwall’s metaethic is his positioning at the very centre of the account the second-personal relationship, which in the quotation directly above (as elsewhere) he models on the relationship of *address*: one person speaking to a second, who is listening to the first.⁶⁰ (Where this should be contrasted with speaking *about* a second person⁶¹ – which is the way the Contractarian reading of the Second-Personal standpoint (above) misconstrues second-personal reasons.) As Darwall writes just a page before: “second personal reasons always derive from agents’ relations to one another”,⁶² where this relation is one of *address*. So, since “first-personal thought [...] lacks an addressing, second-personal aspect” generally, and since “[o]ne can occupy a first-person[al] perspective [...] without explicitly addressing anyone” – as Darwall himself makes very clear (above) – first-personality therefore lacks the essence of Darwall’s account of moral obligation: the bipolar relationship between addressor and addressee, obligator and obligated. So, it is a curious thing why Darwall would want to claim that “the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint” at all, as it would appear to undermine his own account of second-personality.

I believe that this curious suggestion can be ascribed to confusion surrounding the traditional ‘agent-relative’/‘agent-neutral’ distinction. To quote Darwall’s curious remark in full:

⁵⁶ See Ch.1, s.1.2.

⁵⁷ Korsgaard 1996, p.134, as cited in my Ch.1, s.5.3, and Ch.2, s.1.2.

⁵⁸ Darwall 2006, p.27, as cited in the introduction to this chapter.

⁵⁹ Darwall 2006, pp.9-10.

⁶⁰ See my Ch.1, s.1.2.

⁶¹ Dullstein 2012, p.239.

⁶² Darwall 2006, p.8.

*Since second-personal reasons are always fundamentally agent-relative, the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint [...].*⁶³

That is, Darwall appears to believe that second-personal reasons are ‘versions of’ first-personal reasons because second-personal reasons are agent-relative, and (implicitly) because first-personal reasons are agent-relative. I have a number of responses to this.

Firstly, first- and second-personal reasons’ both being agent-relative strikes me as a poor reason to conclude that second-personal reasons are versions of first-personal reasons – why not conclude, instead, that first-personal reasons are versions of second-personal reasons? Indeed, this is precisely Korsgaard’s reaction to Darwall; as she writes: “[B]ecause of the reflective structure of human consciousness, I think that every rational agent stands in what Darwall would call a second-personal relation to herself – she has a second-personal voice within.”⁶⁴ I do not here wish to explore the Korsgaardian account of second-personal reasons, but raise the Korsgaardian case only to make the point that the shared agent-relativity of first- and second-personal reasons gives us no better grounds to conclude, as Darwall does, that second-personal reasons are versions of first-personal reasons than it does to conclude, as Korsgaard does, that first-personal reasons are versions of second-personal reasons.⁶⁵

Secondly, I question Darwall’s concluding that “the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint” on the basis of first- and second-personal reasons’ shared agent-relativity⁶⁶ because Darwall appears here to be trading with a somewhat protean concept of agent-relativity. In a footnote to his first mention of agent-relativity in *The Second-Person Standpoint*, Darwall references the Nagelian definition of agent-relativity: agent-relative reasons are ones

⁶³ Darwall 2006, p.9 (emphasis added).

⁶⁴ Korsgaard 2007, p.11.

⁶⁵ Indeed, a cursory analysis suggests that there are better grounds to favour Korsgaard’s (2007) position rather than Darwall’s (2006). If either kind of reason is ‘a version of’ the other, then Korsgaard’s position, which takes first-personal reasons to be versions of second-personal ones, is *prima facie* more sensible. If second-personal reasons are, as Darwall puts it, reasons that “derive from agents’ relations to one another” (2006, p.8) – where these are *any* agents – then first-personal reasons could be those special kinds of second-personal reasons where the agent with whom one stands in relation happens to be *oneself*. As Korsgaard (1996, p.104) puts it: “(T)he reflective structure of human consciousness establishes a relation here, a relation which we have to ourselves”, a relation between the “thinking self” and the “doing self”. Although this Korsgaardian position is not one I wish to support, it is at least easy to make sense of her suggestion that first-personal reasons are special kinds of second-personal ones: those that obtain when the first person’s “thinking self” stands in (second-personal) relation to his/her “doing self”. Contrast this with the Darwallian position, which is – the other way around – that second-personal reasons are versions of first-personal ones. If first personal reasons are reasons possessed by an agent because of some feature of that agent’s self (such as my desire to obtain my PhD; see Ch.1, s.3.1), then it’s unclear how second-personal reasons could ever constitute a subset of first-personal reasons, since they will only ever comprise of reasons owed by the agent to him-/herself, which are not second-personal as Darwall has defined them.

⁶⁶ Darwall 2006, p.9, as cited above.

“whose formulation includes an ineliminable reference to the agent for whom they are reasons”,⁶⁷ where this definition will cover both first- and second-personal reasons. Recall that I have instead chosen to call such reasons ‘person-regarding’ (again, where both first- and second-personal reasons will count as person-regarding).⁶⁸ I trust that my motivation in abandoning the old ‘agent-relative’/‘agent-neutral’ terminology is now clear: consider Darwall’s later use of the term ‘agent-relative’, just a couple of pages after its use in the Nagelian, person-regarding sense cited above:

As second-personal reasons always derive from agents’ relations to one another, they are invariably fundamentally agent-relative.⁶⁹

Note that, if *this* is what makes second-personal reasons ‘agent-relative’ – their “always deriv[ing] from agents’ relations to one another” – then *first-personal* reasons will not count as ‘agent-relative’, since first-personal reasons do not “derive from agents’ relations to one another” but derive instead from features of an agent’s *self*, such as his/her (rational, reflectively endorsed) desires. This latter use of ‘agent-relative’, which excludes first-personal reasons, clearly deviates from Darwall’s prior employment of the term, which, following Nagel, includes first-personal reasons (and second-personal reasons).

It is this slipperiness of ‘agent-relativity’ (and of ‘agent-neutrality’⁷⁰) that has led me to adopt the terminology of ‘world-’ and ‘person-regarding’ instead. I suggest, therefore, that in his claim that “the [S]econd-[P]erson[al] stance is a version of the first-person[al] standpoint” Darwall is most charitably interpreted as claiming that second-personal reasons are, like first-personal reasons, *person-regarding*. As person-regarding, second-personal reasons are different to world-regarding reasons and similar to first-personal reasons in that they depend for their existence on features of an agent’s self. However, second-personal reasons differ from their person-regarding counterparts – namely, first-personal reasons – in that they depend for their existence on not just brute or self-reflective features of the agent’s self (such as his/her desires) but in particular on the agent’s relationship to another agent *qua* agent – as Darwall puts it “second-personal reasons always derive from agents’ relations to one another”.⁷¹ Thus, I suggest that Darwall’s curious remark is better read as claiming that

Since second-personal reasons are always fundamentally [*person-regarding*], the [S]econd-[P]erson[al] stance is a version of [the same class of reason of which] the first-person[al] standpoint [is a version, namely, person-regarding reasons]⁷²

And his later explanation of the first-/second-personal relationship becomes:

⁶⁷ Darwall 2006, p.6, fn.9, following Nagel 1986, pp.152-153.

⁶⁸ See Ch.1, s.1.1.

⁶⁹ Darwall 2006, p.8.

⁷⁰ See Lerm 2013.

⁷¹ Darwall 2006, p.8.

⁷² Darwall 2006, p.9.

What the [S]econd-[P]erson[al] stance excludes is [...] first-personal thought that lacks an addressing, second-personal aspect.

Thus, although second-person[al] address is always also [*person-regarding*], it is never merely first-personal. One can occupy a first-person[al] perspective [...] without explicitly addressing anyone.⁷³

If we read Darwall in this way, then his remarks are consistent with the rest of what he has to say about second-personality. Even though both are person-regarding, first-personal reasons are distinct from second-personal reasons in that only in the latter case is a reason *addressed* by the second person to the first person (rather than merely being ‘addressed’ by the first person (to himself) *about* the second person, as in the Contractarian case).

If we read Darwall in this way – as observing this proper distinction between first- and second-personal reasons – then my pointing out that Mill’s Point entails that some obligations are *first-personal*, as I did above,⁷⁴ is indeed a strike against Darwall. Since second-personal reasons are *not* “versions of” first-personal ones, the fact that Mill’s Point entails first-personal obligations when an agent holds himself accountable “by the reproaches of his own conscience” means that Mill’s Point does not entail the requisite *second-personal* reasons.

1.4. Loosening the accountability-obligation connection

I raise one final possible defence of Darwall against my “reproaches of his own conscience” objection, although I will not deal with it here; I deal with it in the following chapter, where a similar objection may be raised. The objection is this: my “reproaches of his own conscience” objection has effectively been that an agent’s holding him-/herself accountable entails that the agent was first-personally, not second-personally, obligated. But, why think this? It depends on the presumption that the individual with the authority to hold the agent accountable is the *same* individual to whom the agent was obligated; that is, that I hold myself accountable “by the reproaches of [my] own conscience” because I was obligated to *myself*, that is, first-personally. But, why jump to this conclusion? Why could it not be that, although I hold myself accountable, I am obligated to *another* – that is, *second-personally*? My assuming, as I have done in mounting this “reproaches of his own conscience” objection, that an agent’s liability to being held accountable by him-/herself entails that he/she was obligated to him-/herself – that is, was obligated first-personally – was, according to this defence, too quick; an agent’s liability to being held accountable by him-/herself is compatible with his/her being obligated to another – that is, second-personally obligated. And, that an agent was obligated *to another* (never mind by whom that agent is liable to being held accountable) is all that

⁷³ Darwall 2006, pp.9-10.

⁷⁴ See s.1.1, above.

Darwall needs in order to establish his Second-Personal conclusion that moral reasons have a bipolar nature.

In the following chapter, I go on to consider a similar kind of response to my upcoming moral community objection, according to which I have construed the accountability-obligation connection of Mill's Point too 'tightly', such that I have presumed that an agent's liability to being held accountable by an individual entails that the agent is obligated to *that particular* individual. The objection contends that the accountability-obligation connection is really looser than this, such that an agent's liability to being held accountable entails only that the agent is obligated to *some* individual, who need not be the same individual with the authority to hold the agent accountable. All Darwall needs, after all, in order to establish his Second-Personal conclusion that moral reasons have a bipolar form, is that the agent is obligated *to* another – any other. If this looser accountability-obligation connection is correct, an agent's holding himself accountable "by the reproaches of his own conscience" is no problem for Darwall, as I have argued it is; it may be that only the agent has the authority to hold himself accountable, even though he is obligated not to himself but to another – that is, second-personally. Thus, Darwall's Mill's-Point argument for his Second-Personal account is not undermined by the presence of first-personal obligation, as I have argued in this section, since first-personal accountability needn't entail first-personal obligation.

However, I go on to counter this upcoming defence against my moral community objection; the proposed looser accountability-obligation connection, I argue, isn't plausible. By extension, I will counter this present defence against my "reproaches of his own conscience" objection in the same way. Let us turn, then, to my second major objection to Darwall's Mill's Point argument: the moral community objection.

2. The moral community objection

In section 1 of this chapter, I objected to Darwall's Mill's-Point argument on the grounds that, if he allows that we may be liable to being held accountable by ourselves only, then his conclusion that obligation is second-personal does not follow. In this section, I continue the line of thought that the kind of individuals by whom Darwall allows that agents may be held accountable causes problems for Darwall's Mill's-Point argument. Whereas, in the previous section, this problematic individual was the agent him-/herself, in this section it is the general moral community. It is obvious that the latter aspect of Mill's Point – namely, an agent's liability to being held accountable by the general moral community – avoids the problem faced by the allowance that an agent may hold him-/herself accountable: since liability to being held accountable by the general moral community clearly entails

that the agent is held accountable by individuals *other* than him-/herself, this aspect of Mill's Point avoids the charge of first-personality.⁷⁵ However, I argue that liability to being held accountable by the general moral community faces problems of its own.

Recall that Mill's Point is that there is a "conceptual connection between moral obligation and responsibility."⁷⁶ That is, (part of⁷⁷) what it means to say that I am morally obligated to, say, keep my promise to Sally is that, if I were to break my promise to Sally, Sally would have the authority to hold me morally accountable for my bad behaviour. This holding accountable may take various forms,⁷⁸ but let us here, for the sake of expedience, say that Sally's holding me accountable for breaking my promise to her takes the form of her telling me off for my bad behaviour. According to Darwall's Mill's-Point argument, my liability to being told off by Sally reveals that my obligation was *to* Sally; that is, that it was second-personal.

Interestingly, however, Darwall allows that it is not only the *victim*⁷⁹ of the moral wrong – Sally – who has the authority to hold the perpetrator accountable, but any member of the moral community. As Darwall was quoted in the introduction to this chapter:

Matters of moral obligation are *the moral community's* "business," [...]. When we are morally obligated, we are not morally free to act otherwise; *members of the moral community* have the authority to hold us responsible if we do.⁸⁰

And, in a more recent paper, Darwall is even more explicit that "the victim of wrongdoing [...] has no special standing others do not have to blame the wrongdoer or to hold him responsible through (what P.F. Strawson called "impersonal") reactive attitudes such as indignation."⁸¹ That is, if I break my promise to Sally, it is not only *Sally* – the victim – who has the authority to hold me accountable; any member of "the moral community" might do so – and might do so with all the fullness of the authority of Sally herself, since Sally "has no special standing others do not". Thus, some third-person observer, Barry, might just as well tell me off for breaking my promise to Sally as Sally herself might do.

⁷⁵ See s.1, above.

⁷⁶ Darwall 2006, pp.27-28, as cited in the introduction to this chapter.

⁷⁷ See fn.9, above.

⁷⁸ Mill 1863/1998, Ch. 5, as cited by Darwall 2006, p.27, as cited in my s.1 above.

⁷⁹ I do not, however, mean to imply by the use of the term 'victim' that this point applies only to cases of serious moral wrongdoing, such as physical assault, rape or murder – the sufferers of which we would ordinarily call *victims*. These points do apply to sufferers of less serious moral wrongs too, such as deception, queue-barging or promise-breaking. Although we wouldn't ordinarily (non-sarcastically) call someone a '*victim* of queue-barging', the term is used here as a useful shorthand for: a sufferer of a moral wrong.

⁸⁰ Darwall 2006, p.27, citing Baier 1966, p.223 (emphases added), as cited in the introduction to this chapter.

⁸¹ Darwall 2010a, p.262, referencing Strawson 1962/1968.

In Ch.4 I go on to argue against Darwall's reading of Strawson on the point of impersonal reactive attitudes.

Intuitively, this seems correct insofar as we do indeed hold agents accountable even when we are not ourselves the victims of their wrongdoings. (Here a key example is our practice of criminal law.) However, I will go on to question whether Darwall's Second-Personal account is indeed the best account of this practice of holding agents accountable. Barry surely does possess the authority to tell me off for breaking my promise, but I will question whether the second-personality of obligation is the best account of this authority. While the Second-Personal account does indeed strike me as a plausible account of *Sally's* authority to hold me accountable – where Sally is the specific individual to whom I have made a promise⁸² – it does not strike me as a plausible account of *Barry's* authority to do so, yet Darwall holds (as above) that a member of the general moral community, such as Barry, possess exactly the same fullness of authority to hold me accountable as Sally does herself.

This moral community objection will be separated into two parts, the first comprising the remainder of this present chapter and the second comprising the entirety of the following chapter. Although these two parts of the moral community objection are split across chapters, I do not regard them as capturing two distinct arguments. (The chapter is split only because the detail of the argument would otherwise have made it a very long chapter.)

Chapter 4 will scrutinize the logic of Darwall's Mill's-Point inference. According to Darwall, we can infer that my obligation to Sally was second-personal because I am liable to being held accountable by Sally if I break my promise to her. However, if Darwall allows – as he does – that a member of the general moral community, such as Barry, also has the authority – the very same authority – to hold me accountable if I break my promise to Sally, then, by the same logic, we should be able to infer that I am also second-personally obligated to *Barry* to keep my promise. But, since I have not made a promise to Barry (but to Sally), it is far from clear that I am so obligated to Barry. Darwall's allowance that we are liable to being held accountable not only by the victims of our actions but also by members of the general moral community, therefore, causes problems for his Mill's-Point argument: it demonstrates that the inference from Sally's authority to hold me accountable to the conclusion of the second-personality of my obligation to her is not well-founded.

And, the root of the problem (outlined above) that will be the subject of the following chapter is, I think, the discordance between second-personality, on the one hand, and, on the other, our liability to be held accountable by an entire moral community – not only by those, such as Sally, to whom obligations are (purportedly) second-personally owed, but also by those third-person observers to whom no obligations are owed, such as Barry. The Mill's-Point inference does not go through, as I

⁸² I noted the intuitive plausibility of such an account when I originally introduced the case of Sally and my making a promise to her in s.1, above.

will argue for the remainder of this chapter, because the notions of second-personality and of liability to being held accountable by an entire moral community are essentially in tension with one another.

2.1. Discordance

Darwall's Mill's-Point argument infers from the fact that an agent is liable to being held accountable by an individual to the conclusion that the agent was therefore obligated to that individual, second-personally. If, therefore, Darwall allows – as he, following Mill, does allow – that an agent may also be held accountable by any member of the general moral community, then it must, by the same logic, also follow that the agent is second-personally obligated to each one of those members of the general moral community. Now, there is nothing logically incoherent about the idea that an obligation may be owed to all others – every single member of the general moral community – but it does seem to dilute the idea of the *second-personal*. What is 'second-personal' about owing an obligation to *everybody*? On the other hand, there is something strikingly plausibly second-personal about making a promise to an individual, and therefore being obligated to *that* individual, and therefore being liable to being held accountable by *that* individual. Making a promise to Sally and being held accountable to it *by Sally* appears to be very relevantly different to making a promise to Sally and being held accountable to it by somebody else, like Barry. The idea that only Sally has the authority to hold me accountable for the promise I made to her seems to *fit* the idea of the second-personal much better than does the suggestion that a third-person observer such as Barry also has such authority. But, this is in tension with our intuition that members of the general moral community do have the authority to hold agents accountable, even when they are not themselves the victims of those agents' actions.

This complaint is, admittedly, a rather vague one, amounting to not much more than the mere sense that second-personality doesn't fit very well with the idea of an entire moral community's authority to hold an agent accountable.⁸³ It is, however, very much a *strong* sense of discord, and it is a sense that at least Lavin and Wallace appear to share. As he was cited in the introduction to this chapter, Lavin asks: "But does the idea of [...] "the moral community" as a maker of demands and an occupier of the second-person[al] standpoint [make sense]?"⁸⁴ As I have noted, the suggestion that an entire moral community might have second-personal authority over an agent is not logically incoherent, and to that extent it does 'make sense'; however, it does not make *as much* sense as the suggestion

⁸³ This complaint is bolstered by the fact, for which I shall go on to argue, that Realism can account for our practice of holding each other accountable just as well as Darwall's Second-Personal account can – and that Realism is in accordance with the idea of universal moral obligation, where the Second-Personal account is not. This is, however, the subject of Ch.5.

⁸⁴ Lavin 2008, as cited in the introduction to this chapter.

that the individual with the authority to hold me accountable is the one to whom I am obligated, and that she possesses this authority *because* I am obligated to her. (And, to this end, Lavin notes how Darwall employs examples only of one-on-one second-personal relationships to “[talk] us into the idea of the second-person[al] standpoint”,⁸⁵ such as that of the foot-treader⁸⁶ or the sergeant and her private.⁸⁷) Wallace appears to have in mind the same worry when he writes:

the person who is wronged by you has a privileged basis for complaint against you, an objection to your conduct that is not shared by mere observers to what was done. The notion that someone in particular has been wronged by your action is conceptually connected to the idea that the wronged party has special ground for complaint [...].⁸⁸

That is, if my obligation to keep my promise is owed, second-personally, *to Sally*, as Darwall would claim it is, then this would seem to commit us to the claim that, if I break my promise, it is Sally, “the person who is wronged by [me]”, who “has a privileged basis” or “special ground for complaint” – not Barry, or any other member of the general moral community. Yet Darwall, following Mill, allows – as we intuitively would want to allow – that any member of the general moral community besides Sally also has a basis for complaint against me, and indeed that Sally “has no special standing others do not have to blame the wrongdoer or to hold him responsible [...]”.⁸⁹ There is a tension here, between Darwall’s account of a moral reason as an obligation to a second person, on the one hand, and, on the other hand, the intuition – one Darwall, following Mill, readily accepts – that we are liable to being held accountable for our moral transgressions even by those who are not themselves the ones wronged by our actions.

To this complaint of discordance between second-personality and accountability to an entire moral community, I find in Darwall’s writings subsequent to *The Second-Person Standpoint* two possible responses. The first possible response is not explicitly cast as a rejoinder to the present complaint, but it nevertheless appears to be worthy of consideration as such, although I will ultimately find it to be unhelpful. The second possible response, on the other hand, is explicitly presented as such, and is one I will find to be more successful than the first. However, as I will ultimately go on to argue in Chapter 5, it is a response that I think is also open to the Realist.

2.1.1. Deriving individual authority from representative authority

In a paper subsequent to *The Second-Person Standpoint*, ‘Moral obligation: form and substance’, Darwall appears to have come to the realization that (what I have called) the Mill’s-Point argument will not work. He writes:

⁸⁵ Lavin 2008, as cited in the introduction to this chapter.

⁸⁶ Darwall 2006, pp.5-7, as referenced in the introduction to this chapter.

⁸⁷ Darwall 2006, p.259-273, as referenced in the introduction to this chapter.

⁸⁸ Wallace 2007, p.29.

⁸⁹ Darwall 2010a, p.262, as cited in s.2 above.

[T]he form of moral obligation *period*, as we might call it, involves a fundamental answerability we have to everyone, ourselves included, as representative persons, hence their and our representative authority[.] [On the other hand], bipolar obligations *to others* [...] implicate a *different* standing or authority that others have as individuals to hold us answerable, hence imply[ing] their individual authority.⁹⁰

That is, Darwall accepts that there appears to be a difference in kind between what he gives the useful terms ‘individual authority’ and ‘representative authority’; that is, (respectively) between (a) the kind of authority to hold me accountable that Sally possesses as the specific “individual” to whom the promise was made, and (b) the kind of authority to hold an agent accountable that anybody, like Barry, possesses just in virtue of being a “representative” of the general moral community – but where this is a “*different* standing or authority” to that which “others have as individuals to hold us answerable”. Given this distinction between individual and representative authorities, it becomes clear that the Mill’s-Point inference will not entail the second-personal nature of all moral obligation. As Darwall writes:

Assuming that the form of moral obligations period involves a fundamental answerability to one another and ourselves as representative persons, therefore, does not commit us to the assumption that these obligations are, or even that they entail any that are, *owed to one another*.⁹¹

That is, Darwall appears to admit that his Mill’s-Point argument is unsuccessful. Just because moral reasons are ones which “[involve] a fundamental answerability” – that is, are ones for which agents are liable to be held accountable – it does not follow “that these obligations are, or even that they entail any that are, *owed to one another*” – that is, Mill’s Point does not entail the conclusion that moral reasons are second-personal. This appears to be a ground-breaking concession for my purposes: the Mill’s-Point argument fails, and Darwall admits it.

However, this is no straightforward victory for me yet, because Darwall also promises in ‘Moral obligation’ to show that:

an argument can be made from the equal *representative authority* implicit in this account of moral obligation’s form [...] to a basic reciprocal, bipolar obligation that entails every person’s *individual authority* to demand such treatment.⁹²

That is, even though Darwall concedes a distinction between the kinds of authority possessed over me by Sally and by Barry, he aims to show in this paper that the latter entails the former: that the representative authority of the general moral community entails that of the individual. Darwall’s concession, therefore, of a distinction in kind between Sally’s authority to hold me accountable for breaking my promise to her and Barry’s authority to do so does not spell the downfall of the Mill’s-Point argument. Instead, Darwall aims to deduce the second-personality of my obligation to Sally on

⁹⁰ Darwall 2010c, p.33 (some emphases removed).

⁹¹ Darwall 2010c, p.33 (some emphases removed).

⁹² Darwall 2010c, p.36.

the basis of Barry's representative authority to hold me accountable. If Darwall can derive individual authority from representative authority, then the appearance of any discordance between Sally's and Barry's authorities to hold me accountable will have been dispelled. This new argument promises to reach the same conclusion as the original Mill's-Point argument – namely, that there exist second-personal reasons – albeit by a more circuitous route than originally envisaged. In the end, however, as I will go on to explain, this unfortunately depends on a misreading of Darwall's use of 'individual authority'.

Before I go on to present Darwall's proposed derivation of individual authority from representative authority, let me note that this is very much a secondary aim of the paper presently under consideration. As that paper's title – 'Moral obligation: form and substance' – would suggest, its primary objective is to derive the substance of (at least some) moral obligations on the basis of their form – where, of course, Darwall takes their form to be second-personal. In Darwall's own words:

Beginning from an analysis of moral obligation's *form* that I defend in *The Second-Person Standpoint* as what we are answerable for as beings with the necessary capacities to enter into relations of mutual accountability, I argue that this analysis has implications for moral obligation's *substance*.⁹³

That is, the primary objective of 'Moral obligation' takes for granted the account of second-personal reasons previously propounded by Darwall in *The Second-Person Standpoint*, and on that basis aims to derive the content of (at least some) moral obligations. The primary objective of the paper is to derive (at least some) substantive moral norms from the Second-Personal account of their nature, not to derive individual authority from representative authority, which is merely a secondary aim of the paper.

The paper's primary argument, for substantive moral norms, is strongly reminiscent of Kant's *Groundwork for the Metaphysic of Morals*.⁹⁴ Very briefly, this primary argument of the paper is as follows. Darwall begins by noting that our practice of holding each other accountable presupposes that those held accountable are free moral agents. Referencing Strawson, he writes:

A central lesson of Strawson's discussion is that [...] [i]t is impossible warrantably to hold others answerable without allowing them the interpersonal space to take responsibility for themselves by making their own autonomous choices.⁹⁵

And, on the basis of this – the form of our second-personally holding each other accountable – it follows that we ought "not to usurp others' moral agency or otherwise undermine the conditions of

⁹³ Darwall 2010c, p.31.

⁹⁴ Kant 1785.

Curiously, however, Darwall does not in this argument reference Kant, but references instead the Hobbesian construal of the "golden rule" (Hobbes 1651/1994, Ch.15, par.35, as referenced by Darwall 2010c, pp.41-42).

⁹⁵ Darwall 2010c, p.43, referencing Strawson 1968.

free moral choice.”⁹⁶ That is, a substantial moral norm is entailed just by the form of moral reasons as those for which we are second-personally answerable: “if there are any moral obligations at all, then there must exist a basic *pro tanto* obligation not to undermine one another’s moral autonomy.”⁹⁷ And, Darwall adds optimistically, we will be able to derive further contentful moral norms as we extend similar reasoning “significantly wider, to include, for example, obligations of mutual aid and non-maleficence, as well as others.”⁹⁸

But, note that achieving this primary aim of ‘Moral obligation’ clearly takes for granted moral reasons’ second-personal form. Darwall is perfectly explicit, as he was quoted above, that this paper “[begins] from an analysis of moral obligation’s *form* that [he] defend[ed] in *The Second-Person Standpoint* [...]”⁹⁹ This brings to light two problems. Firstly, the primary argument of ‘Moral obligation’ is not an argument available to us at present. Recall that we are currently (still) in the business of trying to establish that there are such things as second-personal reasons in the first place: Mill’s Point is supposed to establish the second-personal nature of moral reasons on the basis of our practice of holding each other accountable for our moral transgressions; but I have questioned whether this argument can work, given the discordance between second-personality and obligation to an entire moral community, again leaving us without good reason for believing that there are indeed such things as second-personal reasons; so we are not yet in a position to accept an argument that takes for granted the existence of second-personal reasons as its starting premise.

The second problem brought to light by the reliance of ‘Moral obligation’ on the existence of second-personal reasons is really more of a clue to a problem than a problem *per se*. As we have noted, the primary argument of the paper takes for granted the second-personal form of moral reasons, and on that basis tries to derive (at least some) substantive moral norms. On the other hand, the secondary aim of that same paper is to derive individual authority on the basis of representative authority – where, as cited at the beginning of this section, individual authority is “a *different* standing or authority that others have as individuals to hold us answerable” to our “bipolar obligations *to others*”.¹⁰⁰ But, note the tension here: the primary aim of the paper takes second-

⁹⁶ Darwall 2010c, p.43.

⁹⁷ Darwall 2010c, p.31.

I must note that the moral obligation “not to usurp others’ moral agency” (Darwall 2010c, p.43) is nothing like the substantive moral obligation I was led to expect by the paper’s introduction. The moral obligation “not to usurp others’ moral agency” remains very general and formal. It is far from as contentful as, for instance, Kant’s similarly-reasoned *Groundwork* derivation of our duty not to commit suicide. Darwall does, however, at least note that “Obviously, this is only a beginning sketch of how one might seek to derive moral obligations” (2010c, p.44).

⁹⁸ Darwall 2010c, p.44.

⁹⁹ Darwall 2010c, p.31, as cited above.

¹⁰⁰ Darwall 2010c, p.33 (some emphases removed), as cited above.

personality for granted,¹⁰¹ yet the secondary aim is to *derive* the individual, second-personal authority by which second persons give “bipolar obligations *to others*”. How could we, in the very same paper, take the former for granted without first having established the latter? This is a clue to a problem – where that problem is an initial misreading of Darwall’s use of ‘individual authority’.¹⁰² As I will go on to explain, in deriving “individual authority” from representative authority, Darwall is *not* in fact attempting to derive second-personal authority, as I originally presented him as doing. Instead, second-personality is taken for granted throughout the paper, by both its primary and secondary arguments. As will become clear, Darwall does not by ‘individual authority’ mean: the special authority of the individual obligator, Sally (an authority not possessed by Barry) to hold me accountable for breaking my promise to her; he instead means: the authority of the individual *obligated*, me, the first person. As such, the secondary argument of ‘Moral obligation’ is ultimately irrelevant to the present discordance objection.

So, what is the secondary argument of ‘Moral obligation’? How does Darwall purportedly derive individual authority from representative authority? The whole argument is presented in one, short paragraph, which appears as the second-last paragraph of ‘Moral obligation’. It is a secondary argument of that paper because it follows from a premise already employed in the defence of its primary argument. That premise is that our practice of holding others accountable commits us to the presupposition of their moral autonomy. As Darwall was cited, above: “It is impossible warrantedly to hold others answerable without allowing them the interpersonal space to take responsibility for themselves by making their own autonomous choices.”¹⁰³ (Which, in the primary argument, Darwall took to entail the substantive moral norm that we ought “not to usurp others’ moral agency or otherwise undermine the conditions of free moral choice.”¹⁰⁴) The key to the secondary argument is that Darwall equates others’ moral autonomy with their “[taking]

¹⁰¹ It may be suggested that ‘Moral obligation’ (Darwall 2010c) does not take for granted the second-personal form of moral reasons, but instead tries to argue for it, by demonstrating as an inference to the best explanation how such a second-personal account can yield what we intuitively take to be our moral obligations. Although such an argument would be welcome, it is not a justified reading of ‘Moral obligation’. If this were the form of argument of that paper, then: (1) Darwall would have to have undertaken to show how other rival accounts fail to yield what we intuitively take to be our moral obligations as well as his Second-Personal account does, but Darwall undertakes to do no such thing, and (2) Darwall would surely have had to demonstrate the derivation of intuitively much more direct moral obligations than just the very general moral obligation “not to usurp others’ moral agency” (Darwall 2010c, p.43; see fn.97, above), such as the obligation not to murder, for instance. ‘Moral obligation’ does not have the form of an inference to the second-personal nature of moral reasons as the best explanation of what we do intuitively take to be our moral reasons; instead, it simply takes for granted that they do indeed have such form – which, I argue, is problematic.

¹⁰² This is a misreading, which, I claim, is not unjustified, given Darwall’s own presentation of the differences between individual and representative authority, as cited at the beginning of this section 2.1.1.

¹⁰³ Darwall 2010c, p.43, referencing Strawson 1968, as cited above.

¹⁰⁴ Darwall 2010c, p.43, as cited above.

responsibility for themselves”.¹⁰⁵ What it is to presuppose your moral autonomy when I hold you morally responsible, therefore, is to allow you “the interpersonal space to take responsibility for [yourself]” – that is, to hold *yourself* responsible. And, so, the secondary argument follows, according to which:

We can hardly warrantedly hold [an agent] responsible as a representative person for taking individual responsibility *for himself* without thereby committing ourselves to according him a distinctive *individual authority* [...].¹⁰⁶

That is, according to this secondary argument, we can deduce an agent’s individual authority on the basis of the representative authority of the general moral community because, in exercising our (representative) authority to hold an agent morally responsible, we are committed to presupposing that agent’s (individual) authority to hold *him-/herself* responsible, where this is equated with the agent’s moral autonomy.

This reveals that what Darwall really has in mind by ‘individual authority’ is just the agent’s authority over *him-/herself*. ‘Individual authority’, as it is used here, is the authority of only one “individual”, since for every agent there is just one individual agent’s self. Importantly, this is *not* the special authority of Sally to hold me accountable for breaking my promise to her (where this is to be contrasted with the authority of a member of the general moral community, such as Barry); rather, it is just *my* ‘individual authority’ to hold myself accountable, which is presupposed when anybody – be it Sally or Barry – holds me accountable.

This is, of course, just Pufendorf’s Point, dressed up in different vocabulary: our holding agents (genuinely) morally accountable entails that those agents hold *themselves* morally accountable.¹⁰⁷ Despite initial appearances, then, this is no new argument at all. It does not attempt to deduce Sally’s “individual authority” to hold me accountable for breaking my promise to her on the basis of Barry’s representative authority to do so; rather it deduces *my* individual authority to hold *myself* accountable on the basis of Sally’s *or* Barry’s representative authority to do so. When we understand ‘individual authority’ correctly, it becomes clear that this secondary argument of ‘Moral obligation’ has nothing to say to the discordance between, on the one hand, the (purported) second-personality of my obligation to Sally and, on the other, the entire moral community’s authority to hold me accountable if I break my promise, because this secondary argument had nothing to say about *Sally’s* special authority to hold me accountable in the first place.

¹⁰⁵ This is a questionable identity claim. I do not, however, question it here, as the argument at hand will ultimately turn out to be irrelevant to the Mill’s-Point argument presently under examination.

¹⁰⁶ Darwall 2010c, p.44.

¹⁰⁷ See Ch.1, s.6.1.2.

Despite initial appearances, therefore, this argument of ‘Moral obligation’ will not assist in dispelling the discordance between, on the one hand, the (purported) second-personality of my obligation to Sally and, on the other, the entire moral community’s authority to hold me accountable if I break my promise. Let us turn, therefore, to a second possible response to this discordance, one which also appears in a paper by Darwall published subsequent to *The Second-Person Standpoint*, and one which will prove to be relatively more successful.

2.1.2. Forgiveness as special authority

I have complained that there is a discordance between, on the one hand, the (purported) second-personality of my obligation to Sally to keep the promise that I have made to her and, on the other hand, my liability to being held accountable by an entire moral community, consisting of people such as Barry, to whom I have made no promises. This is exacerbated by the fact that Darwall insists that there is absolutely no difference in kind between Sally’s authority to hold me accountable for breaking my promise to her and Barry’s authority to do so, claiming that Sally “has no special standing others do not have to blame the wrongdoer or to hold him responsible [...]”¹⁰⁸ Although, intuitively, any member of the moral community does have the authority to hold me accountable when I enact a moral wrong – even when that member of the moral community is not him-/herself the direct victim of my trespass – it does nevertheless also appear to be the case that the victim possesses a *special* authority to hold me accountable. Wallace puts the worry like this:

the person who is wronged by you has a privileged basis for complaint against you, an objection to your conduct that is not shared by mere observers to what was done. The notion that someone in particular has been wronged by your action is conceptually connected to the idea that the wronged party has special ground for complaint [...].¹⁰⁹

But, Darwall has a ready response to Wallace’s worry. Although Darwall allows that “the victim of wrongdoing [...] has no special standing others do not have to blame the wrongdoer”,¹¹⁰ he *does* also allow that the victim has some “privileged” or “special” authority, as Wallace has required.¹¹¹ And, this special authority is, according to Darwall, the “distinctive standing to [...] *forgive an injury*”.¹¹² That is, the victim does possess some special standing – but it is not, as Wallace presumes, the special standing to hold accountable (since, Darwall claims, any member of the moral community has the authority to do that). Rather, it is the special standing to choose whether to *forgive* the

¹⁰⁸ Darwall 2010a, p.262, referencing Strawson 1968, as cited in s.2.1, above.

This claim would appear to be at odds with Darwall’s “admission” (considered in s.2.1.1, above) that there is a distinction to be drawn between Sally’s “individual authority” and Barry’s “representative authority”. However, when we read ‘individual authority’ correctly – as referring not to Sally’s authority over me but to my own authority over myself – then the apparent contradiction dissipates.

¹⁰⁹ Wallace 2007, p.29, as cited in s.2.1, above.

¹¹⁰ Darwall 2010a, p.262, as cited above.

¹¹¹ Wallace 2007, p.29, as cited above.

¹¹² Darwall 2010a, p.262 (emphasis added).

wrongdoer or not. This “distinctive standing” of the victim is – as Darwall is fond of pointing out – reflected in the practice of tort law, where “it is up to the injured party [i.e. the victim] whether or not to bring a tort action at law”.¹¹³ On the other hand, when it comes to criminal law, “criminal actions are appropriately brought by the people and their representatives”¹¹⁴ – that is, by the moral community – and it is *not* “up to the injured party” whether or not to bring such an action.

So, Darwall appears to achieve the best of both worlds here: a moral wrongdoer may be held accountable not only by the victim of his/her moral wrong but also by any member of the moral community, and in such a way that retains Wallace’s requirement of the victim’s special authority – where this is the special authority to choose whether or not forgive (or bring a tort action against) the wrongdoer. Call this Darwall’s ‘forgiveness response’. If this response is acceptable, it will dispel my worries of discordance between second-personality and liability to being held accountable by an entire moral community. On this forgiveness account, the (purported) second-personality of my obligation to Sally is not captured by the fact – as we originally took it to be – that Sally possesses an authority Barry does not to hold me accountable if I break my promise to her; rather, Sally and Barry possess the same authority to hold me accountable, but the second-personality of my obligation to Sally – the specialness of my standing in a relation of obligation *to Sally*, not to Barry – is captured by the fact that Sally, and not Barry, has the authority to *forgive* me my broken promise if she chooses to do so.

However, I have some doubts about the acceptability of this forgiveness response. In what follows, I raise a few quick objections, but I postpone the final assessment of Darwall’s forgiveness response until Chapter 5.

Firstly, it’s unclear to me that it is the case that it is only the obligator – for instance, Sally – who possesses the authority to forgive. Is it not possible that a third party besides the specific individual to whom the obligation is (purportedly) owed might be able to forgive the agent? As Paul Hughes notes, “Despite the widespread assumption that only the primary victim of wrong has the standing or authority to forgive wrongdoers, a number of philosophers have challenged it.”¹¹⁵ There are two primary candidates as cases of such third-party forgiveness: cases in which the State officially pardons a convicted criminal, and cases in which God forgives a sinner.¹¹⁶ Take the latter, for instance: it is not an uncommon belief that “the Son of Man has authority on earth to forgive

¹¹³ Darwall 2010c, p.33, fn.2.

¹¹⁴ Darwall 2010c, p.33, fn.2.

¹¹⁵ Hughes 2014, s.5.3, referencing Murphy 2009, Pettigrove 2009, and Radzik 2010.

¹¹⁶ Hughes 2014.

sins”.¹¹⁷ And, God is commonly held to have this authority to forgive sins even when those sins (such as rape, murder, or even breaking promises) are – at least on the face of it – perpetrated against fellow human beings rather than against *God* himself (where more plausible candidates of sins perpetrated against God himself might be sins such as idolatry or blasphemy). That is, God is apparently a third party to many of our sins,¹¹⁸ yet he nevertheless is held to possess the authority to forgive them. Could a third party such as Barry not, likewise, possess the authority to forgive me for breaking my promise to Sally?

In presuming that it is only the obligator who possesses the authority to forgive, Darwall is surely drawing upon Strawson, who lists forgiveness as a “personal” reactive attitude,¹¹⁹ rather than an “impersonal” reactive attitude, where personal reactive attitudes are those evinced by the parties actually involved rather than by third-party observers.¹²⁰ But, again, why should we accept that forgiveness is only a personal reactive attitude, and not an impersonal one too? If God, a third-person observer, can forgive our sins against our fellow man (as above), then there’s at least a *prima facie* case for forgiveness’s possibly being an impersonal reactive attitude too. The role of forgiveness in Strawson’s ‘Freedom and resentment’ is really just as an illustration of personal reactive attitudes (where Strawson also lists “gratitude, resentment, forgiveness, love, and hurt feelings”¹²¹ as examples of the same), so it is unnecessary for Strawson there to defend the claim that forgiveness is a personal reactive attitude. If, however, Darwall wishes forgiveness to do the strong argumentative work it is supposed to do in defending his position from Wallace’s complaint, then the bare appeal to Strawson’s authority is insufficient. At the very least, more needs to be said as to why, according to Darwall, it is only Sally, and not Barry (or even God), who possesses the special authority to forgive me if I should break my promise to her.

Furthermore, Strawson lists not just forgiveness as a personal reactive attitude, but also “gratitude, resentment, forgiveness, love, and hurt feelings”¹²² – and where no indication is given that this list of examples is supposed to be exhaustive. So, we might ask, even if it can be demonstrated that

¹¹⁷ Matthew 9:1-8, New International Version.

¹¹⁸ Hughes considers the objection that “any sin against another person is ipso facto a sin against God the creator and, thus, that God’s forgiveness of human wrongdoing is really for wrongs done to Him, making God’s forgiveness of human wrongs not a genuine case of third-party forgiveness.” However, Hughes responds to this objection as follows:

[I]t is simply false to say that a wrong done to a particular person is actually a wrong done to someone else, or to God, and not to that person. Perhaps all injuries done by human beings to other human beings are in some sense also wrongs done to God, but that is consistent with the observation that wrongs endured by particular people are done to them. (Hughes 2014, s.5.3.)

¹¹⁹ Strawson 1962/2003, p.196ff.

¹²⁰ Strawson 1962/2003, p.205ff.

¹²¹ Strawson 1962/2003, p.196.

¹²² Strawson 1962/2003, p.196, as cited above.

forgiveness is only ever a personal reactive attitude and never an impersonal one (as above), why is it the authority to *forgive* in particular that makes Sally's authority over me "privileged" and "special"? Why is it not, say, Sally's hurt feelings? Unless such an account is available, Darwall's forgiveness response is strikingly *ad hoc*.

Secondly, I find Darwall's forgiveness response dubious not only in its apparently *ad hoc* appeal to the authority to forgive in particular, nor in its unsupported assertion that this is an authority possessed only by Sally, but in a more fundamental way than either of these complaints. Darwall has responded to Wallace's worry by accounting for a scenario in which Sally possesses the same authority as Barry for holding me accountable, but where Sally also has the *additional* authority to forgive me if she should choose to do so. But, it is not clear to me that this addresses the original worry. That worry, as far as I understand it, is that Sally's authority to hold me accountable is *different in kind* to Barry's – and this is a worry that is not addressed by claiming that Sally possesses an *extra* authority (to forgive) over and above the same authority (to hold accountable) which she shares with Barry. Rather, the worry is that Barry really doesn't share the same kind of authority as Sally at all: there seems – strongly, even – to be something different about Sally's holding me accountable compared to Barry's doing so, because it is Sally, and not Barry, *to whom* I am (purportedly) obligated.

However, strong as this seeming may be, the onus is of course on me and my fellow objectors, such as Wallace, to suggest *what* it is that is different about Sally's authority. Until we can suggest an alternative, Darwall's forgiveness account is the best going. This is not, however, a project I will embark upon here. For now, I simply flag the feeling that Sally's special authority over me is not an additional authority (to forgive) but an authority altogether different in kind to Barry's.

I have here raised some quick doubts about Darwall's forgiveness response. They do not yet amount to an objection strong enough to dismiss Darwall's forgiveness response to my worry regarding discordance, however. But, I do not lay this matter to rest just yet. In Chapter 5, I go on to argue that Realism can provide an account of the victim's special authority to forgive just as well as – if not better than – Darwall's own Second-Personal account. Insofar as discordance (or the dissolution thereof) is concerned, therefore, Darwall's Second-Personal account and Realism are on a par. But, before I arrive at that argument, I first go on to refine my discordance worry into a more rigorous objection, which I present as the Sally/Barry counterexample, in the following chapter. Darwall's forgiveness response may be able to dispel worries of discordance, but it is insufficient to defend his Mill's-Point argument against the Sally/Barry counterexample.

3. Conclusion

In this chapter, I have examined Darwall's so-called Mill's-Point argument. This is the argument that we can infer the existence of second-personal reasons as the best explanation of our practice of holding each other accountable. However, I have argued that Darwall's Second-Personal account struggles to account for those broader aspects of Mill's Point, namely, that an agent may hold him-/herself accountable, as might a member of the general moral community.

In the first section of this chapter, I argued that Darwall's allowance (following Mill) that we might hold ourselves accountable by the "reproaches of [our] own conscience" supports a first-personal account, rather than Darwall's desired Second-Personal account. In the second section of this chapter, I began the argument that Darwall's Second-Personal account struggles to accommodate that other aspect of Mill's Point, which is that non-victim members of the general moral community might hold an agent accountable. This moral community objection has, so far, taken the form of a charge of "discordance": the sense, shared by several reviewers, that the concept of the second-personal – that is, owing an obligation to a second person – is not harmonious with the fact that members of an entire moral community might have the authority to hold an agent accountable.

In the following chapter, I go on to refine this moral community objection into the more technical complaint that allowing that any member of an entire moral community has the authority to hold an agent accountable – even when that agent is not second-personally obligated to that particular member of the moral community – undermines the very logic of Darwall's Mill's-Point argument. This objection will be illustrated by my "Sally/Barry counterexample", which is the subject of the following chapter.

4.

The Sally/Barry countercase

According to Darwall, because a genuinely morally accountable agent is one who holds him-/herself accountable,¹ it follows that any case of a second person holding a first accountable presupposes that those first and second persons share equal, reciprocal authority. That is, any giving of any second-personal reason entails Morality as Equal Accountability.² However, I have criticized this argument in two ways. Firstly, I have argued that the argument is a *non-sequitur*: a second person's holding a first accountable presupposes only that those first and second persons share an equal authority over *the first person*, which does not get us all the way to Morality as Equal Accountability.³ Secondly, I have criticized the argument on the grounds that we should be sceptical of the existence of the so-called second-personal reasons at all. This was an argument I began in in Chapter 2.⁴

In Chapter 3, I considered an additional argument for the existence of second-personal reasons, Darwall's Mill's-Point argument. This is the argument that, because we commonly hold each other accountable for our moral transgressions, it follows that those transgressed obligations are ones owed to others – that is, are second-personal. However, I have criticized this Mill's-Point inference in two broad ways so far. Firstly, I have questioned Darwall's allowance that we might hold ourselves accountable by the “reproaches of our own conscience”:⁵ if we hold ourselves accountable, then it doesn't follow that our moral obligations are owed to others second-personally.⁶ Secondly, I have questioned Darwall's allowance that we might be held accountable not only by those wronged by our moral transgressions – for instance, Sally, to whom I have made a promise and subsequently broken it – but also by members of the general moral community, like Barry, to whom I have made no promises. I suggested that second-personality – owing an obligation to a second person – is not plausibly compatible with liability to being held accountable by an entire, general moral community.⁷

¹ See Ch.1, s.6.1.2.

² See Ch.1, s.6.2.

³ See Ch.2, s.1.

⁴ See Ch.2, s.2.

⁵ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27.

⁶ See Ch.3, s.1.

⁷ See Ch.3, s.2.1.

In this present chapter, I continue the moral community objection of Chapter 3.⁸ Whereas, previously, I argued for the sense that Darwall's Second-Personal account is discordant with the claim that an agent is liable to being held accountable by any member of the entire moral community,⁹ I aim here to argue, more forcefully, that the very logic of Darwall's Mill's-Point argument is flawed. In particular, I argue that Darwall's inference from my liability to being held accountable by Sally to the second-personality of my obligation to Sally is not justified, and this is confirmed by Darwall's allowance that an agent may be held accountable by members of the general moral community, like Barry, to whom I owe no obligation. I make my case by way of what I will call the 'Sally/Barry counterexample', as presented below.

If the Sally/Barry counterexample is successful, it will block Darwall's Mill's-Point argument, thereby again leaving us without any good grounds on which to accept that there are such things as second-personal reasons in the first place.

1. A counterexample to Darwall's Mill's-Point argument

The Sally/Barry counterexample itself is straightforward, although my subsequent defence of it will be intricate. The counterexample is just this: I make a promise to Sally, and I am subsequently obligated to keep that promise to Sally. Sally, plausibly, has the authority to hold me accountable, should I break my promise to her. And, according to Darwall's Mill's-Point inference, this reveals that my obligation to keep my promise is a second-personal obligation owed *to Sally*.¹⁰ So far, so good. However, Darwall also allows that a member of the general moral community – for instance, Barry – likewise possesses the authority to hold me accountable for breaking my promise to Sally. Now, if my liability to being held accountable by Sally entails that I am second-personally obligated to her, then my liability to being held accountable by Barry must similarly entail that I am second-personally obligated to *him*. That is, if the logic of Darwall's Mill's-Point inference is sound, and if both Sally and Barry possess the authority to hold me accountable, then it must follow that I am second-personally obligated to Barry just I am second-personally obligated to Sally to keep my promise to her. However, I have made no promises to Barry, and so, I maintain, I am not obligated to him (let alone obligated to him *second-personally*). When it comes to Barry, my liability to being held accountable by him does not entail my obligation to him. Since the Mill's-Point inference from accountability to obligation does not go through in Barry's case, it cannot go through in Sally's case either, and so, I argue, it does not support Darwall's conclusion that I am second-personally obligated (to either Sally

⁸ See Ch.3, s.2.

⁹ See Ch.3, s.2.1.

¹⁰ See the introduction to my Ch.3.

or Barry). That is, we are again left without any good reason to believe that there are in fact such things as second-personal reasons. This is the Sally/Barry counterexample.

Below, I consider two possible lines of defence against the Sally/Barry counterexample. The first possible line of defence tries to accommodate the Sally/Barry counterexample, by arguing that the fact that I am not obligated to Barry in particular is not a problem for Darwall's Mills-Point argument. However, I go on to argue in response that Darwall cannot accept this 'looser' accountability-obligation connection. The second possible line of defence, on the other hand, attempts to block the Sally/Barry counterexample altogether, by contending that, despite appearances to the contrary, I actually *am* second-personally obligated to Barry. However, I also reject this suggestion since I am unable to find any plausible second-personal account of my proposed obligation to Barry. I begin with the first possible line of defence against the Sally/Barry counterexample, below.

2. The first line of defence: loosening the accountability-obligation connection

As a first line of defence against the Sally/Barry counterexample, one may argue that I am wrong to assume that Barry's authority to hold me accountable would, on Darwall's account, entail that I owed an obligation to *Barry* – or, indeed, that any one person's authority to hold me accountable would entail that I was obligated to *that particular person*. This would be to construe the accountability-obligation too tightly, as I will put it. Darwall isn't committed to the claim that any one person's authority to hold me accountable would entail that I was obligated to *that* particular person; all Darwall needs, this defence goes, in order to establish his Second-Personal account is that an agent's liability to being held accountable entails that that agent was obligated to *somebody*, be it to that individual with the authority to hold the agent accountable or to some other individual. Call this suggestion, that an agent's liability to being held accountable by a particular individual entails only that the agent was obligated to *some* (possibly a different) individual, the 'looser accountability-obligation connection'. If this looser account is correct, then my pointing out, as I did above,¹¹ that Barry has the authority to hold me accountable for breaking my promise to Sally even though I owe no such obligation to Barry is no strike against Darwall. All Darwall needs in order to establish his Second-Personal account is that I was obligated to *somebody* – in this case, Sally (but not Barry). This is the first line of defence against the Sally/Barry counterexample.

We have already encountered a similar defence of Darwall. In the preceding chapter, I argued that Darwall's allowance that an agent may hold himself accountable "by the reproaches of his own

¹¹ See s.1, above.

conscience”¹² had the implication that some moral obligations (those for which the agent has the authority to hold himself accountable) are not second-personal, as Darwall claims, but *first-personal*.¹³ But, I went on to point out that this “reproaches of his own conscience” objection depends on the assumption that the agent is obligated to the individual by whom he/she is liable to be held accountable; that is, because the agent is liable to being held accountable by him-/herself, I took it to follow that the agent is therefore obligated to him-/herself, first-personally. Once this is made explicit, the objection may follow that the assumption of such a tight accountability-obligation connection is wrong.¹⁴ More loosely, just because an agent may hold him-/herself accountable, it doesn’t necessarily mean that he/she was obligated *to him-/herself*, first-personally – he/she may have been obligated to another, second-personally, even though he/she (is the only individual who) has the authority to hold him-/herself accountable. If this looser accountability-obligation connection is correct, then my “reproaches of his own conscience” objection of the preceding chapter fails, as does this present moral community objection.

So, is this looser understanding of the accountability-obligation connection correct? When I first presented the moral community objection,¹⁵ I mentioned that our practice of criminal law is a key example of members of the general moral community holding agents accountable. For instance, a prison warden has the authority to lock a convicted robber up, even though the prison warden himself was not robbed (by this particular robber). A proponent of the looser accountability-obligation connection, however, may claim that I have understated this legal case. This case doesn’t demonstrate that the prison warden *also* has the authority to hold the robber accountable – as if the victim of robbery herself has such authority, and the prison warden is one more individual who also has such authority. What has been overlooked is that, in this case, the victim *doesn’t* have the authority to hold the robber accountable; if the victim herself were to take the robber away and lock him up, the victim would be guilty of abduction – she would be doing something *wrong*, something she doesn’t have the authority to do. But, where the victim herself does not have the authority to hold the robber accountable, the prison warden does. So, the argument goes, what this legal case (indeed, probably *all* legal cases) demonstrates is that members of the moral community – in this case, the prison warden – have the authority to hold wrongdoers accountable *even when the actual victim of the wrongdoing has no such authority*. And, this is support for the looser accountability-obligation connection: an agent (the robber) may be obligated to one individual (the victim of robbery) even though that individual does *not* have the authority to hold the agent accountable but

¹² Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited in the introduction to my Ch.3.

¹³ See Ch.3, s.1.

¹⁴ See Ch.3, s.1.4.

¹⁵ See Ch.3, s.2.

another individual, a member of the moral community (the prison warden), *does* have the authority to hold the agent accountable. Call this ‘the prison warden argument’, in support of the looser reading of Mill’s Point.

The prison warden argument is offered in support of the proposed looser accountability-obligation connection, which will deflect my Sally/Barry counterexample¹⁶ and “reproaches of his own conscience”¹⁷ objections. According to this looser accountability-obligation connection, (sometimes) one individual (the prison warden) has the authority to hold a wrongdoer (a robber) accountable even though no obligation was owed to that individual (the prison warden) but was owed to a third individual (the victim of robbery) who has no authority to hold the wrongdoer accountable. According to this first line of defence, since an obligation was owed *to* an individual (the victim) – even if this individual has no authority to hold the wrongdoer accountable – this is sufficient for Darwall to establish second-personality.

In defence of the Sally/Barry counterexample, as well as my “reproaches of his own conscience” objection, I have a number of responses to this first line of defence: firstly, I will argue that at least one of the premises of the prison warden argument is false and the victim actually *does* have the (moral) authority to hold the robber accountable; secondly, I will argue that, even if the conclusion of the prison warden argument can be established, it is irrelevant to, and therefore does not deflect, my Sally/Barry counterexample; and, thirdly, I will argue that, even if the prison warden argument is sufficient to establish the proposed looser accountability-obligation connection, this looser accountability-obligation connection is insufficient to establish second-personality, and thus unacceptable to Darwall.

2.1. Halving the looseness

According to the prison warden argument,¹⁸ an individual (the prison warden) may possess the authority to hold an agent (the robber) accountable even when that agent was not obligated to that particular individual; moreover, the individual may possess this authority to hold an agent accountable even when the actual victim of the agent’s misdeeds does *not* possess such authority.

But, firstly, is it true, as the prison warden argument claims, that the victim *doesn’t* have the authority to hold the robber accountable? I argue that this premise of the prison warden argument is false. While I am willing to concede that the victim doesn’t have the *legal* authority to lock the

¹⁶ See s.1, above.

¹⁷ See Ch.3, s.**Error! Reference source not found.**, especially s.**Error! Reference source not found.**.

¹⁸ See s.2, above.

robber up, recall that there are various means of holding accountable. As Darwall, quoting Mill, was quoted in Chapter 3:

“we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.” We may think there are strong moral reasons for people to do something and “dislike or despise them for not doing” it; but unless we think “blame” or some other form of “punishment” is warranted, perhaps just the feeling of guilt, we do not think it “a case of moral obligation” [...].¹⁹

The means of holding accountable, therefore, include: legal punishment (“by law”), social disapproval (“the opinion of his fellow creatures”, “dislike or despise”, “blame”) and guilt (“the reproaches of his own conscience”) – and we are given no directive to treat this as an exhaustive list. Thus, even though it may be beyond the authority of the victim to hold the robber accountable “by law”, she may well nevertheless be able to hold him accountable by her opinion: her disliking, despising and blaming him. To use Strawson’s term, the victim may nevertheless be able to hold the robber accountable by taking up the reactive attitude of resentment towards him.²⁰

Since it strikes me as uncontroversial that a victim of robbery would indeed have the authority to resent the robber who violated his obligation to her not to steal, it follows that the robber is both (*ex hypothesi*) obligated to the victim and (as I have just argued) liable to being held accountable by her. That is, accountability and obligation come together as far as the victim is concerned. So, the accountability-obligation connection is at least only half as loose as the first line of defence²¹ suggested it was: it is not the case that the robber owes an obligation to the victim not to steal from her while the victim has no authority to hold the robber accountable (while the prison warden does); the victim *does* have this authority – perhaps not to hold the robber accountable *legally*, but at least to hold him accountable by other means such as resenting him, which Darwall allows for. So, the proposed looser accountability-obligation connection turns out to be only half as loose as suggested, since accountability and obligation come together as far as the victim is concerned.

Do accountability and obligation also come together as far as the prison warden is concerned too? If so, as I will now go on to argue, the proposed looser accountability-obligation connection isn’t merely half as loose as proposed, but not loose at all – and, indeed, as tight as I originally took it to be: an agent’s obligation to an individual entails his/her liability to being held accountability by that particular individual, and, *vice versa*, an individual’s authority to hold an agent accountable entails that that agent was (is) obligated to that particular individual.

¹⁹ Darwall 2006, p.27, citing Mill 1863/1998, Ch.5, as cited in the introduction to Ch.3.

²⁰ Strawson 1962/1968, as referenced by Darwall 2006, p.3.

²¹ See s.2, above.

2.2. Removing the looseness

I have argued against Darwall that, if he allows that members of the moral community may hold agents accountable, then his Mill's-Point inference from liability to being held accountable to the second-personality of obligation is not warranted, since there are some members of the moral community (for instance, Barry) who do have the authority to hold agents accountable but to whom those agents are not obligated.²² The first line of defence presently under consideration is that the accountability-obligation connection is actually looser than my moral community objection supposes, such that the dual facts of Barry's having the authority to hold me accountable for breaking my promise to Sally and my not being obligated to Barry to keep my promise (since I have made no promises to Barry) do not present a problem for Darwall's argument, and the prison warden argument was presented in support of this looser accountability-obligation connection.²³ I argue in this section that the prison warden argument does not establish the conclusion it needs to (even if its premises are true²⁴).

In order to deal with the Sally/Barry counterexample, this first line of defence needs to show that, on the proposed looser accountability-obligation connection, there are some cases in which an agent is liable to being held accountable by an individual even when no obligation is owed by the agent to *that* individual. (Such that, on the looser accountability-obligation connection, my being liable to being held accountable by Barry even though I owe no obligation to Barry is not inconsistent with Darwall's argument.) However, this is *not* what the prison warden argument establishes. The prison warden argument (if it works²⁵) establishes that an agent (the robber) may have an obligation to an individual (the victim) even though that individual (the victim) does not have the authority to hold the agent (the robber) accountable (but some member of the moral community – the prison warden – does have the authority to do so). But, this is the *opposite* of what the argument needs to establish if it is to be a successful defence against the Sally/Barry counterexample.

To be successful, this first line of defence needs to establish that (a) an agent (me) may be liable to being held accountable by an individual (Barry) even when no obligation is owed by that agent to that individual, but what the prison warden argument actually establishes (if it works) is that (b) an agent (the robber) may owe an obligation to an individual (the victim of robbery) even when that

²² See s.1, above.

²³ See s.2, above.

²⁴ In s.2.1, above, I argued that at least one premise of the prison warden argument is false: the victim *does* have the (moral) authority to hold the wrongdoer accountable by resenting him (even if the victim does not have the legal authority to personally lock the wrongdoer up).

²⁵ Of course, given the argument of the preceding section (2.1), I *don't* think that the prison warden argument works; I think its premise that the victim of robbery does not have the (moral) authority to hold the robber accountable is false.

individual does not have the authority to hold the agent accountable. Since the prison warden argument establishes (aims to establish) the *opposite* of what it needs to establish in order to accommodate my Sally/Barry counterexample, its conclusion is irrelevant.

In response to this charge of irrelevance, it may be argued that I have missed the point of the prison warden argument. In claiming, as I did in the preceding paragraph, that the prison warden argument establishes that (b) an agent may owe an obligation to an individual even when that individual does not have the authority to hold the agent accountable, I have focused on the agent's (the robber's) relationship to the *victim* – when the significance of the prison warden case actually lies in the agent's relationship to the *prison warden*. And, the argument goes, when we direct our attention to the appropriate aspect of the prison warden argument, we see that it does indeed establish that (a) an agent may be liable to being held accountable by an individual even when no obligation is owed to that individual: the *prison warden* has the authority to hold the robber accountable, even though the robber owes no obligation to the prison warden. Thus, when we focus on the appropriate aspect of the prison warden argument, it establishes precisely the conclusion it needs to establish: that (a) an agent (the robber) may be liable to being held accountable by an individual (the prison warden) even when no obligation is owed to that individual (the prison warden).

But, why think that the robber owes no obligation to the prison warden? It is true, *ex hypothesi*, that the prison warden was not himself robbed (by this particular robber), but it does not follow on that basis the robber does not owe the prison warden the obligation not to steal from him. It is true, *ex hypothesi*, that the robber has not *violated* his obligation to the prison warden, but it would be a mistake to deduce on this basis that no such obligation is owed. Indeed, we have good reason for thinking that the robber *is* obligated to the prison warden not to steal. The obligation not to steal strikes me as one of those obligations owed to all individuals. The obligation not to steal is owed to the beneficiary not in virtue of any special feature of the beneficiary, such as his/her usefulness to mankind, his/her closeness to the obligated or his/her social status; it is surely one of those obligations that is just owed to all individuals. (Contrast this with the obligation to keep promises, which, I have presumed,²⁶ is owed only to those to whom the agent has actually made promises.²⁷) It seems to me that the robber actually *is* obligated to the prison warden not to steal from him.

²⁶ See Ch.3, s.1.

²⁷ In section 3.2, below, I examine a possible objection to this presumption.

Thus, once again we see accountability and obligation coming together: the agent is (*ex hypothesi*) liable to being held accountable by an individual and (I have argued) is obligated to that *same* individual. This goes both for the victim of robbery,²⁸ and for the prison warden.

Of course, the correlation that I have demonstrated between one's authority to hold accountable and having an obligation owed to one – even in both the case of the victim of robbery and that of the prison warden – is insufficient to positively establish that there is indeed a tight connection between accountability and obligation. However, it is sufficient to defeat the prison warden argument in objection to it. Neither focusing on the robber's relationship to the victim nor on the robber's relationship to the prison warden establishes the conclusion the prison warden argument needs to: that (a) an agent may be liable to being held accountable by an individual even when no obligation is owed to *that* individual. So, the Sally/Barry case stands as an objection to Darwall's Mill's-Point argument: if, as in Barry's case, an individual may hold an agent accountable without the agent having any obligation to *that* individual, then, in the case of Sally, it does not follow on the basis of Sally's having the authority to hold me accountable that I had an obligation to Sally, and thus it doesn't follow that I had any *second-personal* obligation at all. Mill's Point will not entail second-personality.

Furthermore, I go on to argue in the following section that, even if the prison warden argument – or some other, similar argument, yet to be considered – were able to establish the requisite looseness of the accountability-obligation connection, this looser accountability-obligation connection would be of no help to Darwall, since the looser connection is insufficient to support Darwall's Mill's-Point argument.

2.3. Tightening the connection

Not only does the prison warden case leave us without a reason to take the proposed looser accountability-obligation connection seriously, we have positive reasons for believing that this is *not* the accountability-obligation relationship envisaged by Darwall, as I will now argue. Recall that the first line of defence against the Sally/Barry counterexample, currently under consideration, is to propose a looser accountability-obligation connection, according to which an individual may have the authority to hold an agent accountable even though no obligation was owed to *that particular* individual but to some further individual, and that the owing by the agent of the obligation to this further individual is sufficient to conclude second-personality.²⁹ This looser accountability-obligation

²⁸ See s.2.1, above.

²⁹ See s.2, above.

connection, however, is insufficient to do the work Darwall needs Mill's Point to do, and, as such, it is not an acceptable defence against the objections I have raised in this chapter.

Recall that Darwall Mill's-Point argument works (is supposed to work) like this: Darwall begins with Mill's Point – that is, that there is a conceptual connection between accountability and obligation – and then infers that, because we are held accountable *by* others, we are obligated *to* others; that is, moral obligation is second-personal.³⁰ Construing this accountability-obligation connection 'tightly', as I have done, it is the inference that: because we are held accountable by others, we are obligated to *those same* others. (Hence, if I can show, as I have tried to do with the Sally/Barry case of promising, that we may be liable to being held accountable by some others – like Barry – even though we are not obligated to those same others, then I have presented a counterexample.³¹) In opposition, the first line of defence suggests that the inference ought rather to be read 'loosely', as: because we are held accountable by others, we are obligated to *some (any)* other, who may or may not be the individual(s) by whom we are liable to being held accountable (which would be consistent with my Sally/Barry case and in which case I have presented no counterexample).³² The prison warden argument was presented in defence of this looser accountability-obligation connection, but I have demonstrated that that argument is not successful.³³ With the prison warden argument rejected, not only are we without good reason to accept that looser accountability-obligation connection, I will now go on to show that Darwall in fact cannot accept this looser reading.

There are two reasons why Darwall cannot accept the looser accountability-obligation connection, which claims that an agent may be liable to being held accountable by an individual (such as Barry) without actually being obligated to that individual. The first reason is that such a looser accountability-obligation connection would neuter Darwall's Mill's-Point argument; the second is that the looser accountability-obligation connection is actually at odds with Darwall's Second-Personal account, and as such cannot be employed in its defence.

Firstly, Darwall cannot accept the looser accountability-obligation because – rather than defending his Mill's-Point argument against the Sally/Barry counterexample – it would actually undermine it. According to this looser reading, my liability to being held accountable by Barry doesn't (necessarily) entail my obligation to *Barry*, but to some other. However, if this is the case, then I don't see much reason for thinking that my obligation is *to anybody* at all, let alone to Barry. And, if the looser-accountability-obligation connection leaves us without good reason for believing that our obligations

³⁰ See the introduction to my Ch.3.

³¹ See s.1, above.

³² See s.2, above.

³³ See ss.2.1 and 2.2, above.

are owed to others (whomever those others may be), then the Mill's-Point argument fails to establish the existence of *second-personal* reasons. Indeed, if the looser reading is correct, then we are left not only without a reason to believe that we are second-personally obligated to Barry (as is the contention's claim), but we also left without a reason for believing that we are second-personally obligated to *Sally*. If the accountability-obligation connection is so loose that I may be liable to being held accountable by an individual without being obligated to that individual, then my liability to being held accountable by Sally needn't entail that I am obligated to her either, contrary to the aims of the Mill's-Point argument.

Indeed, the looser accountability-obligation connection is so loose that it allows second-personality to slip through its fingers, leaving the door wide open to Realism. The Realist account of Mill's Point would be that: because I am liable to being held accountable by you, I am obligated, *full-stop* – that is, where my obligation is just a state of the world, and not a second-personal matter of the relationship I have to another individual. If my liability to being held accountable by you doesn't (necessarily) entail that I am obligated to *you* (as the tight reading would have it), then I see no reason for thinking that it entails my obligation to anybody (as the looser reading would have it) rather than my obligation, *full-stop* (as the Realist reading would have it). In the following chapter, I go on to advocate precisely this Realist interpretation of Mill's Point. For now, however, I simply make the point that, if the tight reading of Mill's Point is substituted for the looser one, then the intuitive force of Darwall's Mill's Point argument is forfeited. On the other hand, if the tight reading of Mill's Point is retained, Darwall's Mill's Point argument is susceptible to my Sally/Barry counterexample.

The second reason that Darwall cannot accept the looser accountability-obligation connection is not only that it undermines the force of his Mill's-Point argument, but that it is positively in contradiction to his own Second-Personal account. That Second-Personal account, recall, is that moral reasons are not a matter of the way the world is, but are, rather a matter of the kinds of relationships we stand in to other people.³⁴ And, the pertinent kind of relationship – the relationship that generates second-personal moral reasons – is that in which the second person has the authority to hold the other accountable. This is the great novelty of Darwall's Second-Personal account: what it is to be subject to a moral reason *is* just to be subject to a second person's authority; that second person doesn't come to possess authority over the first because the first is already under a moral obligation to the second – rather, on Darwall's ground-breaking Second-Personal account, the second person *gives* the first a moral reason, by way of his/her second-

³⁴ See my Ch.1, s.1.2.

personal authority over the first. The upshot of this is that it is impossible, on Darwall's Second-Personal account, for an individual like Barry to possess second-personal authority over an agent and yet for that agent to not be obligated to that individual; this is because *what it is* to be under a moral obligation on Darwall's Second-Personal account is just to be under a second person's authority and thereby to be liable to being held accountable by that second person. As such, Darwall cannot consistently accept the looser accountability-obligation connection, according to which Barry may have the second-personal authority to hold me accountable and yet I am not obligated to Barry, as this is in contradiction to Darwall's Second-Personal account of moral reasons itself.

Thus, the first line of defence against the Sally/Barry counterexample fails. The counterexample cannot be explained away by positing a looser connection between accountability and obligation, because, firstly, we have little reason for thinking that there is such a proposed looser connection,³⁵ secondly, even if there is such a proposed looser connection, it doesn't provide the necessary support for Darwall's Mill's-Point argument (as above), and thirdly, it would actually contradict Darwall's own Second-Personal account of moral reasons. As such, my moral community objection stands: if Darwall allows that it is (sometimes) not only the victim of a wrongdoing (Sally) who has the authority to hold the wrongdoer accountable but also some third-party member(s) of the moral community (Barry), yet the wrongdoer has no obligation to *that* member of the moral community (Barry), then Darwall is not warranted in inferring on the basis of the fact that Sally has the authority to hold the wrongdoer accountable that the wrongdoer is obligated *to* Sally. That is, Darwall's allowing that a wrongdoer may be liable to being held accountable by a member of the moral community to whom he/she is not obligated means that Mill's Point does not entail second-personality.

3. The second line of defence: rejecting the counterexample

There is a second possible line of defence against my moral community objection, however. Whereas the first line of defence tried to accommodate the Sally/Barry case by loosening the connection between accountability and obligation,³⁶ this second line of defence retains the tight accountability-obligation connection and simply rejects the Sally/Barry case as a counterexample to begin with. I have argued, in the preceding section,³⁷ that the tight reading of Mill's Point presents the better case for second-personality. Recall that the tight reading is that the accountability-obligation connection is such that, if I am liable to being held accountable by a particular individual,

³⁵ See ss.2.1 and 2.2, above.

³⁶ See s.2, above.

³⁷ See s.2.3, above.

then it follows that I am obligated to *that* particular individual. What makes the Sally/Barry case a counterexample (if it is one) to Darwall's Mill's-Point argument is that Barry has the authority to tell me off for breaking my promise to Sally even though I have no such obligation to Barry (because I made no promises to Barry). Thus, the Sally/Barry case may be rejected as a counterexample if we accept that Barry has the authority to tell me off for breaking my promise to Sally, while rejecting that I have no obligation to Barry – that is, insisting that I actually *am* obligated to Barry to keep my promises.

This is essentially the same tactic I employed, above,³⁸ in arguing that, even though the prison warden himself was not robbed (by our particular robber), the robber is both liable to being held accountable by the prison warden and is obligated to the prison warden not to steal from him, since the obligation to steal is owed to all individuals. In that previous discussion, I asserted that, where the obligation to not steal is owed to all individuals, the obligation to keep promises is not and is owed only to those to whom the agent has made promises. This second line of defence, then, is essentially to argue that, just as the obligation to not steal is owed to all individuals, the obligation to keep promises is also owed to all individuals. And, if the obligation to keep promises is owed to all individuals, then it is owed to Sally as well as to those members of the moral community, like Barry. So, the argument goes, my Sally/Barry case does not present a counterexample to Darwall's Mill's-Point argument: if it's true that I am both liable to being held accountable by Barry and that I have an obligation to Barry to keep my promises, then this casts no doubt over Darwall's inference on the basis of the fact that I am liable to being held accountable by an individual that I am second-personally obligated to that individual. If the obligation to keep promises is owed to all individuals, then my Sally/Barry case is no counterexample against Darwall's Mill's-Point argument.

3.1. Relevant descriptions

But, how could it be that the obligation to keep promises is owed to all individuals? How could it be that I have an obligation to keep my promises to Barry when I have made no promises to Barry? There *is* no promise to Barry, so how could I be obligated to keep it?

The key to seeing that it *is* possible that I may be so obligated lies in how the obligation is described. As G.E.M. Anscombe criticized Mill: "Mill also, like Kant, fails to realize the necessity for stipulation as to relevant descriptions"³⁹ – a stipulation which is necessary, given that any action might be variously described. For instance, as Donald Davidson has taught us, the very same action might just

³⁸ See s.2.2, above.

³⁹ Anscombe 1958, p.2.

as well be described as “alert[ing] a prowler to the fact that I am home” as it is “flip[ping] the switch, turn[ing] on the light, and illuminat[ing] the room.”⁴⁰

This is significant because here our natural way of describing the obligation at issue may be misleading – and misleading in my favour, I admit. The phrase “my obligation to keep my promise to Barry” is ambiguous: the “to Barry” could be understood as describing the *promise* (i.e. a promise to Barry) or it could be understood as describing the *obligation* (i.e. an obligation to Barry, about promising). I suggest, however, that this phrase is much more likely to be read in the former sense than in the latter sense: when one speaks of the “obligation to keep my promise to Barry”, one is referring to the obligation to keep that particular promise which was the one made to Barry, and not the one made to Susan, for instance. And, if the phrase is understood in this former sense – that is, as an obligation about promises-to-Barry – then, since there *are no* promises-to-Barry, it seems to follow obviously that I have no relevant obligations. (And, since I have no obligation to keep my promises to Barry – yet Barry (*ex hypothesi*) nevertheless has the authority to hold me accountable for keeping promises, this apparently presents a counterexample to Darwall’s Mill’s-Point argument.)

But, the above reading is not the only one available. The phrase “my obligation to keep my promise to Barry” might also be read in the latter sense; that is, as describing not (only) an obligation about promises-to-Barry but a second-personal obligation-*to*-Barry about promising.⁴¹ And, when the phrase is understood correctly to pick out my second-personal obligation-*to*-Barry about promising, then it is less obvious that this is an obligation I don’t have. Granted, I have made no promises *to Barry*, but the obligation “about promising” (if there is one) need not be an obligation to keep my-promises-to-Barry but an obligation to keep promises generally, such that one ought always to keep any promise (regardless of to whom it was made). And, having such an obligation to keep my promises generally is consistent with my not actually having made any promises to Barry: even though I have made no promises to Barry, I may nevertheless be obligated to Barry (and, indeed, to all individuals) to keep my promises generally. Thus, the fact that (*ex hypothesi*) I have made no promise to Barry doesn’t necessarily entail that I have no obligation-*to*-Barry to keep promises, as I

⁴⁰ Davidson 1963, pp.685-700.

⁴¹ Hence, elsewhere in this discussion I have tried to use the phrase “my obligation to Barry to keep my promise”, which is awkward but hopefully less prone to ambiguity. Even despite this measure of care, however, I suspect that the phrase is still misunderstood as describing a promise-to-Barry rather than an obligation-*to*-Barry.

I suggest, also, that the tendency to read the expression as “my obligation to keep my promises-to-Barry”, rather than as “my obligation-*to*-Barry to keep promises”, is the same tendency that causes us to mistake first-personal reasons *about* second persons as genuinely second-personal reasons. (See Korsgaard’s complaint of this nature against the Neo-Kantian Analogical argument in Ch.1, s.5.3, and see my complaint of this nature against the Darwallian version of the Analogical argument in Ch.2, s.1.2, and against Contractarianism in Ch.2, s.2.3.3.2.1, and in Ch.3, s.1.3.)

may nevertheless be obligated to Barry in a general way, as I am obligated to all individuals, to keep promises when I make them.

If the obligation to keep promises does indeed have this general form, then the Sally/Barry case may be rejected as a countercase on the grounds that, as in the prison warden case,⁴² the agent is actually both liable to being held accountable by Barry (and by Sally) and is second-personally obligated to Barry (and to Sally) to keep promises generally.

3.2. Is there a general obligation to keep promises?

It is correct that the possibility of such a general obligation to keep promises shows that my not having made any promises-to-Barry does not necessarily entail that I have no obligation-to-Barry concerning promising. But, the mere *possibility* of such a general obligation is not sufficient to defend Darwall from the Sally/Barry countercase. In order to defend Darwall from this objection, it needs actually to be the case that there *is* a general obligation to keep promises that is owed second-personally to all individuals. Recall that Darwall's Mill's-Point argument is that: because we are liable to being held accountable by others, we are obligated, second-personally, to those others; that is, Darwall infers second-personality on the basis of liability to accountability. My objection has been that Darwall is not justified in making this inference if there are members of the moral community (such as Barry) – which Darwall seems to allow – who have the authority to hold an agent accountable, even when no obligation is owed by that agent to *those* members. Thus, in order to effectively block this objection, it must be shown that not only is it *possible* that the general obligation to keep promises is owed second-personally to all individuals – but that it *is* plausibly owed to all individuals. The mere possibility is insufficient to vindicate Darwall's inference to the best explanation.

Below, I consider two possible accounts of the general obligation to keep promises. I will argue, however, that none of them is acceptable to Darwall. Thereafter, I will go on to consider an account of pluralized addressed, which will not yield a general obligation to keep promises as such, but which might plausibly do the same work as the response presently under consideration. I will argue, however, that this account of pluralized address is also unacceptable.

3.2.1. A Realist account of the general obligation to keep promises

How could it be that an agent has a general obligation to keep promises, even to those to whom he/she has made no such promises? It strikes me that Realism provides the most natural explanation of such a general obligation: if it is just part of the fabric of the universe that one ought to keep promises, then it follows that I ought to keep my promises even when I have made no

⁴² See s.2.2, above.

promises to Barry. It follows in just the same way, for instance, that it follows on the Realist account that the foot-treader ought not cause you undue harm even when you haven't actually spoken the command, "Get off my foot!" to the him.

However, a Realist account of the general obligation to keep promises clearly cannot be employed in support of Darwall's Second-Personal account; that would be to break into the Second-Personal circle from outside it.⁴³ I raise the Realist suggestion here, however, because it is one I will pursue in depth in the following chapter.

The pertinent question, of course, is not whether there is *an* account of promising available according to which one is generally obligated to keep promises even when one has made no promises; rather, the pertinent question is whether there is available a *second-personal* account of promising according to which I may be obligated to Barry to keep my promises, even when I have made no promises to Barry. If such a second-personal account of the general obligation to keep promises is available, then the second line of defence will have successfully rejected my Sally/Barry counterexample, by establishing that not only does Barry have the authority to hold me accountable for breaking my promise to Sally, but I am also under a general obligation to Barry (as I am to all other members of the moral community) to keep any promises I make.

3.2.2. A Contractualist account of the general obligation to keep promises

Scanlon, whom Darwall generally follows closely, devotes an entire chapter of *What We Owe to Each Other* to developing a Contractualist account of promising.⁴⁴ And, unlike Realism, Contractualism is compatible with the second-personal approach. As such, Contractualism would seem to present a fruitful avenue in the search for a second-personal account of the proposed general obligation to keep promises.

According to Scanlon's account, we ought to keep our promises because we could not reasonably reject *Principle F*, which is that:

If (1) A voluntarily and intentionally leads B to expect that A will do X (unless B consents to A's not doing so); (2) A knows that B wants to be assured of this; (3) A acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) B knows that A has the beliefs and intentions just described; (5) A intends for B to know this,

⁴³ Darwall 2006, p.12, as cited in my Ch.2, s.2.3.2.

Likewise, there are available several other non-second-personal accounts of the general obligation to keep promises; Kant (1785, Ch.1), for instance, derives the wrongness of false promising from the Categorical Imperative. But, of course, these other non-second-personal accounts will also be objectionable on the grounds of attempting to break into the Second-Personal circle from outside it.

⁴⁴ Scanlon 1998, Ch.7, pp.295-327.

and knows that B does know it; and (6) B knows that A has this knowledge and intent; then, in the absence of special justification, A must do X unless B consents to X's not being done.⁴⁵

Note how this Contractualist account is compatible with the second-personal approach: it captures the idea of reciprocal recognition, with its requirements not only that A recognize B (in intentionally leading A to form an expectation) but also that B recognize A as doing so, as “[having] this knowledge and intent”. If Contractualism – unlike Realism – is compatible with second-personality, and if Contractualism has a ready account of promising, such as Scanlon’s *Principle F*, then it does indeed appear plausible that we might have a general obligation to keep promises.

However, there are several problems with this Contractualist line of response. Firstly, note that Scanlon’s *Principle F* (above) does not offer an account of the proposed *general* obligation to keep promises – that is, it does not offer an account of the proposed obligation to individuals such as Barry to keep promises *even when no promises have been made to Barry*. To use the terms of *Principle F*, what is currently under investigation is the possibility that A, having made a promise to B, might nevertheless be obligated *to C* to keep his/her promise (to B). Yet *Principle F* makes no mention of such an obligation to C. If we accept that *Principle F* could not reasonably be rejected, then all that has been established is that, if A makes a promise to B, then A is obligated *to B* to keep his/her promise – but that is not the issue under contention. That Scanlon’s *Principle F*, therefore, could not reasonably be rejected does nothing to suggest that we are subject to what I have called the *general* obligation to keep promises, such that I might be obligated to Barry to keep my promises even though no promises have actually been made to Barry. Moreover, Scanlon is explicit that:

Unlike an obligation to comply with a just institution that provides some of the public goods, the obligation to keep a promise *is owed to a specific individual* who may or may not have contributed to the practice of promising. In addition, *the only expectations that are directly relevant are those created by the promiser and promisee at the time the promise is made.*⁴⁶

Even if Contractualism, unlike Realism, is compatible with a second-personal account, therefore, a hallmark Contractualist account of promising such as Scanlon’s in fact suggests that we are not under a general obligation to keep promises, such that I am obligated to Barry to keep my promises even though I have made no promises to Barry. In fact, it explicitly claims the opposite: that a promisor is obligated only to the individual to whom the promise was made.

We should note, furthermore, that this Scanlonian account of promising is one Darwall in fact *rejects*. In his review of Scanlon’s work, ‘Contractualism, root and branch’, published just prior to *The Second-Person Standpoint*, Darwall argues that: “Scanlon’s claim is that the wrongness of breaking promises is to be explained by the wrongness of violating *Principle F* [...], however, [...] I

⁴⁵ Scanlon 1998, p.304.

⁴⁶ Scanlon 1998, p.316 (emphases added).

doubt that this is so”,⁴⁷ adding that: “*Principle F* is not compelling”,⁴⁸ because “*Principle F* is impotent to explain the wrongness of breaking promises [...]”⁴⁹ Since Darwall himself rejects the Scanlonian account of promising, therefore, Scanlonian Contractualism is unlikely to supply the plausible account of our proposed general obligation to keep promises necessary to defend Darwall’s Second-Personal account from my Sally/Barry counterexample.

Of greater importance, however, is *why* Darwall rejects Scanlon’s account of promising – and this will remind us of a familiar point. To make his case, Darwall devises “a fairly elaborate case” in which I deliberately lead you to form the belief that I will attend your party by sending not you but some third person an email saying that I will attend your party, knowing all the while that you (for whatever reason) regularly check my email.⁵⁰ The content of this email satisfies clause 1 of *Principle F*, which is that “A voluntarily and intentionally leads B to expect that A will do X (unless B consents to A’s not doing so)”,⁵¹ and Darwall goes on to describe how subsequent emails between me and my other friend – all of which you read – satisfy clauses 2-6 of *Principle F*.⁵² However, in yet another email – which you also read – I go on to state that I have changed my mind and no longer plan on attending your party, and this information is made known to you, via the email, before you have relied on it for catering purposes, etc. So, Darwall concludes that:

Even though conditions (1) through (6) [of *Principle F*] are satisfied, then, it nonetheless seems that, if I have informed you otherwise and you have not relied on it, you have no remaining claim to my actually coming to your party. I conclude that there is a case where it would not be wrong to violate *Principle F* [...].⁵³

That is, Darwall presents a case in which all of Scanlon’s six *Principle F* clauses are met and yet it is not true that “A must do X unless B consents to X’s not being done”,⁵⁴ according to Darwall, I do nothing wrong if I do not attend your party, under these stipulated conditions.⁵⁵

⁴⁷ Darwall 2006a, pp.210-211, referencing Scanlon 2001, p.245. (This is the same *Principle F* that appears in Scanlon 1998, p.304, as cited above.)

⁴⁸ Darwall 2006a, p.211, referencing Scanlon 2001, p.245. (This is the same *Principle F* that appears in Scanlon 1998, p.304, as cited above.)

⁴⁹ Darwall 2006a, p.214, referencing Scanlon 2001, p.245.

⁵⁰ Darwall 2006a, p.212.

⁵¹ Scanlon 1998, p.304, as cited above.

⁵² Darwall 2006a, pp.212-213.

⁵³ Darwall 2006a, p.213.

⁵⁴ Scanlon 1998, p.304, as cited above.

⁵⁵ Let me note, for the record, that it is not obvious to me that, if I gave you my assurance (via a spied-upon email) that I would be attending your party, and later changed my mind, that “you [would] have no remaining claim to my actually coming to your party” (Darwall 2006a, p.213). Even if this was not an email addressed to you, but one sent with the *intention* that you would read it and form the relevant expectation on that basis (as it was, *ex hypothesi*), then my intuition is fairly strongly that I *would* be doing something wrong by later opting not to attend the party. That is to say, I am unconvinced by Darwall’s original counterexample to Scanlon’s *Principle F*. Nevertheless, what is presently of greater interest to me is Darwall’s diagnosis of the error of *Principle F*.

According to Darwall, this counterexample to *Principle F* obtains when the principle is read “causally”, and it is on this causal reading that “*Principle F* is not compelling”.⁵⁶ This causal reading is one according to which all that is required in order that A ‘assures’ B that A will do X is that A intentionally produces in B the relevant belief (and that B knows that A does so). And, this causal reading leaves room for the possibility – as Darwall has pointed out – that A could intentionally produce in B the relevant belief without A actually having *addressed B*, as might obtain in a situation where B is known by A to spy on A’s emails. And, naturally, as this emphasis on address would suggest, what Darwall claims is absent from the Scanlonian account of promising is the *second-personal*.⁵⁷ To quote Darwall:

Scanlon’s claim is that the wrongness of breaking promises is to be explained by the wrongness of violating *Principle F*. Unless, however, we interpret the idea of “providing assurance” in a (second-personal) way that already presupposes the authority of the assured to claim compliance, or assume, at least, that assurer and assured present themselves to one another as presupposing this in common, I doubt that this is so.⁵⁸

When we interpret the idea of providing assurance not causally, as a matter of producing relevant beliefs, but second-personally, then, claims Darwall, “[i]t is part of the very idea of a promise or an assurance that the addresser gives the addressee to understand that she thereby has a claim to the addresser’s following through”.⁵⁹ That is, since B’s authority to hold A to his/her promise is built into the very idea – second-personally construed – of A having made a promise to B, then Darwall’s email counterexample is blocked. According to this counterexample, “Even though conditions (1) through (6) are satisfied, then, it nonetheless seems that, if I have informed you otherwise and you have not relied on it, you have no remaining claim to my actually coming to your party”,⁶⁰ but, on a second-personal reading, it simply isn’t possible that I might have made you a promise that I will attend your party and that, when I back out, “you have no remaining claim to my actually coming to your party”. This is because, second-personally understood, “the idea of “providing assurance” [...] already

⁵⁶ Darwall 2006a, p.211, referencing Scanlon 2001, p.245.

⁵⁷ Darwall’s (2006a) remarks in ‘Contractualism, root and branch’ about Scanlon’s *Principle F* are clearly cast as a criticism of Scanlon: Scanlon presents a causal principle, when he should instead have presented a second-personal principle. However, in *The Second-Person Standpoint* (2006), Darwall appears to present Scanlon as something of a champion of the second-personal (even if not by name) – in which case Darwall is clearly reading Scanlon second-personally, rather than causally. It’s unclear to me why, in ‘Contractualism, root and branch’ (2006a), Darwall seems to think that the causal reading of *Principle F* is the one intended by Scanlon and is therefore grounds for criticism of Scanlon, when, in *The Second-Person Standpoint* (2006), Darwall clearly thinks that Scanlon is most naturally read not causally but second-personally. However, regardless of whether or not Darwall’s (2006a) ‘Contractualism, root and branch’ criticism of Scanlon is deserved, my present point depends on the fact – which is not up for doubt – that Darwall finds the second-personal reading of *Principle F* to be preferable to the causal reading.

⁵⁸ Darwall 2006a, pp.210-211, referencing Scanlon 2001, p.245.

⁵⁹ Darwall 2006a, p.214.

⁶⁰ Darwall 2006a, p.213, as cited above.

presupposes the authority of the assured to claim compliance, or assume, at least, that assurer and assured present themselves to one another as presupposing this in common [...].”

So, Darwall rejects Scanlon’s *Principle F*, because it allows for cases, such as Darwall’s email counterexample, in which it is not wrong to break a promise. On the other hand, if we read *Principle F* not causally but second-personally, as Darwall recommends, then such counterexamples are blocked. Perhaps, then, a Darwallian, second-personally-interpreted *Principle F* can provide an account of the general obligation to keep promises.

However, a Darwallian reading of *Principle F* will fare no better than the Scanlonian version in this present context. There are several reasons why this Contractualist account will not succeed as an account of the proposed general obligation to keep promises. Firstly, if Darwall’s only amendment to the Scanlonian account of promising is that we read it second-personally rather than causally, then it is true of the Darwallian account – as it was true of the Scanlonian account (above) – that this is not an account of what I have called the *general* obligation to keep promises. Recall that, in order to block my Sally/Barry counterexample, what is sought is an account of the proposed *general* obligation to keep promises, such that I might be obligated to Barry to keep my promises even though I have made no promises to Barry. But, as we noted above, *Principle F* supplies no such account, on neither the causal nor the second-personal readings thereof. *Principle F* covers only the relationship between the promisor (“A”) and the promisee (“B”); it makes no mention of any obligations potentially owed to third-person bystanders (“C”) such as Barry. As such, *Principle F* is essentially irrelevant to the present Sally/Barry counterexample, regardless of whether we read it causally or second-personally.

Secondly, note how Darwall explicates his suggestion that we read *Principle F* second-personally:

we [should] interpret the idea of “providing assurance” in a (second-personal) way *that already presupposes the authority of the assured to claim compliance*, or assume, at least, that assurer and assured present themselves to one another as presupposing this in common [...].⁶¹

However, this is precisely an authority, as I have been at pains to make clear, that we *cannot* presume. This complaint will by now be familiar. It emerges in the context of the larger project at hand: the aim of Darwall’s *Morality as Equal Accountability*, is to *establish* that we all share equal, second-personal authority.⁶² Darwall initially tries to do this by way of Pufendorf’s Point, according to which, if I hold you genuinely morally accountable, you too must hold yourself genuinely morally accountable. And, this purportedly entails that you and I have a shared, second-personal

⁶¹ Darwall 2006a, pp.210-211, as cited above (emphasis added).

⁶² See Ch.1, s.6.2.2.

authority.⁶³ However, I have argued that this is a *non-sequitur*: all that has been demonstrated is that you and I share an authority *over you* – it is yet to be seen that you share the requisite reciprocal authority *over me*.⁶⁴ As such, we are not in a position to simply presume, as Darwall’s second-personal reading of *Principle F* recommends, “the authority of the assured to claim compliance”. Who are *you* to “claim compliance” from *me* if I don’t feel like coming to your party anymore? Because Morality as Equal Accountability is yet to be vindicated, Darwall is unable to appeal to its claims in the defence of *Principle F* currently under consideration.

Darwall also endorses Fichte’s Point, which argues on the basis of our giving each other second-personal reasons that we can transcendently deduce our shared, equal, second-personal authority.⁶⁵ However, I have questioned whether we can accept the starting point of that supposed transcendental deduction: are there really such things as second-personal reason – reasons we literally *give* each other – or do we simply draw each other’s attention to reasons already in existence?⁶⁶ To this end, we have explored Darwall’s Mill’s-Point argument, which argues on the basis of our practice of holding each other accountable that we are therefore obligated *to* each other – that is, are second-personally obligated.⁶⁷ However, I have presented my Sally/Barry counterexample in service of my moral community objection to Darwall’s Mill’s-Point argument.⁶⁸ If a defender of Darwall is to block my Sally/Barry counterexample, therefore, he/she cannot do so by appealing to a point that the Mill’s-Point argument is (ultimately) meant to defend, namely, our shared, equal second-personal authority. That would be to beg the question. Unless we know that there are actually such things as second-personal reasons, the Fichtean transcendental deduction of our shared, equal moral authority can’t even get started – and it is the Mill’s-Point argument that is supposed to establish the second-personal nature of our moral obligations.

If we seek, as we do, an account of the proposed general obligation to keep promises in order to block my Sally/Barry counterexample to Darwall’s Mill’s-Point argument, then Darwall is not yet in a position to supply that account. Such a second-personal reading of *Principle F* would “[presuppose] the authority of the assured to claim compliance”⁶⁹ – yet that is precisely an authority we are currently in the business of trying to establish. This is why, as Darwall is perfectly explicit in his review of Scanlon, and as I have been at pains to make clear, that we cannot appeal to Contractualism in order to vindicate the Second-Personal account. Contractualism takes for granted

⁶³ See Ch.1, s.6.1.2.

⁶⁴ See Ch.2, s.1.

⁶⁵ See Ch.1, s.6.2.1.

⁶⁶ See Ch.2, s.2.

⁶⁷ See Ch.3.

⁶⁸ See s.1, above.

⁶⁹ Darwall 2006a, pp.210-211, as cited above.

that we share an equal moral authority; Morality as Equal Accountability (if the arguments for it are successful) can demonstrate the truth of that Contractualist tenet.⁷⁰ Rather than Morality as Equal Accountability relying on Contractualist accounts of promising (or of anything else), it is Contractualism that needs second-personality for its vindication. As Darwall concludes his review of Scanlon:

[...] I have been arguing that *Principle F* is impotent to explain the wrongness of breaking promises unless we interpret it in second-personal terms. I have been suggesting throughout this review essay that [C]ontractualism is itself best grounded in the equal second-personal authority that you and I must presuppose whenever we take up the second-person point of view. If this is right, equal second-personal authority is the root from which [C]ontractualism, and its impressively branching theory of rights, grows.⁷¹

I conclude, therefore, that a Contractualist account of promising does not give us grounds to believe that there is a general, second-personal obligation to keep promises, such that I may be obligated to Barry to keep my promises even though I have made no promises to Barry. Contractualism does not – indeed, cannot – support second-personality. On the other hand, second-personality may well turn out to support Contractualism, as Darwall suggests, above – but second-personality will first have to be established. If my Sally/Barry counterexample stands, then Darwall’s Mill’s-Point argument for second-personality fails. We have no good reason to believe that there are any such things as second-personal reasons in the first place.

In the following section, I turn my attention away from any substantive account – Realist or Contractualist – of promising, general or otherwise, and towards the nature of second-personal address instead. These more general considerations will offer further support for the idea that there is no general obligation to keep promises such that I might be second-personally obligated to Barry to keep my promises even though I have made no promises to Barry, which will in turn strengthen my Sally/Barry counterexample against Darwall’s Mill’s-Point argument.

3.3. Pluralized address

It might be suggested that a second-personal reason could be given in ‘plural form’. That is, even if there is no general second-personal obligation to keep promises (as explored above⁷²), it may nevertheless be the case that I am obligated to Barry to keep my promise to Sally, even though I have made no promises to Barry – if it is the case that my promise was made to Sally *plurally*. When I second-personally address Sally with the words, “I promise...”, I thereby (purportedly) come to owe a second-personal reason to Sally to keep the promise I have made to her – and, if this second-

⁷⁰ See Ch.2, s.2.3.3.2.2.

⁷¹ Darwall 2006a, p.214.

⁷² See ss.3.2.1 and 3.2.2, above.

personal address is pluralized, I will thereby (purportedly) come to owe the same second-personal reason to another second person, such as Barry, to keep the promise that I have made to Sally.

If such pluralized second-personal address obtains, therefore, I have presented no countercase to Darwall's Mill's-Point argument:⁷³ Sally has the authority to hold me accountable to the promise I made to her, and I am (purportedly) second-personally obligated to Sally; Barry also has the authority to hold me accountable to the promise I have made to Sally, *and* I am (purportedly) second-personally obligated to Barry to keep my promise to Sally by way of my pluralized second-personal address of Sally, even though I have made no promises to Barry directly. That is, in the cases of both Sally and Barry, I am liable to being held accountable by that second person as well as (purportedly) being second-personally obligated to that second person; hence, Darwall's Mill's-Point inference – that, on the basis of our liability to being held accountable by another, we are second-personally obligated to that other – goes through in both the cases of Sally and of Barry. If my second-personal address of Sally is pluralized, therefore, my Sally/Barry countercase gains no foothold on Darwall's Mill's-Point argument.

The idea of pluralized address does receive some attention in *The Second-Person Standpoint*, where Darwall makes a handful of scattered references to it.⁷⁴ For instance, in the very early pages of the book, he writes:

I claim that to understand moral obligation as related to moral responsibility in the way we normally do, we have to see it as involving demands that are “in force” from the moral point of view, that is, from the (first-person *plural*) perspective of the moral community.⁷⁵

Note how the first part of this remark alludes to Mill's Point (even if not by name): the idea that moral obligation is conceptually “related to moral responsibility”, such that since the latter is a second-personal affair, the former must be second-personal in nature too. As such, despite only a handful of references being made to pluralized second-personal address in the main text of *The Second-Person Standpoint*, the idea seems to be a promising one for getting to grips with the Mill's-Point argument.

But, we should ask, what *is* it for a second-personal reason to be addressed in ‘plural’?

Unfortunately, the notion of pluralized second-personal address receives no real further analysis than the bare mention of it, as above. Nevertheless, it does strike me as an intriguing notion and one certainly deserving of further exploration. But, this is not an exploration I will embark upon here. This is because, just on basis of the few remarks about pluralized address Darwall does

⁷³ See s.1, above.

⁷⁴ See Darwall 2006, pp.9-10, 102, 177, 182, 185 (fn.8), 198, 201, 276 (fn.39).

⁷⁵ Darwall 2006, p.9 (emphasis added).

proffer, there are already apparent several considerations that count against the idea, many of which will by now be familiar.

Firstly, almost all of Darwall's references to pluralization are not of the pluralized address of second-personal reasons but of *first-personal* ones.⁷⁶ (And, those references that do not specify the kind of pluralization as second-personal do not specify it all, which would suggest that the kind of pluralized address under discussion in the non-specified cases is, as in all the specified cases, first-personal rather than second-personal.) This is apparent even from the very first mention of the notion, as it was cited above:

I claim that to understand moral obligation as related to moral responsibility in the way we normally do, we have to see it as involving demands that are "in force" from the moral point of view, that is, from the (*first-person plural*) perspective of the moral community.⁷⁷

That is, instead of the above remark helping us to get to grips with the Mill's-Point argument that moral obligation's being "related to moral responsibility" is best understood in terms of both being second-personal, the above remark seems rather to recommend that they are (plurally) *first-personal* – quite the opposite of the Mill's-Point conclusion. I doubt, therefore, that Darwall's handful of references to pluralization – where this is first-personal pluralization – will be helpful in blocking my Sally/Barry counterexample.

In response, it might be argued that we should read the above reference to first-personality in light of Darwall's claim that "the second-person[al] stance is a version of the first-person[al] standpoint";⁷⁸ however, I have already presented my case against taking this claim literally.⁷⁹ Since, in Darwall's own words, "[o]ne can occupy a first-person[al] perspective [...] without explicitly addressing anyone",⁸⁰ if second-personality is to be modelled on the relationship of address,⁸¹ then "[w]hat the second-person[al] stance excludes is [...] first-personal thought that lacks an addressing, second-personal aspect."⁸² Darwall's handful of references to first-personal pluralization, therefore, do not shed light on what pluralized second-personal address might be, nor whether it actually obtains. If there is such a thing as pluralized second-personal address, Darwall does not discuss it in *The Second-Person Standpoint*.

Although Darwall does not provide an account of pluralized second-personal address, he does, however, provide a detailed and useful analysis of second-personal address *per se*. And, I think that

⁷⁶ See Darwall 2006, pp.9-10, 102, 177, 185 (fn.8), 198, 201, 276 (fn.39).

⁷⁷ Darwall 2006, p.9, as cited above (emphasis added).

⁷⁸ Darwall 2006, p.9, as cited in my Ch.3, s.1.3.

⁷⁹ See Ch.3, s.1.3.

⁸⁰ Darwall 2006, p.10.

⁸¹ Darwall 2006 throughout; see, for instance, p.3.

⁸² Darwall 2006, pp.9-10.

careful attention to this analysis will reveal the improbability of pluralized second-personal address (at least on the scale required to block my Sally/Barry counterexample). This will follow on the basis of the *reciprocal recognition* essential to second-personal address.

The idea of reciprocal recognition is explicitly captured in Fichte's Point,⁸³ but the notion is already present in Darwall's modelling the second-personal relationship on the relationship of address.⁸⁴ When one person addresses another, that person recognizes the other – but, if *address* is to have taken place, the first person must also have recognized the second reciprocally, that is, as addressing the first. If I pass you in the corridor and you call out, "Hey, handsome!" you will not have *addressed me* unless I recognize you-as-recognizing-me-as-handsome; if I don't have reciprocal recognition of you in this way, then, though I may hear your words, I will think you are not talking to me but to the good-looking guy behind me, and you will not have successfully addressed me.⁸⁵ The nature of second-personal address, therefore, necessitates reciprocal recognition by the first person of the second-as-recognizing the first. This leads me to believe that second-personal address is not easily pluralized, as I go on to explain.

Having a conversation is, of course, one key example of the reciprocal recognition involved in second-personal address. So, if the second-personal relationship is to be modelled on the relationship of address, then the amount of people with whom an individual may plausibly have a conversation (i.e. address second-personally) should be a good guide as to how many people an agent may plurally stand in the second-personal relationship of moral obligation with. We should be able to get at least a rough idea of how many people to whom we might plausibly be plurally second-personally obligated on the basis of intuitive considerations about how many people with whom it is possible to have a conversation. I will argue that this reveals that the scope of pluralized second-personal address falls far short of the entire general moral community, such that it is improbable that my (purportedly) plurally second-personally addressing a promise to Sally entails that I am also obligated to Barry.

With how many people is it possible to have a conversation? One can of course have a conversation with another, single person, recognizing that that other recognizes one-as-recognizing-that-other. One can even have a conversation with two other people, recognizing each of those two others as recognizing one-as-recognizing-each-of-those-two-others. One can possibly even have a conversation with twenty other people, recognizing each of those twenty others as recognizing one-

⁸³ See Ch.2, s.6.2.2.

⁸⁴ Darwall 2006 throughout; see, for instance, p.3.

⁸⁵ Thanks to Jeremy Wanderer for sharing this illustration with me. I believe it comes from Kukla & Lance (2009).

as-recognizing-each-of-those-twenty-others. But, even this may be a stretch. Bear in mind that the claim is not that twenty people might have conversations amongst themselves, but that one person might have a conversation with a full twenty others on the other end of the line, so to speak. It's not entirely obvious to me that one speaker might have a conversation with *twenty* listeners.

The situation starts to look even more improbable when we consider that moral reasons are, according to Darwall, second-personal, and that moral reasons are universal.⁸⁶ Consider the moral reason to not cause me undue harm, for instance. If there are seven billion people on the planet, and the moral reason to not cause me undue harm is universal, then, on the present proposal, this means that I have second-personally addressed each one of those 6 999 999 999 in demanding that I not be unduly harmed. Bearing in mind that second-personal address necessarily involves reciprocal recognition, this means that for all those 6 999 999 999 people possessing a moral reason not to cause me undue harm, I must have recognized each one of them as recognizing me-as-recognizing-each-one-of-6 999 999 999-people. Is this possible? Is it possible to have a conversation with seven billion people? Is it even possible to have a conversation with seven hundred people? I struggle to see how it could be. And, note that this is only exacerbated by the fact, if it is a fact, that future persons might owe us obligations (to, for instance, respect our memory, not wantonly demolish our monuments, not turn our burial sites into parking lots, etc.). If future persons owe us obligations, and if that obligation is second-personal, as Darwall claims, then the implication is that we second-personally address a *massive* – possibly infinite! – amount of people, given how many people are surely to follow us in the future. I struggle even more to see how one might recognize each of an infinite amount of people as recognizing one-as-recognizing-each-of-an-infinite-amount-of-people. It is surely beyond our capacities to have a conversation with an infinite amount of people.

Thus, I conclude that, while it may be possible to plurally second-personally address a limited amount of people – perhaps as many as twenty – it is not possible to plurally second-personally address very large numbers of people, such as the other 6 999 999 999 people on earth or the possibly infinite amount of people to follow me in the future. Yet, if such people owe me obligations, such as not to cause me undue harm or to respect my memory, and if those obligations are second-personal, as Darwall claims moral obligations are, then Darwall's account requires that such largescale pluralized second-personal address *is* possible.

What of the obligation to keep promises, though? Could it not be addressed second-personally and plurally, such that it is owed to both Sally and Barry, even though I make a promise only to Barry and not to Sally? Sally and Barry amount to just two people. And, since I have admitted the possibility of

⁸⁶ See Ch.1, ss.1.2 and 1.3.

second-personally addressing two people (above), it may appear that I have conceded the possibility of plurally addressing, and thereby coming to owe an obligation to, both Sally and Barry.

However, these appearances are misleading. “Barry” is just a useful label for: any member of the general moral community. The relevance of considering whether I might be obligated to “Barry” to keep my promise to Sally is that, according to Mill’s Point, I am liable to being held accountable for breaking my promise to Sally by *any* member of the general moral community – where this includes Barry, as well as Susan, and Lindiwe, and Marco, and Solomon, etc., etc. Since, according to Mill’s Point, I am liable to being held accountable for breaking my promise to Sally not only by Barry, but also by Susan, and by Lindiwe, and by Marco, and by Solomon, etc., in order to block my Sally/“Barry” counterexample, it will have to be demonstrated that I am obligated not only to Barry but also to Susan, and to Lindiwe, and to Marco, and to Solomon, and indeed to any member of the general moral community by whom I am liable to be held accountable. And, if I am liable to being held accountable by any member of the general moral community, as I am according to Mill’s Point, then this will amount to roughly 6 999 999 999 individuals. In short, my concentration throughout this discussion on just the characters Sally and “Barry” makes it no more plausible that the obligation to keep promises is plurally second-personally addressed than that the universal obligation not to cause undue harm (as above) is.

However, there is a much more critical reason why the possibility of pluralized second-personal address (if, indeed, it is possible at all) will not be successful in blocking my Sally/Barry counterexample. The suggestion depends on a very subtle – but nevertheless fatal – confusion. We have been considering the suggestion that, in making a promise to Sally, I second-personally address Sally plurally, such that I (purportedly) come to have a second-personal obligation not only to Sally, but also to a further second person, Barry, even though I have made no promises to Barry. That is, the considerations of this section have taken me to be the addressor (of a plurality of second persons) as well as taking me to be the obligated party. But, notice that this deviates from the relationship between second-personal address and the giving of second-personal reasons as originally presented. According to this original presentation of the relationship between second-personal address and the giving of second-personal reasons, one party does the addressing, resulting in a second party’s being obligated. On the account originally presented, second-personal reasons are given in address; that is, second-personal reasons are given by a second person *to a first* when the *second* addresses the first (be it plurally or otherwise). (Recall Darwall’s foot-treader scenario, according to which, when *you* tell the foot-treader to get off your foot – that is, when *you*, the second person, address the first person, the foot-treader – *you* thereby give the *foot-treader* a reason to remove his foot from on top

of yours.⁸⁷) That is, on the account originally presented, a second person does the addressing, resulting in the first person's being subject to a (second-personal) reason; whereas, the possible response to my Sally/Barry counterexample currently under consideration has been one according to which it is the *same* individual who both does the addressing and is subject to obligation: one individual (me) who both (plurally) addresses a promise to Sally and is subject to the obligation (purportedly to both Sally and Barry) to keep her promise to Sally.

In other words, the considerations of this present section 3.3 confuse the making of a promise with the giving of a reason. For a promisor to make a promise to the promisee is surely for the promisor to address the promise, however this is *not* the (purportedly) *reason-giving* address pertinent to the present discussion. Instead, the relevant address is the reciprocal address *of the promisor by the promisee* when the promisee makes a claim against the obligated promisor, saying, for instance, "But you have to come to the party – you promised!" Claiming, therefore, as the suggestion of this present section has done, that *my promise* to Sally might be plurally addressed to both Sally and Barry (and indeed any other second person) is simply irrelevant to the present discussion, since, on the original account, it is not my making a promise (be it plurally or otherwise) that gives me a second-personal reason in the first place; rather it is *Sally's* (and possibly *Barry's*) demanding (having the authority to demand) that I keep my promise that gives me a second-personal reason to keep it. (Likewise, it is *your* demanding that the foot-treader get off your foot that gives him a second-personal reason to do so.) Therefore, even if it is the case that, in addressing a promise to Sally, I do so plurally, thereby also addressing Barry (as well as any other second person), it simply doesn't follow that I am thereby obligated to Sally as well as to Barry to keep my promises, since, in addressing the promise, I was never (on Darwall's account) *obligating myself* to either Sally or Barry in the first place. Rather, on Darwall's Second-Personal account, it is *Sally* who obligates me when she addresses me second-personally by demanding that I keep my promise to her.⁸⁸

Thus, even though the suggestion that my promise might have been addressed plurally, not only to Sally but also to Barry, is misconceived as a response to my Sally/Barry counterexample, I have nevertheless accorded the idea of pluralized address due consideration. Ultimately, I judge it highly improbable that pluralized second-personal address might take place between a single individual

⁸⁷ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁸⁸ In response, it may be argued that, likewise, *Barry* obligates me second-personally when Barry demands that I keep my promise to Sally. Therefore, the argument may go, it *does* follow that Barry both has the authority to hold me accountable and I am obligated to him, and thus my Sally/Barry counterexample is blocked. However, this is to take for granted precisely the issue under contention in this project. Why assume that *Barry* has the authority to demand that you keep your promise to Sally? Barry is a third-person observer, a stranger on the street – who is *he* to make any demands of you at all, especially when it comes to the private matter of your promise to Sally?

and a very large number of people, given the reciprocal recognition that is an essential aspect of second-personal address. The pluralization of second-personal address does not support the idea that I owe an obligation to Barry to keep my promise to Sally, even though I have made no promises to Barry. It seems, then, that we must accept that the obligation to keep a promise is owed only to those to whom promises have actually been made. Here, Wallace, for one, agrees with me:

The fact that promisees have a compelling objection to principles that permit you to flout expectations you have deliberately led them to form about your own [behaviour] is connected to the relational idea that *promissory obligations are owed specifically to the promisee*; it is the recipient of the promise who would be wronged when it is broken and who is in a privileged position to complain when let down in this way.⁸⁹

In conclusion, I argue that the initial intuition was correct and I owe no obligation-to-Barry to keep my promise, since I have made no promises to Barry. Since, however, (*ex hypothesi*) Barry has the authority to hold me accountable for breaking my promise to Sally, the Sally/Barry case stands as an objection to Darwall's Mill's-Point argument. If Darwall's Mill's-Point argument infers from the fact that I am liable to being held accountable by Sally to the conclusion that I am therefore second-personally obligated to Sally, then the same logic must apply in the case of Barry: since I am also liable to being held accountable by a member of the general moral community, such as Barry, for breaking my promise to Sally, then it must follow that I am also second-personally obligated to Barry – yet, I have made no promises to Barry, so I am surely not obligated to him, nor is there available a plausible account of a general obligation to keep promises (such that I may be obligated to Barry even though I have made no promises to him),⁹⁰ nor is it plausible that I am obligated to Barry by way of my 'plurally' second-personally addressing my promise to Sally (above). Since it is the case that I am not obligated to Barry, yet I am nevertheless liable to being held accountable by Barry for breaking my promise to Sally, it follows that Darwall's Mill's-Point argument fails: the fact that I am liable to being held accountable by Sally cannot entail that I am therefore second-personally obligated to Sally.

I expand on this argument in section 3.4, below, where I consider the nature of the third person's (such as Barry's) holding the first accountable. If we have reason to doubt that a third person holds a first accountable *second-personally*, then we have similar reason to doubt that the obligation entailed by this authority to hold accountable is itself second-personal.

3.4. Impersonal reactive attitudes

In this chapter, I have posited the Sally/Barry counterexample as an objection to Darwall's Mill's-Point argument, and I have examined some attempts to block that counterexample by insisting that I actually

⁸⁹ Wallace 2007, p.34 (emphasis added).

⁹⁰ See s.3.2, above.

am obligated to Barry even though I have made no promises to him. Yet, I have found all of those attempts to be unsuccessful. Now, I wish to mount the beginnings of a different kind of objection to Mill's-Point argument – one which will be sustained across the following chapter.

The Sally/Barry countercase challenged the logic of Darwall's Mill's-Point inference by pointing out that, if my liability to being held accountable by Sally entails that I am second-personally obligated to Sally, then my liability to being held accountable by Barry should also entail that I am second-personally obligated to Barry – yet I am not so obligated to Barry. This new objection challenges not the logic of Darwall's Mill's-Point inference *per se*, but its premise. The Mill's-Point argument infers of the basis of *the second-personality of holding an agent accountable* to the second-personality of the obligation to which the agent was held accountable; this new objection begins to question whether holding an agent accountable is even a second-personal matter in the first place.

Here P.F. Strawson appears to share my intuition. As we have noted,⁹¹ one way in which we can hold agents accountable is by evincing “reactive attitudes”, like resentment or indignation, towards them. This range of reactive attitudes Strawson sorts into the “personal” and the “impersonal”.⁹² Personal reactive attitudes are experienced by the “transagents”⁹³ themselves – that is, by those actually involved in the interaction: in particular, victims. Impersonal reactive attitudes, on the other hand, are those experienced by third-personal observers of the interaction – or, as I have put it, members of the general moral community. To apply this to the working example of this chapter: whereas Sally will experience a *personal* reactive attitude to my breaking my promise to her, according to Strawson, Barry will experience only an *impersonal* reactive attitude. Both hold me accountable, but Sally does so personally, whereas Barry does so only impersonally. And, as Strawson writes: “Because of this impersonal or vicarious character, we give them different names.”⁹⁴ That is, Strawson notes, as I do, an important difference between a victim's holding an agent morally responsible and any other member of the moral community's doing so, and this prompts Strawson to draw a distinction between these two different kinds of reactive attitudes. Given this distinction, even if personal reactive attitudes are second-personal (as Darwall holds), it does not follow straightforwardly that impersonal reactive attitudes will be second-personal too.

⁹¹ See s.2.1, above.

⁹² Strawson 1962/1968, pp.72-74 and 84-85, as referenced by Darwall 2006, pp.66-67.

⁹³ Darwall 2006, p.66.

⁹⁴ Strawson 1962/2003, p.205.

Indeed, if Barry's holding me accountable is not a personal reactive attitude but an *impersonal* one, then it would appear, *a fortiori*, not to be *second-personal* either.⁹⁵

Darwall, however, insists that a third-person observer's impersonal reactive attitude *would* be second-personal. Note how his position compels him to do so. Since he wishes to argue that we can infer the second-personal nature of moral reasons on the basis of our practice of holding each other accountable (i.e. the Mill's Point argument),⁹⁶ Darwall must construe the accountability-obligation connection tightly, as I have argued; that is, he is committed to claiming that those to whom we are obligated are those by whom we are liable to being held accountable, and vice versa.⁹⁷ And, if morality is universal,⁹⁸ then we must all be liable to being held accountable by each other – or by the “moral community”, as Darwall puts it.⁹⁹ But, since Darwall's major claim has been that “moral obligations and demands are quite generally second-personal”,¹⁰⁰ if moral obligations are those to which we are liable to being held accountable by the moral community, then it must be the case that a member of the moral community's holding one accountable is a *second-personal* matter; if it weren't, then it wouldn't follow (by a Mill's-Point inference) that moral reasons are universal *and* second-personal. So, Darwall's position compels him to claim that impersonal reactive attitudes – that is, those reactive attitudes held by members of the moral community – are second-personal too.

So, what independent motivation does Darwall offer for his claim that Strawsonian impersonal reactive attitudes are second-personal? I find in Darwall's work two occasions of such motivation. In the first instance, having defined Strawson's distinction between the personal and impersonal reactive attitudes, Darwall goes on to write in *The Second-Person Standpoint*: “I claim that reactive attitudes are always implicitly second-personal”¹⁰¹ (where, given the context of this remark, Darwall is apparently referring to both personal and impersonal reactive attitudes). Unfortunately, however, no thorough explanation follows as to why impersonal reactive attitudes are (purportedly) second-personal. What does follow is a discussion of indignation (which is an impersonal reactive attitude¹⁰²). However, Darwall's objective in that discussion is not to defend his presumption that impersonal reactive attitudes, like indignation, are second-personal; instead it is to again make the

⁹⁵ Of course, one should not, however, read too much into a mere label. Strawson also describes impersonal reactive attitudes as “sympathetic or vicarious [...] or disinterested or generalized” (1962/2003, p.205), and some of these terms are much less suggestive than “impersonal”.

⁹⁶ See the introduction to my Ch.3.

⁹⁷ See s.2.3, above.

⁹⁸ Darwall 2010c, p.40. See also my Ch.1, s.1.3.

⁹⁹ Darwall 2006, p.27.

¹⁰⁰ Darwall 2006, p.8.

¹⁰¹ Darwall 2006, p.67.

¹⁰² Strawson 1962/2002.

point (“Strawson’s Point”) that desirability is not a reason of “the right *sort*”¹⁰³ for holding agent accountable. As Darwall writes with reference to indignation:

If we come to believe that someone does not deserve blame, say, because he could not possibly have known the true character of what he was doing or because he was under extreme duress, then this will reduce or even defeat our indignation toward him. But if we learn that attempting to hold him accountable would be undesirable, say, because it will provoke him further, this will hardly undermine indignation.¹⁰⁴

Unfortunately this point does nothing to explain why “believ[ing] that someone [does or] does not deserve blame” is a *second-personal* matter. Indeed, not only is an explanation lacking, the passage above seems to suggest quite the opposite of the second-personality of indignation. If what’s at issue is a *belief* about an agent’s culpability, as the above excerpt suggests, then this seems good reason to conclude that impersonal reactive attitudes, like indignation, are *not* second-personal. Much of the force of Darwall’s argument in *The Second-Person Standpoint* depends on the distinction he draws between second-personal authority and epistemic authority.¹⁰⁵ Whereas epistemic authority is the authority merely to counsel on the basis of one’s knowledge of the facts, second-personal authority is the authority to *command*.¹⁰⁶ Recall, for instance, Darwall’s foot-treader scenario.¹⁰⁷ This scenario is meant to bring into relief the idea that, if you literally give the foot-treader a (second-personal) reason to remove his foot from on top of yours, you thereby exercise your *second-personal authority* to make such demands; on the other hand, if you merely draw the foot-treader’s attention to the fact of your pain, you exercise only your *epistemic authority*. As Darwall originally described this non-second-personal response to the foot-treader:

Structurally, the situation would be entirely analogous to a purely *epistemic* case, for example, one in which you give him reasons to believe that you are in fact in pain. Were he to credit the way things seem from the perspective of his desire, he would accept a state-of-the-world-regarding and agent-neutral reason for removing his foot. [...] [I]n “giving” him the reason in this way, you wouldn’t so much be addressing it to him as getting him to see that it is there anyway, independently of your getting him to see it or even of your ability to do so. There are two points here. First, in pointing to the reason, you would be directing him epistemically rather than practically [...]. *Qua* this form of reason-giving, you would be asking him to agree, as it were, that there is a reason for him to do something rather than asking him to agree to do it. Any claims you might make would thus be on his beliefs about practical reasons and not directly on his will. Second, your being able to give him the reason would not depend in any way on his seeing you as trying to give it to him or as having any [second-personal] competence or authority to do so.¹⁰⁸

And, as we have established, this kind of world-regarding, epistemic case is *not* what Darwall has in mind as the second-personal form of moral reasons: drawing an agent’s attention to the (moral)

¹⁰³ Darwall 2006, pp.15-16, citing Strawson 1962/1968, p.72.

¹⁰⁴ Darwall 2006, p.68.

¹⁰⁵ This is a distinction which McMyler (2011) has queried.

¹⁰⁶ See my Ch.2, ss.1.4 and 2.3.2.

¹⁰⁷ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

¹⁰⁸ Darwall 2006, pp.5-7 (emphases altered).

facts is merely to exercise one's epistemic authority, rather than to "give a person a distinctive kind of (normative) reason for acting"¹⁰⁹ by an exercise of one's second-personal authority. If, therefore, we should "reduce or even defeat our indignation" towards an individual because "we c[a]me to *believe* that [he/she] does not deserve blame", we would be proceeding not on our second-personal authority but on our *epistemic* authority, on our knowledge of the facts that "he could not possibly have known the true character of what he was doing or [that] he was under extreme duress".¹¹⁰ That is, if our being indignant or not towards an individual is a matter of our knowledge of the facts – our epistemic authority – then it is *not*, as Darwall nevertheless insists, a second-personal matter.¹¹¹ As such, Darwall's first motivation ultimately offers very little support for his claim that impersonal reactive attitudes are second-personal – indeed, it would appear to support quite the opposite conclusion.

Let us examine, therefore, the second occasion of Darwall's motivation for the claim that impersonal reactive attitudes are second-personal. This finds its home in a paper published subsequent to *The Second-Person Standpoint*. In that paper, Darwall again makes the claim that impersonal reactive attitudes are second-personal, writing quite explicitly: "Nonetheless, however 'impersonal', blame is not an 'objective' attitude in Strawson's sense. It is just as 'interpersonal' or 'second-personal' as personal reactive attitudes like resentment or guilt".¹¹² This time Darwall aims to substantiate his claim by footnoting the following excerpt from Strawson:

The same abnormal light which shows the agent to us as one in respect of whom the personal attitudes, the personal demand, are to be suspended, shows him to us also as one in respect of whom the impersonal attitudes, the generalized demand, are to be suspended.¹¹³

However, note that this again does nothing to support Darwall's assertion that impersonal reactive attitudes are *second-personal*. All this quote establishes is that, when personal reactive attitudes ought to be suspended, impersonal reactive attitudes ought also to be suspended – but that is far from implying that both are second-personal. In the case Strawson has in mind, above, both personal and impersonal reactive attitudes ought to be suspended because the *participant stance* is altogether inappropriate; as the original text by Strawson continues: when an individual "is seen in this light, he is not seen [...] as a morally responsible agent".¹¹⁴ But, failing to engage with an individual second-personally is not (necessarily) the same as failing to see him/her as a moral agent; one can interact with an individual as a moral agent without thereby interacting with that individual

¹⁰⁹ Darwall 2006, p.4 (emphasis removed), as cited in my Ch.1, s.1.2.

¹¹⁰ Darwall 2006, p.68, as cited above (emphasis added).

¹¹¹ I have raised essentially the same worry in Ch.2, s.1.4.

¹¹² Darwall 2010c, p.38, referencing Strawson 1962/1968.

¹¹³ Strawson 1962/1968, p. 87, as cited by Darwall 2010c, p.38, fn.7.

¹¹⁴ Strawson 1962/2003, p. 207.

second-personally (as I explain, below). As such, the fact that the suspension of both personal and impersonal reactive attitudes come together when the participant stance is inappropriate offers no support for Darwall's insistence that impersonal reactive attitudes are second-personal.

Darwall's above argument appears to rely on the hasty assumption that any person-on-person action is necessarily *second-personal*,¹¹⁵ but that simply isn't true. Recall, as we noted in the preceding section,¹¹⁶ that second-personal interactions are modelled on relations of address: a second person recognizing a first, who is conscious of the second as recognizing him/her. In cases of second-personal address, the first's consciousness of the second as recognizing the first is essential. To use the same example, as above: If I pass you in the corridor and you call out, "Hey, handsome!" you will not have *addressed me* unless I recognize you-as-recognizing-me-as-handsome; if I don't have reciprocal recognition of you in this way, then I will think you are not talking to me but to the good-looking guy behind me, and you will not have successfully addressed me.¹¹⁷ The first person's reciprocal recognition of the second-as-recognizing-the-first is essential to second-personal interaction. But, not all person-on-person exchanges have this form; we often interact with others in ways that don't require the other's recognition of us-as-recognizing-them. Monika Dullstein captures this nicely with her distinction between talking *with* another personal and talking *about* another person,¹¹⁸ where the former is second-personal and the latter not. Both talking with a person and talking about a person are person-on-person actions, but only the former is also a second-personal action because only the former involves the first person's reciprocal recognition of the second as addressing the first, whereas talking about a person requires no recognition by the subject that he/she is the topic of conversation.

Darwall in fact presents a similar case of non-second-personal, person-on-person action in his examination of the notorious Milgram experiments. Very briefly, the Milgram experiments involve three people: one "authority figure", one "learner", and one "teacher". The learner is required to master some simple memory tasks, and, when the learner gets an answer wrong, the teacher is instructed by the authority figure to administer a shock to the learner. Unbeknownst to the teacher, he/she is him-/herself actually the subject of the experiment, the aim of which is to see how far a subject will go in administering shocks to a clearly pained (but acting) "learner" on the instruction of the authority figure. One set-up of the experiment has the learner in a separate but adjacent room,

¹¹⁵ See, for instance, Darwall's remark that an impersonal reactive attitude "is just as 'interpersonal' or 'second-personal' as personal reactive attitudes like resentment or guilt" (Darwall 2010c, p.38, referencing Strawson 1962/1968, as cited above).

¹¹⁶ See s.3.4, above.

¹¹⁷ Kukla & Lance (2009), as referenced by Jeremy Wanderer (personal communication), as referenced in s.3.3, above.

¹¹⁸ Dullstein 2012, p.239.

where his/her cries of pain can be heard through the wall, but from where he/she is unable to see or hear the subject/teacher. A second set-up instead seats the learner right in front of the subject/teacher, and even requires the subject/teacher to physically press the learner's hand against a metal plate in order for the shock to be delivered. Interestingly, when the learner is moved from the separate room to be seated in front of the subject/teacher, the subject's/teacher's tendency to continue administering shocks to the learner decreases dramatically.¹¹⁹

Fascinating as these experiments and their results may be, my interest here is instead in Darwall's interpretation of those results. As he writes:

The most significant change in the overall rate of disobedience came when the learner was moved into the same room with the subject[/teacher]. The subject[/teacher] could then see the effects of what he was doing. But also importantly, *he was aware, for the first time, of the learner's awareness of him*. Even if audible protests were experienced as claims in the first two setups, the conditions made reciprocal recognition impossible.¹²⁰

That is, Darwall allows that the case in which the subject shocks the learner (thereby "recognizing" him) but where the learner's separation by a solid wall has "made reciprocal recognition impossible" is *not* a case of second-personal interaction, even though it is clearly a case of one person acting on another. On the other hand, when the learner was seated in front of the subject and the subject "was aware, for the first time, of the learner's awareness of him", the two were purportedly second-personally engaged. It is clear, by Darwall's own admission, therefore, that not all person-on-person action is second-personal. So it follows that we would be hasty in assuming that impersonal reactive attitudes are necessarily second-personal: we may hold a reactive attitude towards an agent in a way that does not require that agent to reciprocally recognize us-as-holding-that-reactive attitude.

In fact, I think Darwall's treatment of the Milgram experiments may support an even stronger conclusion than just the *hastiness* of the assumption that impersonal reactive attitudes are second-personal. We must attend to the events of the Milgram experiments: regardless of where the learner is seated, every time the learner gets an answer wrong, the subject/teacher is required to administer a punishment – a shock – to the learner. That is, the Milgram experiments involve not just any kind of person-on-person action, but those in which the subject/teacher apparently *holds the learner responsible* for his/her incorrect answers – and this goes for even that version of the experiment in which the learner is seated in a separate room. And, if Darwall is correct that the learner's physical separation "made reciprocal recognition impossible",¹²¹ then the actions of the subject/teacher on the learner in these versions of the experiment are not second-personal (as

¹¹⁹ Milgram 1974, as referenced by Darwall 2006, pp.164-5.

¹²⁰ Darwall 2006, p.166 (emphasis added).

¹²¹ Darwall 2006, p.166, as cited above.

above). But, note that they are nevertheless cases of the subject/teacher *holding the learner responsible* for his/her incorrect answers – that is, they are cases of the subject/teacher *non-second-personally holding the learner responsible*. Darwall’s treatment of the Milgram experiments, therefore, not only concedes that there are non-second-personal, person-on-person actions, but apparently also concedes that there are cases of non-second-personally holding others accountable.¹²² With this concession on the table, I see very little reason to accept, as Darwall insists, that impersonal reactive attitudes are second-personal.¹²³

According to Darwall, following Mill, an agent is liable to being held accountable not only by his/her victim but by any member of the general moral community. And, according to Darwall, we can infer on the basis of Mill’s Point – that we are liable to being held accountable by others – that we are therefore obligated *to* those others, second-personally. There is nothing logically inconsistent about the suggestion that any agent is second-personally obligated to *all* members of the general moral community. But, this suggestion does appear to undermine the strong and plausible sense in which I have a second-personal obligation *to Sally* to keep the promise that I made to her, and to which *Sally* therefore has the authority to hold me accountable. This appears to be importantly different to the scenario in which Barry holds me accountable for breaking the promise that I made to Sally, since I have neither made nor broken any promises to Barry. Here, Strawson appears to share my intuition, leading him to distinguish between the personal and impersonal reactive attitudes. Although Darwall asserts that impersonal reactive attitudes, like Barry’s, will (also) be second-personal, I have argued in this section that Darwall has not made a good case for that claim. Rather, Darwall’s suggestion that our evincing the impersonal reactive attitude of indignation is founded on our knowledge of the fact that the agent is deserving of blame implies that this impersonal reactive

¹²² I introduced this argument with some reservation, claiming that “Darwall’s treatment of the Milgram experiments *may* support an even stronger conclusion” (as above; emphasis added). This is because I think it may be argued in response that the subject’s/teacher’s shocking the learner is not a case of *genuine moral accountability* at all, but rather only a case of moulding behaviour by negative reinforcement (see my Ch.1, s.6.1.2); that is, the kind of response we offer to badly behaved dogs or very young children. And, it is true that shock therapy, as a form of aversion therapy, is intended to function by negative reinforcement. However, the learners in Milgram’s experiments are neither dogs nor very young children, but mentally unimpaired adult human beings – that is, the kind of being we normally do hold genuinely morally accountable. As such, it’s unclear that the shocks administered by the subject/teacher are not modes of holding the learner genuinely morally responsible: it may well be that they are intended not only to shape the learner’s behaviour, but to get the learner to *understand* that the answer he/she gave was wrong (and therefore not to make the same mistake again in the future). It is, therefore, at least possible to read the Milgram experiments as cases of normal interactions between unimpaired, adult human beings, which involve holding each other genuinely morally accountable. If this is the correct reading of the Milgram experiments, then it is one according to which Darwall has conceded a case of *non-second-personally* holding another accountable.

¹²³ I hope to make an even better case for this in the following chapter, where I try to develop a non-second-personal account of holding accountable from a Realist perspective.

attitude is not second-personal at all, since second-personal commands issue from one's second-personal authority, not one's epistemic authority. Likewise, Darwall's observation that the teacher/subject in (a version of) Milgram's experiment does not stand in a second-personal relationship of reciprocal recognition with the learner – while the teacher/subject nevertheless apparently holds the learner responsible for his/her incorrect answers – is to concede that an individual may hold an agent responsible non-second-personally. Not only, therefore, does Darwall provide us with little reason to accept that impersonal reactive attitudes are, as he asserts, (also) second-personal, we in fact have reason to conclude that they may be *non-second-personal*.

If, therefore, Darwall's Mill's Point argument infers on the basis of our practice of holding each other accountable to the nature of moral reasons, and if we have little reason to accept that impersonal reactive attitudes are second-personal, then it follows that our mode of holding each other accountable isn't necessarily second-personal, and therefore we have similarly little reason to accept that we are *second-personally* obligated to all members of the general moral community. We surely *are* all each obligated to not, for instance, cause undue harm to any other – but, I have argued, the conclusion that this is a *second-personal* obligation is not warranted.

4. Conclusion

Mill's Point is that

“we do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.” [...] [U]nless we think “blame” or some other form of “punishment” is warranted, perhaps just the feeling of guilt, we do not think it “a case of moral obligation” [...]. When we are morally obligated, [...] members of the moral community have the authority to hold us responsible if we do.¹²⁴

And, on this basis, Darwall concludes: “Moral obligations are thus *to others* [...]”;¹²⁵ that is, moral obligations are thus second-personal. However, I have argued over the course of this and the preceding chapter that Darwall's inference does not go through. Darwall, following Mill, allows that an obligated agent is one liable to being held accountable, not only by the victim of his/her wrongdoing, but also by the agent him-/herself – “by the reproaches of his own conscience” – or by members of the general moral community. These broader aspects of Mill's Point, however, are not easily explained by a Second-Personal account.

In particular, my moral community objection has been that, if Darwall aims to infer that my obligation to Sally is second-personal on the basis of the fact that Sally has the authority to hold me

¹²⁴ Darwall 2006, p.27, citing Mill 1863/1998, Ch.5, as cited in the introduction to Ch.3.

¹²⁵ Darwall 2006, p.27, as cited in the introduction to Ch.3.

accountable, then Barry's stipulated authority to hold me accountable for breaking my promise to Sally must, by the same logic, entail that I am obligated to Barry – and yet I have not made any promises to Barry, so I am surely not obligated to him.¹²⁶ To this end, I examined several possible accounts of the proposed general obligation to keep promises, such that I might have been obligated to Barry even though I have made no promises to him,¹²⁷ as well as the proposal that my promise to Sally was addressed plurally, such as to include address to Barry,¹²⁸ but I found none of these responses to be acceptable. Since Barry may have the authority to hold me accountable even though I owe no obligation to him, the logic of Darwall's Mill's-Point inference fails: Sally's authority to hold me accountable will also not justifiably entail that I am second-personally obligated to Sally.

Lastly, I went on to propose the further objection to the logic of Darwall's Mill's-Point argument, which was that it is questionable whether our practice of holding each other accountable is second-personal in the first place.¹²⁹ If, therefore, Darwall wishes to infer from our practice of holding each other accountable to the nature of our moral obligations, it will not follow that our moral obligations are second-personal if it is not that case that our practices of holding each other accountable are always second-personal.

If Darwall's Mill's-Point argument fails, we are left without any good grounds on which to conclude that there are indeed such things as second-personal reasons. And, if there are no such things as second-personal reasons, then Darwall's Morality as Equal Accountability will not follow.¹³⁰ This criticism of Darwall's Mill's-Point argument is continued in the following chapter, where I go on to argue that a Realist account is likely able to succeed where Darwall's Second-Personal accounts falters.

¹²⁶ See s.1, above.

¹²⁷ See s3.2, above.

¹²⁸ See s.3.3, above.

¹²⁹ See s.3.4, above.

¹³⁰ See Ch.2, s.2.

5.

The Realist objection

Over the course of the preceding two chapters,¹ I have argued that Darwall's Mill's-Point argument struggles to account *second-personally* for those broader aspects of Mill's Point, namely, that an agent may be liable to being held accountable by him-/herself only, or by members of the general moral community. In this present chapter, I will go on to argue, furthermore, that *Realism* can account for Mill's Point, including all its broader aspects, and that it can do so at least as well as – if not better than – Darwall's own Second-Personal account. Insofar as Darwall's Mill's-Point argument is an inference to the best explanation² – and if Realism provides an explanation at least as good as that of Darwall's Second-Personal account – Darwall's Mill's-Point argument does not go through. Second-Personality is not the best explanation of Mill's Point if Realism is at least as good as it, if not better.

However, I do not in this chapter limit myself to the Mill's-Point argument. I aim to show more broadly that Realism has explanatory power at least equivalent to Darwall's Second-Personal account. Throughout this project, I have made several other gestures towards Realism. For instance, in Chapter 2, we considered the suggestion that a command is valid when it is supported by independent reasons³ – where these independent reasons are most likely *Real* reasons. That is, your commanding the foot-treader to get off your foot is valid because the foot-treader already has a Real, world-regarding reason to get off your foot, in order to relieve your pain. Realism, therefore, also enjoys the advantage over Darwall's Second-Personal account of providing a straightforward account of validity.

For all these reasons, I will argue, *contra* Darwall, that Realism is actually preferable to the Second-Personal account. I find this challenge particularly compelling, given that it is Realism specifically against which Darwall pits his Second-Personal account in the foot-treader scenario of the opening pages of *The Second-Person Standpoint*.⁴ I must note, however, that I would not want to endorse Realism *per se*. I think that Realism is ultimately mistaken (as I will try to argue in the following

¹ See Ch.3 and 4.

² Darwall 2006, p.76, fn.26, as cited in the introduction to my Ch.3. See also Darwall 2010c, p.32.

³ See Ch.2, s.2.3.1.

⁴ See Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

chapter⁵). But, insofar as Mill's Point in particular is concerned, even though it might not be the case that Realism is *the best* explanation of Mill's Point, all that is required for my extended critique of Darwall's Mill's-Point argument⁶ is that Second-Personality is *not the best* account of Mill's Point. If I am able to demonstrate, as I plan to do in this chapter, that Realism offers an account of Mill's Point at least as good as – if not better than (even if only minimally better than) – Darwall's Second-Personal account, then Darwall's Mill's-Point inference to the best explanation fails: Second-Personality is not the best explanation of Mill's Point if another account is at least as good as it.

Before I turn to that argument for Realism, I remind the reader that by a "Realist" I have in mind primarily the sort of theorist against whom Darwall originally pits his Second-Personal account in the foot-treader scenario presented in the opening pages of *The Second-Person Standpoint*;⁷ that is, I have in mind here primarily the most common variety of Realist, the Consequentialist.⁸

1. Arguments for Realism

In this project, I have criticized Darwall's arguments for his Second-Personal account primarily on three grounds. Firstly, I have contended that his argument for Morality as Equal Accountability does not go through: Pufendorf's Point does not establish that, when a second person holds a first accountable, those second and first persons share an equal, reciprocal moral authority over each other; rather, I have contended, all that is revealed is that the second and first persons share an authority over *the first* only, and this is insufficient to account for the universality of moral reasons.⁹ Then, I have argued that we have no good reason for believing that there are such things as second-personal reasons in the first place, and this broad criticism has taken two forms, which I list here as my second and third primary criticisms of Darwall's arguments for his Second-Personal account. That is, secondly, I have argued that we have no good reason to believe that there are such things as second-personal reasons unless they are *valid* reasons, yet Darwall offers us no plausible, non-question-begging, second-personal account of validity.¹⁰ And, thirdly, I have considered Darwall's Mill's-Point inference to the existence of second-personal reasons on the basis that they best explain our practice of holding each other accountable, and have argued that this inference is dubious

⁵ See Ch.6.

⁶ See Ch.3-present.

⁷ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁸ See my Ch.1, s.2.

⁹ See my Ch.2, s.1.

¹⁰ See my Ch.2, s.2.3.

insofar as Darwall admits that we may be held accountable not only by the victims of our transgressions, but also by ourselves as well as members of the general moral community.¹¹

In this chapter, I argue that Realism outperforms Darwall's Second-Personal account on all three scores. That is, I argue that Realism can account for Pufendorf's Point, and that it provides the most straightforward account of validity, and that it provides a better explanation of our liability to being held accountable not only by the victims of our transgressions but also by ourselves as well as by members of the general moral community. I begin with the Realist account of validity, below.

1.1. Realist validity

On the Realist account, morality is "state-of-the-world-regarding",¹² or, as I have put it for short: world-regarding.¹³ Of course, this doesn't mean that just natural states of the world, such as the disaster wrought by a hurricane, are of moral significance; it includes states of the world resulting from agents' actions – to use Darwall's own example, it includes the bad state of affairs that would result from an agent's treading on your foot and causing you pain.¹⁴ On both Darwall's Second-Personal and Realist accounts of the foot-treader scenario, you have the authority to object to the foot-treader's causing you pain,¹⁵ and your giving/"giving"¹⁶ him a reason to get off your foot would be valid – as, intuitively, it surely would be. However, these two metaethics give very different accounts of what that valid authority would consist in. I argue that the Realist account is preferable.

The Realist account of validity is disarmingly straightforward. It can be modelled on the Correspondence theory of truth. Your utterance, "There is a cat on the mat", is true when there is indeed a cat on the mat; that is, when the world is such that there is a cat and it is on a mat. Likewise, on the Realist account, your telling the foot-treader to get off your foot is valid when there is indeed a reason for him to get off your foot; that is, when the world is such that the foot-treader is standing on your foot and it is causing you pain. (On the other hand, to quote Fleischacker, "[i]f you were instead to issue the command 'Move your foot!' quite arbitrarily, when my foot was nowhere near yours and doing nothing that I could reasonably see as disturbing you,"¹⁷ your command would

¹¹ See my Ch.3-4.

¹² Darwall 2006, p.6, as cited in my Ch.1, ss.1 and 2.

¹³ See Ch.1, s.1.1.

¹⁴ Darwall 2006, pp.5-6, as cited in my Ch.1, ss.1 and 2.

¹⁵ Darwall 2006, pp.5-7, as cited in my Ch.1, ss.1 and 2.

¹⁶ Of course, what I mean by the latter term in the present context is not that you *give* – literally, *give* – the foot-treader a reason, as one does on the Second-Personal account; I mean just: pointing out independent reasons, as one does on the Realist account. Because of the way this term (as well as "address") has been appropriated by Darwall for the purposes of distinguishing his Second-Personal approach, it is difficult to find neutral terminology. Darwall (2006, p.6) himself uses "giving" in scare quotes in describing the Realist response to the foot-treader.

¹⁷ Fleischacker 2009, p.120, as cited in my Ch.2, s.1.4.

not have been valid, as there would have been no fact of the world in support of it.) On both the Second-Personal and the Realist responses to the foot-treader, Darwall allows that you may validly object to the foot-treader's standing on your foot; on the Realist account this is just a matter of the world being such that there actually is a Real reason for the foot-treader to remove his foot from on top of yours. As the world is the truth-maker of our utterances, so the world is the validator of our "given" reasons. Thus, Realism offers a very straightforward account of validity.

Indeed, at times it seems that even Darwall himself slips into this way of understanding validity on his own Second-Personal account. For instance, in an interview Darwall remarks:

To go back to the example I gave before: If you're under a moral obligation not to step on my feet, and you step on my feet, I hold you responsible for it and blame you for it. In blaming you for it, I'm demanding that you not do it. I'm implicitly saying, "There is reason for you not to do it [...]"¹⁸

Here, Darwall is ostensibly articulating his Second-Personal account of giving reasons by "demanding" – but his accompanying explanation that what such a demand consists in is "implicitly saying, "There is reason for you not to do it"" is apparently to make a *Realist* appeal: drawing the foot-treader's attention to a reason that is "[t]here", rather than second-personally *giving* him a reason to get off your foot. I think we can probably put this confounding comment down to the fact of its having been made in an interview, where it is very difficult to give articulate, accurate answers while having to 'think on one's feet'. Nevertheless, it exhibits the strong temptation to make sense of validity in Realist terms. When is a command valid? The most natural, tempting answer is: When there is a reason for it – but this is a *Realist* account.

But, this is a temptation that Darwall must resist, as a Realist account of validity it is inconsistent with the rest of his Second-Personal account. Indeed, it's the very point of Darwall's original presentation in *The Second-Personal Standpoint* of the foot-treader scenario that second-personal reasons are *given* by a second to a first person, whereas world-regarding reasons are just "[t]here", as he puts it above, or "there anyway", as he puts it in that original presentation.¹⁹ So, what a second person is doing on Darwall's Second-Personal account when he/she gives a first person a reason *can't* be "implicitly saying, "There is reason for you not to do it [...]"", or drawing the first person's attention to a reasons that is already "[t]here".²⁰ Supposing that this is what the second person does would be to abandon the Second-Personal account by trying to break into the Second-Personal circle from outside it²¹ – what Darwall has explicitly stipulated "there is no way to" do.²²

¹⁸ Darwall 2009, pp.136-137.

¹⁹ Darwall 2006, p.5, as cited in my Ch.1, s.1.2.

²⁰ Darwall 2009, pp.136-137, as cited above.

²¹ See Ch.2, s.2.3.2.

Realism provides an attractively straightforward account of validity – but it is one unavailable to Darwall’s Second-Personal account.

Furthermore, the Realist account also offers a very straightforward account of how a non-victim member of the moral community, Barry, may also have the authority to “stick up for you” by also demanding that the foot-treader remove his foot from on top of yours, even though Barry remains unharmed by the foot-treader’s stomping. Since, on the Realist account, a command is validated by Real states of the world – the world we all share – any individual may in principle have knowledge of them. If Barry is particularly observant, he may notice the foot-treader’s foot planted on top of yours – that is, he may notice the same Real state of the world that gives you the authority to demand that the foot-treader remove his foot from on top of yours – in which case, Barry would also have the authority to tell the foot-treader to get off your foot. Not only, therefore, does Realism provide a very straightforward account of validity; in its world-regardingness it also accounts for the fact that members of the general moral community share the same authority as the victim of wrongdoing: the authority of knowledge of the facts, where these are facts to which any individual may have access. I continue this line of argument in the following section, but first I address an imminent objection.

Darwall, of course, would reject the kind of “epistemic authority” upon which these Realist accounts depend, as we have seen already.²³ A doctor may know all the facts about heart disease, but when the doctor tells you not to eat so much junk food, he’s not *commanding* you to change your diet; he’s merely *counselling* you to do so: your visiting a fast food outlet that very evening doesn’t *wrong the doctor*. Likewise, Darwall would argue, when on the Realist account you (or Barry) tell the foot-treader to get off your foot *because* you have knowledge of the fact of his causing you pain, you aren’t really *giving* the foot-treader a reason to get off your foot; you’re simply drawing his attention to the facts, as your doctor draws your attention to your increased chance of heart disease.

But, this rejection of epistemic authority does nothing to undermine the Realist account. It simply begs the question of *why* we should reject epistemic authority. Darwall, of course, has to reject epistemic authority because it is incompatible with his Second-Personal account, according to which second persons *give* first persons reasons, rather than merely drawing first persons’ attention to them. But, of course, we haven’t yet accepted Darwall’s Second-Personal account. My sustained argument has been that we haven’t yet been given any grounds to accept that there *are* indeed such

²² Darwall 2006, p.12, as cited in my Ch.2, s.2.3.2.

²³ See Ch.2, ss.1.4 and 2.2.

things as second-personal reasons,²⁴ so it is no shortcoming of *Realism* that it fails to account for second-personal reasons. Since Darwall allows on both the Second-Personal and the Realist responses to the foot-treader that you have the authority to tell the foot-treader to get off your foot,²⁵ and since the Realist account is “able to explain [this] significant ethical phenomenon[on]”²⁶ by the epistemic authority to draw first persons’ attention to Real reasons, I see no (non-question-begging) reason why we ought to reject epistemic authority as Darwall does. Since, as I will continue to argue in the following section, Realism can do all the same explanatory work as Darwall’s Second-Personal account, we still have no reason to accept that there is indeed such a thing as commanding over and above the more ordinary process of counselling – the latter of which is easily accounted for by Realism epistemically.

Furthermore, if Darwall rejects epistemic authority, in what does *he* ground the authority to give valid commands? His answer is apparently ‘second-personal authority’, that is, the authority of the Second-Personal standpoint – but we have already seen the problem with that approach: we are without a non-question-begging account of the Second-Personal standpoint that respects Darwall’s stipulation that we not try to break into the Second-Personal circle from outside it.²⁷ On the other hand, Realism, I have argued, provides a glaringly straightforward answer to the question of validity. But, of course, this option is not available to Darwall, because as soon as we ground authority in knowledge of Real moral norms – that is, epistemic authority – we turn our backs on second-personality. The situation reminds me of Friedrich Heinrich Jacobi’s famous criticism of Kant that, without the presupposition of the thing-in-itself, one cannot enter into Kant’s philosophy – but, with the supposition of the thing-in-itself, one cannot remain within it. Likewise, without the Realist account of validity, one apparently cannot enter into Darwall’s Second-Personal account²⁸ – but, with the Realist account of validity, one cannot remain within it.

Realism provides a much more straightforward account of validity than does the Second-Personal account. And, I will argue furthermore – of most relevance to this chapter – that Realism can account for Mill’s Point, including all of its broader aspects, just as well as – if not better than – Darwall’s Second-Personal account.

²⁴ See Ch.2, s.2, and Ch.3-4.

²⁵ See Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

²⁶ Darwall 2006, p.76, fn.26, as cited in the introduction to my Ch.3.

²⁷ See my Ch.2, s.2.3.3.

²⁸ See my Ch.2, ss.2.2 and 2.3.

1.2. Realist accountability

Mill's Point, recall, is that there is a conceptual connection between obligation and accountability, such that an obligated agent is necessarily liable to being held accountable, be this by the victim of the agent's wrongdoing, by the agent him-/herself, or by members of the general moral community.²⁹ Whereas Darwall's Mill's-Point argument is that his Second-Personal account offers the best explanation of Mill's Point, I will argue that the Realist is able to explain Mill's Point just as well as – if not better than – Darwall's Second-Personal account is.

Considering that Mill is a Utilitarian, and Utilitarians are Realists,³⁰ it should hardly be surprising that Realists are able to account for *Mill's Point*. Darwall, however, overlooks this. Consider, for instance, his remarks on Mill's "internal sanction" (conscience):

When we consider that we have done something we think wrong, we don't just have a painful feeling. We have a painful "appearance" that we shouldn't have done what we did, that not doing it was "*in itself* obligatory." We have a putative awareness of having acted contrary to what members of the moral community (and we ourselves as such members) justifiably demand.³¹

Mill's description of an act that is "*in itself* obligatory" strikes me as a clear Realist interpretation of moral value: an act that simply is, in itself, required, as a matter of the way the world is. Notice, however, how in the final quoted sentence Darwall immediately gives this Millian idea a second-personal interpretation: a matter not of wrongness in itself, but of what persons are or aren't allowed to *demand*. This logical leap depends on the oversight of the Realist alternative, and I submit that many of Darwall's arguments depend on a similar oversight.³² My aim in this chapter is to give that Realist alternative due consideration.

So, how can the Realist explain the accountability-obligation connection captured by Mill's Point? The same way the Realist explains everything else: there *just is* a conceptual connection between obligation and accountability. The world is such that electrons are negatively charged, and the world is such that agents who are obligated are agents who are accountable for acting as obligated. And, how is the Realist able to account for those broader aspects of Mill's Point, that the Second-Personal account struggles to explain?³³ That is, how is the Realist able to account for the agent's liability to being held accountable not only by his/her victim, but also by his-/herself or by member of the general moral community? These broader aspects of Mill's Point will follow straightforwardly, I argue, because of the basic difference between the Second-Personal and the Realist accounts.

²⁹ See the introduction to Ch.3.

³⁰ See Lerm 2013, pp.9-11.

³¹ Darwall 2006, p.169, citing Mill 1863/1998, Ch.3, s.1.

³² For instance, in Ch.4, s.3.4, I objected to Darwall's unsupported assumption that any case of holding an agent accountable is a second-personal matter.

³³ See Ch.3 and 4.

Recall that, on the Second-Personal account, your objection to the foot-treader would be “addressed to him [the foot-treader] [...] as the person causing gratuitous pain to another person [...]”³⁴ That is, on the Second-Personal account, your objection to the foot-treader is a function of the relationship between the two of you; namely, *his* standing on *your* foot. And, as we explored in Chapter 3,³⁵ the trouble with this account is how the second-personal relationship – such as that between yourself and the foot-treader – might be expanded to include the whole of the moral community whilst nevertheless retaining its sense as second-personal. How could we explain *second-personally* the fact that a third-person member of the general moral community, Barry, also has the authority to hold the foot-treader accountable for standing on *your* foot? On the other hand, an account that does not attempt to ground moral reasons in a second-personal relationship does not suffer this difficulty. Contrast Darwall’s Second-Personal account of the foot-treader scenario with the Realist one Darwall describes: on the Realist account,

[t]he reason would not be essentially for him [the foot-treader] as the agent causing another person pain. It would exist, most fundamentally, for anyone who is in a position to effect your relief and therefore for him, since he is well placed to do so.³⁶

So, on the Realist account, if an agent stands on your toes, then he has a reason to remove his foot from on top of yours, and he has this reason on the Realist account because “he is well placed to” correct a bad state of the world – your pain – by removing his foot from on top of yours, and not because of any special relationship in which he stands to you. On the Realist account, moral obligations aren’t *to* anybody; they just *are*. The implication of this is that if, on this Realist account, somebody else were also in a position to correct the bad state of the world that is your pain, then he/she would also have an obligation to so act. If, for instance, a nearby doctor were able to relieve your pain by injecting you with a pain-killer, then that doctor would have a reason to do so too.

And, just as the foot-treader as well as the nearby doctor will have reason to relieve your pain when each is in a position to do so, I argue that any individual will have the authority to hold an agent accountable when that individual is in a position to do so. If I am correct that any individual will have the authority to hold an agent accountable when that individual is in a position to do so, then it will follow that it is not only the victim of a wrongdoing who may have the authority to hold an agent accountable, but also the agent him-/herself or even members of the general moral community – that is, those broader aspects of Mill’s Point will follow on this Realist account of the authority to hold agents accountable. They follow precisely because Realism rejects the second-personal relationship that Darwall places at the heart of his account.

³⁴ Darwall 2006, p.7; as cited in the introduction to my Ch.2.

³⁵ See Ch.3, s.2.1.

³⁶ Darwall 2006, p.6, as cited in my Ch.1, s.1.2.

My claim is that, on the Realist account, any individual – the victim, the agent him-/herself, or members of the general moral community – may have the authority to hold an agent accountable when that individual is in a position to do so. But, when is an individual “in a position” to hold an agent accountable? In order to address this question, we must first have an account of what it is for an agent to be *held accountable*. I turn to this question, below.

1.2.1. An account of accountability

Given just how much Darwall’s arguments for his Second-Personal account rely on accountability, the notion is left curiously unanalysed. Like at least one reviewer of *The Second-Person Standpoint*, I “found myself wishing at many points that [Darwall] would make more trouble for his basic concepts”.³⁷ Darwall does, of course, have plenty to say about what “is central” to accountability; for instance: “What is central is simply reciprocal recognition of the standing to make certain demands of one another”.³⁸ But, note that a claim like this, and the many others like it, concerns the *presuppositions* of accountability, what needs to be the case in order for accountability to take place – it does not tell us *what accountability itself consists in*. My argument is effectively that, when we attend to what accountability actually consists in, it becomes clear that “reciprocal recognition of the standing to make certain demands of one another” – that is, second-personality – *isn’t* necessarily presupposed. The Realist account can suffice.

So, what is it to be held accountable? As I have remarked, Darwall does not himself present an account of accountability. However, I do believe that one can be pieced together on the basis of the rest of Darwall’s text, by drawing out the implications of Darwall’s various comments on and appeals to accountability. The emerging account of accountability will not be one that I necessarily advocate as the correct account of accountability; it will be the account of accountability that Darwall, rightly or wrongly, implicitly presumes – *and*, I will argue, it will be one that does *not* necessarily presuppose second-personality, *contra* Darwall. My aim, in other words, will be to defeat Darwall on his own terms.

The obvious place to start in reconstructing Darwall’s tacit account of accountability is with Pufendorf’s Point.³⁹ Recall that, according to Pufendorf – and as endorsed by Darwall – genuine moral accountability “forces a man to acknowledge of himself that the evil [i.e. the punishment], which has been pointed out to the person who deviates from an announced rule, falls upon him justly”, where the agent “is forced of [him]self to weigh [his] own actions, and to judge [him]self

³⁷ Lavin 2008.

³⁸ Darwall 2006, pp.83-84.

³⁹ See my Ch.1, s.6.1.2.

worthy of some censure, unless [he] conforms to a prescribed rule”.⁴⁰ Mere behavioural modification by negative reinforcement, on the other hand, works by “point[ing] out some object of terror” to the subject which “shakes the will with an external force”.⁴¹

The difference between genuine moral accountability and mere behavioural modification, then, is that a certain *inner state* is present in cases of the former and not in cases of the latter. Genuine moral accountability consists in a certain internal state, whereas mere behavioural modification “shakes the will with an external force”. This internal/external distinction is captured by Darwall’s own explanation of Pufendorf’s Point as follows:

To see what Pufendorf is getting at, we need to distinguish between merely external censure or blame [...] and internal blame, that is, the self-reactive attitude of guilt.⁴²

That is, the internal state distinctive of genuine moral accountability is *guilt*.

But, we must take care to understand ‘guilt’ correctly. Darwall doesn’t have in mind just the *feeling* of guilt; as he puts it: “When we consider that we have done something we think wrong, we don’t just have a painful feeling.”⁴³ Over and above the “painful feeling”,⁴⁴ there must be a cognitive component to guilt: the agent’s belief that he/she is guilty. The inner state, therefore, that is essential to genuine moral accountability is the *belief* that one is guilty. Indeed, as Pufendorf puts it (above), genuine moral accountability consists in the agent’s *acknowledgement* or *judgement* that he/she is “worthy of some censure”.⁴⁵ Darwall also cites with approval the following remark by Adam Smith:

Smith writes insightfully that when we resent injuries, what our resentment is “chiefly intent upon, is not so much to make our enemy feel pain in his turn, as [...] to make him *sensible*, that the person whom he injured did not deserve to be treated in that manner” [...].⁴⁶

That is, where the chief objective is not to induce in the guilty agent “a painful feeling”,⁴⁷ but to make the agent *sensible* of his/her guilt. And, amusingly, Darwall also cites the Alanis Morissette song, ‘You Oughta Know’:

And I’m here to *remind* you
Of the mess you left when you went away
It’s not fair to deny me
Of the cross I bear that you gave to me

⁴⁰ Pufendorf 1672/1934, p. 91, as cited by Darwall 2006, pp.111-112, as cited in my Ch.1, s.6.1.2.

⁴¹ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, pp.111-112, as cited in my Ch.1, s.6.1.2.

⁴² Darwall 2006, p.112, as cited in my Ch.1, s.6.1.2.

⁴³ Darwall 2006, p.169.

⁴⁴ Indeed, it’s not clear to me that the “painful feeling” necessarily attends guilt at all; it seems to me that an agent may hold him-/herself accountable just by recognizing his/her culpability, without actually *feeling bad* about it. My argument does not depend on this, however.

⁴⁵ Pufendorf 1672/1934, p. 91, as cited by Darwall 2006, pp.111-112, as cited above.

⁴⁶ Darwall 2006, p.84, citing Smith c.1762/1982, pp.95-96 (emphasis added).

⁴⁷ Darwall 2006, p.169, as cited above.

You, you, you oughta *know*⁴⁸

where, again, the emphasis is on bringing the agent to a state of *knowledge* of his/her guilt. The agent “oughta *know*” that what he/she did was wrong.

So, I suggest on the basis of all these texts that we analyse genuine moral accountability as *the agent’s conviction that what he/she did was wrong, such that he/she may justly be (proportionally⁴⁹) punished for it*. This, I suggest, is what Darwall has in mind by ‘guilt’, rather than the mere “painful feeling”⁵⁰ thereof. An agent is genuinely morally accountable, therefore, when he/she is convicted that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it, and an individual *holds* an agent genuinely morally accountable when that individual brings the agent to such a state of conviction, which is the internal state that distinguishes it from mere behavioural modification. I have arrived at this account of accountability on the basis solely of Darwall’s own text and texts Darwall cites with approval. It strikes me as a perfectly plausible account of accountability.

However, importantly for my purposes, note that nothing in this account of accountability requires, nor even indeed suggests, second-personality. If accountability is the agent’s conviction that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it, then I can hold you accountable – that is, induce in you the conviction of your guilt – by any numbers of ways: sitting down and explaining to you why what you did was wrong, renting out a billboard along your daily commute proclaiming the message “What you did was wrong!”, putting you in prison, or even sitting in a room next door and pressing a button to pass an electric shock through your body.⁵¹ Importantly, not all of these ways of inducing your conviction will require your reciprocal recognition of me-as-holding-you-accountable – that is, not all of these ways of holding you accountable will be second-personal. Because belief can have various causes (which are not limited to testimony), if an agent is held accountable when he/she comes to the belief that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it, then an agent may be held accountable in any number of ways, not all of which will be second-personal.

I have suggested that, just as on the Realist account an agent is obligated to, say, relieve a victim-of-foot-treading’s pain when that agent is in a position to do so, an individual has the authority to hold

⁴⁸ Morissette 1995, as cited by Darwall 2006, pp.73-74, fn.19 (emphases added).

⁴⁹ I have not yet had anything to say in support of proportionality (nor, however, have I had anything to say that would weigh against proportionality). I take for granted that no plausible view of accountability would permit execution as a punishment for pickpocketing or that a murderer deserves only a telling off. As Darwall (2006) has nothing to say on the matter of proportionality, neither do I.

⁵⁰ Darwall 2006, p.169, as cited above.

⁵¹ See Ch.4, s.3.4.

an agent accountable when that individual is in a position to do so.⁵² When will an individual be in a position to hold an agent accountable? If the preceding account of accountability is correct, an individual will be in a position to hold an agent accountable when that individual is in a position to induce in the agent the belief that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it – where this bringing to belief may take various non-second-personal forms. Below, I expand upon what it is for an individual to be in a position to bring an agent to a state of conviction, and how this allows the Realist to account for those broader aspects of Mill’s Point that were such a sticking point for the Second-Personal account.

1.2.2. Being “in a position”

According to Darwall’s Realist account of the foot-treader scenario, the foot-treader has a reason to remove his foot from on top of yours when he is “in a position” to relieve your pain, thereby leaving open the possibility that another agent, such as the nearby doctor, may also be in a position to relieve your pain, and will thereby also have a reason to do so (by, for instance, injecting you with a pain-killer).⁵³ I suggest, likewise, that on the Realist account an individual will have the authority to hold an agent accountable when that individual is “in a position” to do so. If, then, I am correct that genuine moral accountability consists in the agent’s conviction that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it,⁵⁴ it will follow that an individual is in a position to hold an agent accountable when the individual is in a position to bring the agent to this state of conviction. The question that the Realist needs to address, then, is: When is an individual in a position to bring an agent to a state of conviction?

This isn’t a very exciting question. Contrast the Realist’s question, ‘When is an individual in a position to bring an agent to a state of conviction?’, with the question that the Second-Personal theorist has to answer: ‘When does an individual have ‘second-personal authority’?’ Whereas we have struggled to secure a non-question-begging answer to the latter question that does not try to break into the Second-Personal circle from outside it,⁵⁵ the answer to the former question is just the same as the answer to: When is an individual in a position to bring any subject to any state of belief? And, the very standard answer to this question is just: minimally, when the individual knows what he/she is talking about, and has the means of communicating it to his/her audience. So, as a rough but plausible guess, I suggest that, minimally, an individual’s being “in a position” to hold an agent accountable requires: 1. the individual’s knowledge of what I will call “the facts of wrongdoing”: namely, what the agent was obligated to do, and whether or not the agent respected this obligation

⁵² See s.1.2, above.

⁵³ See s.1.2, above.

⁵⁴ See s.1.2.1, above.

⁵⁵ See s.1.1, above, for the most recent version of this complaint.

(and possibly also knowledge of any mitigating circumstances); and 2. the individual's ability to communicate this knowledge to the agent, where this communication may take the form of verbal expression, punishment, reactive attitudes, or possibly even other expressions (if you're Alanis Morissette, a song⁵⁶).

If these are the criteria of being "in a position" to hold an agent accountable, then it will follow that *various* individuals may have the authority to hold an agent accountable; that is, this Realist account will be able to account for Mill's Point and all its broader aspects. Let us use the foot-treader scenario again for the purposes of illustration. You, the victim of foot-treading, will be in a position to hold the foot-treader accountable, because you know that the foot-treader is standing on your foot, and you have the means of communicating this to him. But, a member of the general moral community, like Barry, may also be in a position to hold the foot-treader accountable: if Barry notices that the foot-treader's foot is planted on yours and knows that this kind of arrangement produces pain, and if Barry is around to tell the foot-treader off, then Barry is in a position to hold the foot-treader accountable. That is, if these conditions obtain, it will follow that a member of the general moral community will have the authority to hold the agent accountable, just as the victim him-/herself does. Whereas Darwall's Second-Personal account faced severe difficulty accounting for this broader aspect of Mill's Point,⁵⁷ the Realist can account for it with no trouble at all.

Furthermore, on this Realist account it will follow that the foot-treader has the authority to hold *himself* accountable: if he gradually becomes aware of the uneven surface beneath his foot (namely, the contours of your foot) and that it is causing you pain, he will be convicted of his wrongdoing and thus will hold himself accountable. This is because, on the account of accountability under discussion, what it is to be held accountable is just for one to hold the belief that what one did was wrong, such that one may justly be (proportionally) punished for it,⁵⁸ so, if this is a belief that one comes to by one's own senses, then one has effectively held oneself accountable. And, an agent will, in normal cases,⁵⁹ possess the authority to hold him-/herself accountable in this way. Since any agent is necessarily 'involved' in the act of wrongdoing just as much as any victim is (as above), the agent will be in a position to know the relevant facts of wrongdoing just as well as the victim might, and will thus, on this Realist account, possess the same authority as the victim to hold the agent – that is, him-/herself – accountable. Therefore, the second broader aspect of Mill's Point, according

⁵⁶ See Morissette 1995, as cited in s.1.2.1, above.

⁵⁷ See my Ch.3, s.2, and Ch.4.

⁵⁸ See s.1.2.1, above.

⁵⁹ One notable exception will be if the agent dies shortly after committing the action.

to which an agent is liable to being held accountable not only by the victim of his/her actions but also by him-/herself, also follows straightforwardly on the Realist account.

Thus, this Realist account is able to account for those broader aspects of Mill's Point very neatly. Members of the general moral community may be in a position, and thus possess the authority, to hold agents accountable when those members of the general moral community have knowledge of the facts of wrongdoing and the means of communicating this knowledge to the relevant agent, thereby bringing the agent to the conviction that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it. Likewise, the agent him-/herself may be in such a position, and indeed presumably will be most of the time, since the agent – being the agent – will almost always have access to the relevant facts of wrongdoing, in ways that members of the general moral community may sometimes lack. Thus Realism is able to account for those broader aspects of Mill's Point. The Realist is able to achieve all of this because, unlike Darwall, the Realist does not reject epistemic authority; if the authority to "give" reasons and hold agents accountable to those reasons is grounded in knowledge of facts of the world – a world we all share, and knowledge of which we all in principle have access to – then the subsequent account of the universality of those reasons and the authority of the general moral community to hold agents accountable to those reasons is much more straightforward. Darwall's Second-Personal account is not the best explanation of Mill's Point, since Realism is at least as good as it. As such, (and for several other reasons⁶⁰), I maintain that Darwall's Mill's-Point inference to the best explanation fails.

This amounts to two serious Realist strikes against Darwall's Second-Personal account so far: where the Second-Personal account does not offer a non-circular account of validity,⁶¹ the Realist account thereof is the natural one anyway;⁶² and, whereas Darwall's Second-Personal account struggles to accommodate those broader aspects of Mill's Point,⁶³ they follow straightforwardly as logical implications of a Realist account of accountability. Furthermore, the Realist account is poised to make a third strike against the Second-Personal account: Pufendorf's Point, as I now go on to argue.

1.2.3. Pufendorf's Point

Above, I argued that an agent who comes to believe, on the basis of his/her own senses, that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it, is one who holds him-/herself accountable. Note, though, that this will be the outcome any time any individual brings another agent to a state of conviction; that is, any time any individual holds another agent

⁶⁰ See Ch.3 and 4.

⁶¹ See Ch.2, s.2.3.

⁶² See s.1.1, above.

⁶³ See Ch.3 and 4.

accountable, that agent will necessarily thereby hold him-/herself accountable – and this is just Pufendorf's Point.⁶⁴ If accountability is the agent's belief that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it), then Pufendorf's Point will follow as a matter of the transparency of belief,⁶⁵ as I explain below.

According to my analysis of (Darwall's claims regarding) accountability,⁶⁶ an individual holds another agent genuinely morally accountable when that individual brings the agent to a state of conviction, that is, when that individual leads the agent to believe that what the agent did was wrong, such that he/she may justly be (proportionally) punished for it. And, I have suggested that, on the Realist account, an individual will have the authority to hold another agent accountable in this way when that individual has knowledge of the facts of wrongdoing, and the means of communicating them to the agent.⁶⁷ Now, when it is a case of an agent holding him-/herself accountable, the second criterion – the means of communicating one's knowledge – of course evaporates: if I know that I'm guilty, then I will know that I know that I'm guilty; this knowledge is automatically 'communicated' to myself. All an agent requires, therefore, in order to possess the authority to hold him-/herself accountable is the knowledge of the facts of his/her wrongdoing. Because of the transparency of belief, the agent's knowledge of the facts of his/her wrongdoing will automatically be 'communicated' to him-/herself – that is to say, the agent will thereby hold him-/herself accountable, just in virtue of his/her conviction and the transparency thereof. An agent who believes that he/she is guilty *is* an agent who holds him-/herself accountable, on this account.

So, it follows that, if an agent is held accountable by *another* individual – that is, if an individual brings an agent to a state of conviction, wherein the agent believes that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it – then it will follow, just on the basis of the agent's being brought to this state of belief by another individual, that that agent thereby holds him-/herself accountable. If another individual holds me accountable – that is, brings me to a state of conviction – then I will believe that I am guilty, and will therefore, transparently, come to believe that I believe that I am guilty, and will, in so doing, hold myself accountable also. Far from revealing the ground-breaking second-personal nature of obligation,⁶⁸ therefore, Pufendorf's Point turns out to be no more than an implication of the well-known transparency of belief. Moreover, Pufendorf's Point is easily captured by a Realist account of accountability.

⁶⁴ See Ch.1, s.6.1.2.

⁶⁵ See, for instance, Moran 2001.

⁶⁶ See s.1.2.1, above.

⁶⁷ See s.1.2.2, above.

⁶⁸ See Ch.1, ss.6.1.2 – 6.2.2.

This is the third Realist strike against Darwall's Second-Personal account: not only does Realism provide a much more natural and straightforward account of validity than does the Second-Personal account,⁶⁹ and not only does Realism easily account for those broader aspects of Mill's Point where the Second-Personal account struggles to do so,⁷⁰ Realism is even able to provide an account of Darwall's champion argument for Morality as Equal Accountability: Pufendorf's Point. That is, for all three of the major criticisms I have levelled against Darwall's Second-Personal account, Realism is able to succeed. Considering that Realism is the metaethic against which Darwall explicitly pits his Second-Personal account,⁷¹ I take this to be a compelling indictment of Darwall's view.

However, there is one further criticism I have levelled against Darwall's Second-Personal account, and which I have not yet raised in this present chapter: the suggestion that there is a *discordance* between the notion of second-personality and liability to being held accountable by an entire moral community.⁷² I turn now to consider the possible Realist responses to those arguments.

1.2.4. Forgiveness

In Chapter 3, we considered Darwall's insistence that the particular victim of a misdeed "has no special standing others do not have to blame the wrongdoer or to hold him responsible",⁷³ and that all members of the moral community have equal authority to blame a wrongdoer. In response, we considered Wallace's objection that:

the person who is wronged by you has a privileged basis for complaint against you, an objection to your conduct that is not shared by mere observers to what was done. The notion that someone in particular has been wronged by your action is conceptually connected to the idea that the wronged party has special ground for complaint [...].⁷⁴

However, we saw that Darwall has a ready response to Wallace's worry, which is that, even though Darwall claims that all members of the moral community possess the same authority to blame a wrongdoer, the victim him-/herself *does* possess the "privileged basis" or "special ground for complaint" that Wallace requires – only that it is the "distinctive standing to [...] *forgive an injury*".⁷⁵ I called this Darwall's 'forgive response' to Wallace's complaint.

I did, in Chapter 3, go on to present several objections to Darwall's forgiveness response, in particular arguing that Darwall's forgiveness response is *ad hoc*, as well as querying Darwall's presumption that only victims possess the special authority to forgive wrongdoers, since this

⁶⁹ See s.1.1, above.

⁷⁰ See s.1.2.2, above.

⁷¹ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁷² See Ch.3, s.2.1.

⁷³ Darwall 2010a, p.262, as cited in my Ch.3, s.2.1.

⁷⁴ Wallace 2007, p.29, as cited in my Ch.3, s.2.1.

⁷⁵ Darwall 2010a, p.262 (emphasis added), as cited in my Ch.3, s.2.1.2.

presumption flies in the face of our common practice of State pardons and the widely held belief that God can forgive sins.⁷⁶ I take these considerations (amongst others raised in prior discussion⁷⁷) to be persuasive reasons for rejecting Darwall's forgiveness response. Nevertheless, I wish now to explore whether Realism can again one-up the Second-Personal account. Can Realism provide an account of the victim's "privileged basis" or "special ground for complaint" that Wallace has required?⁷⁸ I will argue that it can. Like Darwall, I will argue that the victim of a moral transgression possesses the special authority to forgive the wrongdoer. Moreover, I will show how this Realist forgiveness response avoids (most of) those objections I earlier levelled against Darwall's forgiveness response. In particular, I will show how the Realist forgiveness response actually follows on the basis of the Realist account of when individuals are in a position to hold agents accountable, and as such avoids the charge of an *ad hoc* response. I will also show how the Realist view can account for the widespread belief that a third-party, such as God, might have the authority to forgive an agent even when that agent has not wronged God himself.

Firstly, we must ask: What *is* forgiveness? According to Darwall,

[f]orgiveness acknowledges the other's responsibility for wronging one, but refrains from pressing claims or "holding it against" him.⁷⁹

That is, to forgive is to forgo punishing the guilty agent.⁸⁰ However, "forgiveness nevertheless functions within the [...] space of holding responsible."⁸¹ Following Strawson, Darwall considers forgiveness a reactive attitude.⁸² To forgive an agent, therefore, is to forgo punishing that agent, while nevertheless holding the agent accountable. With the help of the account of what it is to hold an agent accountable that we drew out of Darwall's text, above,⁸³ we can fill in the details of this picture.

From Darwall's text we have extracted an account of accountability according to which an agent is held accountable when the agent is convicted that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it. And, if an agent is held accountable when the agent is convicted that what he/she did was wrong, such that he/she *may* justly be (proportionally) punished for it, then an individual may bring an agent to this state of conviction – the belief that he/she (the agent) *deserves* punishment – without actually meting out that punishment. This helps us to see

⁷⁶ See Ch.3, s.2.1.2.

⁷⁷ See Ch.3, s.2.1.2, and Ch.4, s.3.4.

⁷⁸ Wallace 2007, p.29, as cited above.

⁷⁹ Darwall 2006, p.72.

⁸⁰ Thanks to Professor David Benatar for pointing out to me that this is not an uncontroversial account of forgiveness. It is, however, the account of forgiveness that Darwall (as cited above) accepts.

⁸¹ Darwall 2006, pp.72-73.

⁸² Darwall 2006, p.72, referencing Strawson 1968.

⁸³ See s.1.2.1, above.

how forgiveness “refrains from pressing claims or “holding it against”” the agent,⁸⁴ but how forgiveness “nevertheless functions within the [...] space of holding responsible.”⁸⁵ A forgiven agent, therefore, is one who realizes that what he/she did was wrong, such that he/she *may* justly have been (proportionally) punished for it – and that he/she is fortunate to have escaped that punishment. So, forgiving is not the same as ignoring a wrongdoing; forgiveness requires the agent’s acknowledgement that a wrong act was committed. Consider, for instance, South Africa’s famous Truth and Reconciliation Commission’s (TRC’s) requirement that perpetrators of Apartheid horrors first confess their misdeeds before receiving amnesty – which was modelled, of course, on the Christian practice of confession of sins. To forgive is not to ignore the wrongdoing, therefore, but to *bring the agent to conviction of his/her wrongdoing – in the absence of punishment.*

According to Darwall, it is the victim of wrongdoing (rather than any member of the general moral community) who possesses the special authority to forgive an agent, and it is this authority to forgive that accounts for what is “special” or “privileged”⁸⁶ about the victim’s position, as Wallace has required. However, I have criticized this Second-Personal forgiveness response on the grounds that it is an *ad hoc* response.⁸⁷ Why is it the authority to forgive in particular that Darwall cites as constituting the victim’s special authority, and not any of the other personal reactive attitudes listed by Strawson?⁸⁸ Below, I argue that, where the Second-Personal forgiveness response is strikingly *ad hoc*, the Realist version of the forgiveness response is an implication of the Realist appeal to epistemic authority, and as such avoids the *ad hoc* charge.

Above, I argued that a member of the general moral community, such as Barry, may possess the authority to hold an agent accountable even when Barry was not himself wronged by that agent’s actions. Barry will possess this authority, on the Realist account of accountability, when he is in a position to hold the agent accountable – that is, when Barry has knowledge of the facts of the agent’s wrongdoing, and the means of communicating this to the agent, with the effect that the agent is brought to a state of conviction (i.e. held accountable).⁸⁹ But, although Barry may have knowledge of the facts of wrongdoing by observing the foot-treader’s foot pressing down on yours, I take it that you, as the victim of foot-treading, will hold an *intimate* knowledge of this wrongdoing, of an order that a member of the general moral community will lack. You are, after all, as the victim of foot-treading, the one who actually suffers the excruciating pain of the foot-treader’s foot

⁸⁴ Darwall 2006, p.72, as cited above.

⁸⁵ Darwall 2006, pp.72-73, as cited above.

⁸⁶ Wallace 2007, p.29, as cited above.

⁸⁷ See Ch.3, s.2.1.2.

⁸⁸ Strawson 1962/2003, p.196.

⁸⁹ See s.1.2.2, above.

crushing yours, whereas Barry merely witnesses the incident. The Realist may plausibly suppose, therefore, that the victim's special authority to forgive is grounded in the victim's special, intimate knowledge of the facts of wrongdoing.

To forgive, recall, is to bring the agent to conviction of his/her wrongdoing in the absence of punishment (above). And, arguably, an individual will be in a privileged position to bring an agent to conviction in the absence of punishment – that is, to forgive the agent – when that individual has special knowledge of the facts of wrongdoing. Since victims, I have suggested (above), will normally possess this special knowledge of the facts of wrongdoing, it will follow on the Realist account that victims will possess the special authority to forgive wrongdoers, whereas ordinary members of the moral community – who will generally lack this special, intimate knowledge – will therefore lack the special authority to forgive. Equipped with the special, intimate knowledge of the facts of wrongdoing, the Realist may argue, the victim will be able to bring the agent to conviction just by *explaining* to the agent the harm that was done: how much it really hurt when you stood on my foot, how I suffer from gout to begin with, how that is the second time today now that that toe has sustained injury, *etc, etc*. Given the victim's special knowledge of the facts of wrongdoing, the victim will be able to bring the agent to conviction without meting out any punishment, but just by non-punitively communicating to the agent the harm that was done. (Here, I think more serious examples of wrongdoing than foot-treading make a much more compelling case: how rape ruined my life, how my self-esteem was completely demolished to the point that I can't even hold down a job as a cashier, how I find myself completely unable to trust another human being, how I am disgusted with myself for resenting my own children, just because they remind me of you, *etc, etc*. These are the sorts of facts to which members of the general moral community plausibly would not have access.) Members of the general moral community, on the other hand, are not privy to all these details of a wrongdoing. Barry knows enough to have the authority to hold the foot-treader accountable – after all, Barry can *see* that the guy is standing on your foot – but Barry may not know enough to bring the foot-treader to a state of conviction just on the basis of explaining to him the harm that was done; that is, Barry and the other members of the moral community will not possess the authority to forgive the agent. Since Barry and the other members of the moral community lack the special, intimate knowledge of the facts of wrongdoing, in order to bring the agent to a state of conviction they will have to rely on more persuasive means of expression, like meting out punishment. On the other hand, since the victim will possess this special, intimate knowledge of the facts of wrongdoing, the victim will be able to bring the agent to a state of conviction even if he/she chooses to forgo punishment – that is, the victim will possess the “distinctive standing to [...] *forgive*

an injury".⁹⁰ Thus, that a victim possesses the special authority to forgive an agent – an authority which members of the general moral community lack – follows on the basis of the Realist account of accountability. The Realist forgiveness response to Wallace's worry is not an *ad hoc* explanation, unlike Darwall's forgiveness response.⁹¹

One further, interesting implication of this Realist account is that it does not in principle limit the authority to forgive to only the victim of wrongdoing. That is, if a non-victim, third-party has special, intimate knowledge of the facts of wrongdoing, of at least the same order as that possessed by the victim, then such a third-party will also be in a position to forgive the agent, that is, to bring the agent to a state of conviction in the absence of punishment. Note that this falls far short of the implausible suggestion that *any* third-person member of the moral community will have the authority to forgive an agent (when that individual was not him-/herself the victim of wrongdoing). It is, rather, the much more modest suggestion that sometimes a third-party member of the moral community will possess the authority to forgive an agent – but only when that third-party individual is in a special epistemic position; that is, when that third-party individual possesses a special, intimate knowledge of the facts of wrongdoing, of at least the same order as the victim's knowledge thereof. I take it this a condition that would not commonly obtain; it would obtain perhaps only in cases in which the third-party member of the moral community is somebody very emotionally close to the victim and is somebody with a highly developed capacity for empathy. So, although the Realist account does in principle allow that non-victim, third-party members of the general moral community may possess the authority to hold an agent accountable, I doubt that this will work out to be the case in practice.

I take it that Darwall would object to this implication of the Realist account that a third-party member of the moral community might possess the authority to forgive an agent (when that individual is not him-/herself the victim of wrongdoing). However, I question (as I did in Chapter 3⁹²) whether this is a reasonable objection, for at least three reasons.

Firstly, even if it is a commonly held belief that only the victim of wrongdoing has the authority to forgive the wrongdoer,⁹³ this is a controversial point in the philosophical literature on forgiveness. At least four philosophers – Paul Hughes,⁹⁴ Glen Pettigrove,⁹⁵ Linda Radzik,⁹⁶ and (in his more recent

⁹⁰ Darwall 2010a, p.262 (emphasis added), as cited above.

⁹¹ See my Ch.3, s.2.1.2.

⁹² See Ch.3, s.2.1.2.

⁹³ See, for instance, Swinburne 1989 and Govier & Verwoerd 2002, as cited by Hughes 2014, s.5.3.

⁹⁴ Hughes 2014, s.5.3. Note that Hughes restricts this admission to God only, denying that State pardons are genuine cases of interpersonal forgiveness.

⁹⁵ Pettigrove 2009, as referenced by Hughes 2014, s.5.3.

work, where he departs from his earlier point of view) Jeffrie Murphy⁹⁷ – have admitted the possibility that a non-victim third party may possess the authority to forgive a wrongdoer. As such, in the absence of further argument, the Realist account of accountability cannot be flatly rejected simply on the grounds that it admits the possibility that a non-victim third party may in principle possess the authority to forgive a wrongdoer, especially given that the requisite conditions of such authority are unlikely to obtain in reality (above). Moreover, below I question whether it really is a commonly held belief that only the victim of wrongdoing has the authority to forgive the wrongdoer, and thus whether it really is the case at all that this Realist account has counterintuitive implications.

My second response to Darwall's projected rejection of the Realist account of accountability on the grounds that it allows that a non-victim member of the general moral community may possess the authority to forgive a wrongdoer is this: I suggest that it is actually commonly held to be the case that some non-victim third parties indeed *do* possess the authority to forgive; for instance, it is believed by many that "the Son of Man has authority on earth to forgive sins".⁹⁸ Far from yielding counterintuitive implications, therefore, the Realist account of accountability seems to do a good job of accommodating the popular intuition⁹⁹ that God has the authority to forgive our sins, even when those sins wrong our fellow man and not God himself.¹⁰⁰ Note how effortlessly the Realist account can do this: the Realist account requires an individual to possess special, intimate knowledge of the facts of wrongdoing, of the same order as that possessed by the victim, before that individual has the authority to forgive the agent (above); and, since God is held to be omniscient, he *will* possess this special, intimate knowledge of the facts of wrongdoing; and so it will follow on the Realist account that God, like the victim of wrongdoing him-/herself, will have the authority to forgive the agent (while those ordinary, non-Divine members of the general moral community most likely will

⁹⁶ Radzik 2010, as referenced by Hughes 2014, s.5.3.

⁹⁷ Murphy 2009, as referenced by Hughes 2014, s.5.3.

⁹⁸ Matthew 9:1-8, New International Version, as cited in my Ch.3, s.2.1.2.

⁹⁹ The idea that God has the authority to forgive sins may be a *popular* intuition, but it is of course not a *universal* one; many people do *not* believe that God has the authority to forgive sins. However, this is not a problem for my present argument. I contend, with reference to those who do not believe that God has the authority to forgive sins, that these are people who believe that *God does not exist at all* (rather than believing that God does exist but lacks the authority to forgive sins). If God does not exist, then of course God is not omniscient, and it will not follow that God has the authority to forgive sins. That is, the belief that God does not have the authority to forgive sins because God does not exist is consistent with the Realist account that I have put forward.

¹⁰⁰ In Ch.3, s.2.1.2, fn.118, I considered the objection that "any sin against another person is ipso facto a sin against God the creator and, thus, that God's forgiveness of human wrongdoing is really for wrongs done to Him, making God's forgiveness of human wrongs not a genuine case of third-party forgiveness." (Hughes 2014, s.5.3.). My response to this possible objection, as before, is to echo Hughes:

[I]t is simply false to say that a wrong done to a particular person is actually a wrong done to someone else, or to God, and not to that person. Perhaps all injuries done by human beings to other human beings are in some sense also wrongs done to God, but that is consistent with the observation that wrongs endured by particular people are done to them. (Hughes 2014, s.5.3.)

not). (Indeed, since God is fully *omniscient*, while the victim of wrongdoing possesses only “special” knowledge of the facts of wrongdoing, this Realist account can also accommodate the popular belief that an omniscient God possesses a *higher* authority than mere mortals – even than the victims of wrongdoing – to forgive.) Insofar as we intuitively do hold that non-victim third parties, such as God, have the authority to forgive, I take it that the Realist account has the edge over Darwall’s account.

As a third response to Darwall’s projected rejection of the Realist account of accountability on the grounds that it allows that a non-victim member of the general moral community may possess the authority to forgive a wrongdoer, I question whether Darwall himself can consistently hold that it is only victims of moral wrongs who possess the authority to forgive. I have already criticized Darwall’s Second-Personal account (specifically his Second-Personal account of Mill’s Point) on the grounds that it embodies a certain discordance:¹⁰¹ that is, while Darwall’s account of moral reasons is that they are obligations owed *to another*, such as Sally, he nevertheless holds that an agent is liable to being held accountable by an entire moral community, including individuals such as Barry, to whom no obligations are owed.¹⁰² Now, Darwall’s insistence that only the victim of a wrongdoing has the authority to forgive does indeed appear to accord better with his analysis of moral reasons as second-personal: that is, a moral reason is owed to an individual, such as Sally, and it is subsequently only that individual – Sally, and not Barry – who possesses the authority to forgive me should I break my promise to her. However, this appearance of consistency is short-lived. Recall that Darwall, following Strawson, considers forgiveness a reactive attitude.¹⁰³ As Darwall was cited earlier, “forgiveness nevertheless functions within the [...] space of holding responsible.”¹⁰⁴ That is to say, Darwall counts forgiveness a *mode of holding accountable*. Now, if Darwall, following Mill, holds that any non-victim member of the moral community has the authority *to hold an agent accountable*,¹⁰⁵ and if Darwall, following Strawson, counts forgiveness a *mode of holding an agent accountable*, then, on the face of it, it follows that Darwall must admit that *any non-victim member of the moral community* has the authority to *forgive* a wrongdoer. It is unclear to me, therefore, how Darwall can consistently maintain that any member of the moral community has the authority to hold an agent accountable and yet also assert that only some of those individuals (namely, victims) have the authority to hold agents accountable in the particular manner that is forgiveness. At the very least, therefore, more argument is required to defend Darwall’s apparently inconsistent

¹⁰¹ See Ch.3, s.2.1.

¹⁰² See my Ch.4.

¹⁰³ Darwall 2006, p.72, referencing Strawson 1968, as referenced above.

¹⁰⁴ Darwall 2006, pp.72-73, as cited above.

¹⁰⁵ Mill 1863/1998, Ch.5, as cited by Darwall 2006, p.27, as cited in my Ch.3.

position here. (This, of course, feeds back into my complaint, above, that Darwall's forgiveness response to Wallace's worry is *ad hoc*.)

I conclude, therefore, that Darwall – especially given the rest of his views – cannot simply dismiss this Realist account of accountability on the grounds that it permits the possibility that some non-victim members of the general moral community may be in a position to forgive. The Realist forgiveness response, therefore, presents a further, serious objection to Darwall's Second-Personal view. Whereas Darwall's forgiveness response is apparently *ad hoc*, the Realist account of the victim's special authority to forgive follows on the basis of the Realist's grounding their account in epistemic authority. And, whereas Darwall's insistence that only the victim of a wrongdoing may have the authority to forgive flies in the face of the popular idea that God has the authority to forgive sins, the Realist account is easily able to accommodate this popular idea on the basis of God's supreme epistemic authority. I maintain that the Realist account of accountability is at least as good as – if not better than – Darwall's Second-Personal account, and, as such, Darwall is not justified in inferring Second-Personality as the best explanation of Mill's Point.¹⁰⁶

There is, however, one aspect of the forgiveness response in which, I will readily admit, the Realist account does not have the edge over the Second-Personal account. The Realist account of the victim's special authority to forgive does, in some sense, seem to miss the point of a discordance objection such as Wallace's. As I understand that objection, it is the claim that, if I have made a promise *to Sally*, and I go on to break my promise *to Sally*, then *Sally* surely subsequently holds me accountable in an entirely different capacity to that in which Barry might hold me accountable for breaking my promise to Sally.¹⁰⁷ Yet, while the Realist has successfully presented an account of the victim's "special" or "privileged" authority that Wallace has required,¹⁰⁸ in the form of the victim's special authority to forgive, this is nevertheless a "special" authority different only in *degree* and not in *kind* to that of the other members of the general moral community. This is quite clearly exhibited in the Realist forgiveness response's appeal to epistemic authority in both the victim's case and that of the ordinary member of the general moral community: Barry has the authority to hold me accountable for breaking my promise to Sally if Barry has knowledge of the facts of wrongdoing (and the means of communicating them to me);¹⁰⁹ Sally too has the authority to hold me accountable, as well as the special authority to forgive me should she choose to do so, because she has *more knowledge* than Barry does. What differentiates Sally's authority from Barry's on such an account is

¹⁰⁶ See Ch.3.

¹⁰⁷ See Ch.3, s.2.1.

¹⁰⁸ Wallace 2007, p.29, as cited above.

¹⁰⁹ See s.1.2.2, above.

just a matter of degree, therefore. And, I doubt that this is an adequate response to Wallace's worry, which, recall, is that "the person who is wronged by you has a privileged basis for complaint against you, an objection to your conduct *that is not shared by mere observers to what was done.*"¹¹⁰

However, although I count this a shortcoming of the Realist account of accountability, I do not consider it damaging to the present argument. Although, I readily admit, the Realist account does not have the edge over the Second-Personal account in this regard, nor does the Second-Personal account prevail over the Realist account when it comes to the kind of special authority granted to the victim; both present accounts according to which the "special" authority possessed by the victim is essentially the same in kind as that possessed by a member of the general moral community. Indeed, this was an objection to Darwall's forgiveness response that I originally raised when I first considered that argument in Chapter 3.¹¹¹ Darwall states explicitly that, on his Second-Personal account, "the victim of wrongdoing [...] has no special standing others do not have to blame the wrongdoer".¹¹² That is, despite her purported special authority to forgive me, when Sally holds me accountable, on Darwall's Second-Personal account, she does so in exactly the same capacity as Barry might – *contra* (my understanding of) Wallace's requirement (above). Thus, while I doubt that the Realist account of accountability meets Wallace's requirement that Sally "has a privileged basis for complaint against you, an objection to your conduct *that is not shared by mere observers to what was done*",¹¹³ I doubt that Darwall's Second-Personal account does either.

As I have already noted,¹¹⁴ I do not in this chapter wish to advocate Realism *per se*. However, what I do wish to do is demonstrate that a Realist account of accountability is a live possibility, but one which Darwall apparently overlooks. As such, Darwall's Mill's-Point argument does not offer strong support for Second-Personality.¹¹⁵ Whereas Darwall's Second-Personal account struggles to account for cases in which an agent is liable to being held accountable by a member of the general moral community, but where the agent is not obligated to such an individual,¹¹⁶ the Realist account suffers no trouble, grounding as it does the individual's authority to hold accountable in his/her epistemic authority.¹¹⁷ Whereas I have questioned whether Darwall's reliance on Pufendorf's Point is consistent with his otherwise second-personal claims,¹¹⁸ a Realist account of Pufendorf's Point

¹¹⁰ Wallace 2007, p.29, as cited above (emphasis added).

¹¹¹ See Ch.3, s.2.1.2.

¹¹² Darwall 2010a, p.262, as cited above.

¹¹³ Wallace 2007, p.29, as cited above (emphasis added).

¹¹⁴ See the introduction to this chapter.

¹¹⁵ See my Ch.3.

¹¹⁶ See Ch.4.

¹¹⁷ See s.1.2.2, above.

¹¹⁸ See Ch.2, s. 1.3.

follows just as a logical implication of the Realist account of accountability and the transparency of belief.¹¹⁹ Moreover, Realism can provide a straightforward, natural account of validity,¹²⁰ whereas no non-question-begging account is forthcoming from the Second-Personal quarter.¹²¹ The Realist can also provide an account according to which the victim will (generally) possess a special authority to forgive his/her wrongdoer, which other members of the moral community will lack, and, unlike the Second-Personal account,¹²² the Realist account can do this in a way that follows on the basis of the rest of its claims and is not an *ad hoc* response (above). Furthermore, where the Second-Personal account must flatly deny the widely held belief that God has the authority to forgive our sins, the Realist account of the authority to forgive can accommodate this popular belief (above). The Realist advantages, therefore, are many, and the Second-Personal advantages few. The only area in which Realism does not surpass the Second-Personal account is in responding fully to Wallace's complaint, by providing an account of Sally's authority to hold me accountable according to which her authority is entirely different in kind to that possessed by Barry – but there the Second-Personal account falls short too (above). I conclude, therefore, that the Realist account is preferable to the Second-Personal account in almost all respects so far considered. If Darwall's Mill's-Point argument is an inference to the best explanation, therefore,¹²³ he is not justified in concluding that moral reasons are second-personal, because second-personality is not the best explanation of Mill's Point; Realism is at least as good – if not better.

Below, I turn to some further, general considerations about the form of the Realist account, which I believe also make it preferable to Darwall's Second-Personal account.

1.3. Realist symmetry

In the preceding two sections, 1.1 and 1.2, I argued that Realism is at least as good as – if not better than – Darwall's Second-Personal account in that Realist explanation is more wide-ranging, capable of explaining more phenomena than the Second-Personal alternative. In this present section, I aim to argue that the Realist explanation of these phenomena is preferable, not only in that it is more comprehensive than the Second-Personal alternative, but also in the sense that the Realist explanation of these phenomena possesses a quality that the Second-Personal alternative does not. Whereas, I have argued, the Second-Personal account is discordant, I will argue that the Realist account is harmonious and symmetrical.

¹¹⁹ See s.1.2.3, above.

¹²⁰ See s.1.1, above.

¹²¹ See Ch.2, s.2.3.

¹²² See Ch.3, s.2.1.2.

¹²³ Darwall 2006, p.76, fn.26, as cited in the introduction to my Ch.3. See also Darwall 2010c, p.32.

In Chapter 3, section 2.1, I put forward the criticism that the Second-Personal account embodies ‘discordance’. That is, the Second-Personal claim that a moral reason is given by a second person to a first, existing in the relationship between the two of them, does not ‘fit’ very well with Darwall’s simultaneous claim that an agent is liable to being held accountable not only by the second person to whom that agent was obligated but by any member of the general moral community. One way to express this point¹²⁴ is that Darwall admits an *agent-neutral* understanding of accountability, such that any member of the general moral community possesses the authority to hold a wrongdoer accountable.¹²⁵ However, Darwall also wishes to advocate – by his own explicit admission¹²⁶ – an *agent-relative* account of the nature of obligation, according to which I have a moral reason in virtue of the relationship in which I stand to you. As Darwall puts the point with reference to the foot-treader scenario, the foot-treader has a reason to remove his foot from on top of yours “as the person causing gratuitous pain to another person”,¹²⁷ rather than as a matter of the way the world is. So, whereas Darwall’s Second-Personal account of the nature of moral reasons is agent-relative, his attendant account of accountability is agent-neutral. We might say, therefore, that the Second-Personal account is asymmetrical.

On its own terms, this isn’t a very compelling objection to the Second Personal account.¹²⁸ However, the objection does gain strength when a comparison is drawn with the Realist account. Whereas the Second-Personal account is asymmetrical, the Realist account is *symmetrical*. The same Realist view accounts both for the nature of moral obligation and our practice of holding each other accountable. On the Realist account, a moral reason is had by an agent in virtue of the way the world is. So the foot-treader, for instance, has a moral reason to remove his foot from on top of yours not “as the person causing gratuitous pain to another person”,¹²⁹ but rather as one “who is in a position to effect your relief”.¹³⁰ And, as we have noted,¹³¹ this has the implication that a nearby doctor who was also “in a position to effect your relief” would also thereby have a reason to relieve your pain (even though the doctor does not stand in the relationship to you of being the one causing you pain). On this Realist account, individuals are *vessels* of moral value, and as long as the vessel can

¹²⁴ I am hesitant to describe the argument in this way, though, given the reservations I have expressed about the ‘agent-relative’ and the ‘agent-neutral’ (see Ch.1, s.1.1, as well as Lerm 2013).

¹²⁵ See my Ch. 3 and 4.

¹²⁶ Darwall 2006, p.8, as cited in my Ch.1, s.1.1.

¹²⁷ Darwall 2006, p.7, as cited in my Ch.1, s.1.1.

¹²⁸ As I noted in Ch.3, s.2.1.

¹²⁹ Darwall 2006, p.7, as cited above.

¹³⁰ Darwall 2006, p.6, as cited in my Ch.1, s.1.1.

¹³¹ See s.1.2, above.

conduct moral value, any vessel will do.¹³² In Darwall's own words, Realism admits a "*state-of-the-world-regarding and agent-neutral*"¹³³ account of moral reasons. Because Realism is a (state-of-the-) world-regarding metaethic,¹³⁴ and because the world is shared by all of us – victim, agent and members of the general moral community alike – it follows that the Realist account of moral reasons is agent-neutral: the doctor may have a reason to relieve your pain just as well as the foot-treader might, for example. And, the world-regardingness of Realism *also* explains why a member of the general moral community or even the agent him-/herself may have the authority to hold an agent accountable just as well as the victim might: just as the doctor might have a reason to relieve your pain if he is in a position to inject you with a pain-killer (even though he is not the one standing on your foot), Barry might have the authority to hold me accountable for my broken promise to Sally when Barry has knowledge of the facts of wrongdoing and the means of communicating them to me¹³⁵ (even though I made no promises to Barry). That is, the Realist account gives rise to both an agent-neutral account of moral reasons *and* an agent-neutral account of accountability. Unlike the Second-Personal account, therefore, the Realist account is *symmetrical*. Insofar as the Second-Personal account is asymmetrical and the Realist account symmetrical, therefore, I argue that Realism is preferable.

I take this argument to be akin to considerations of parsimony, often used to adjudicate between two theories of equivalent explanatory power. That is, the consideration currently at issue is not one of the explanatory power of nor the content of the theory *per se*, but is instead one of the aesthetic form of the theory. Moreover, like considerations of parsimony, my current contention of symmetry is a relative claim: Realism is preferable to the Second-Personal account, I claim, insofar as the Realist account is symmetrical and the Second-Personal account is not. However, I say that my argument that Realism is symmetrical whereas the Second-Personal account is not is only *akin to* – but not the same as – considerations of parsimony, for two reasons. Firstly, a theory is parsimonious insofar as it posits fewer entities than a rival theory in order to do the same explanatory work. And, my present claim is not that Realism posits fewer entities than the Second-Personal account; it is unclear to me which of Realism or the Second-Personal account posits fewer entities.¹³⁶ Secondly,

¹³² This is a better description of Consequentialist Realism than it is of Non-Consequentialist versions thereof. However, as I have noted (Ch.1, s.2), in this project, by 'Realist' I refer to that theorist's non-second-personal response to the foot-treader as described by Darwall (2006, pp.5-6). As I have explained (Ch.1, s.2), this Realist response is Consequentialist.

¹³³ Darwall 2006, pp.5-6, as cited in my Ch.1, s.1.2.

¹³⁴ See Ch.1, s.1.1.

¹³⁵ See s.1.2.2, above.

¹³⁶ At first blush, it may appear obvious that Realism posits more entities than the Second-Personal account, and that the Second-Personal account is therefore more parsimonious. This is because Realism claims that moral reasons issue from the world (Ch.1, s.2), thereby positing such entities as Real moral norms (or, at least,

my present argument is one only akin to, but not the same as, one of parsimony, because such an argument claims that one theory is preferable to another when it is able to do the same explanatory work as the rival theory, and does so by positing fewer entities. However, my claim has not been that Realism and the Second-Personal account are able to do the same explanatory work – instead, my argument of the preceding two sections has been that Realism is able to do *more* explanatory work than the Second-Personal account.¹³⁷ And, I have argued in this section, Realism is able to do it *better*: symmetrically.

Lastly, we should consider the question of the symmetry of a metaethic in the context of Mill's Point. Darwall's Mill's-Point argument, recall, is that we can infer the second-personality of moral reasons on the basis of the fact of our practice of holding each other accountable.¹³⁸ In light of the above, we can see that Darwall's Mill's-Point argument tries to infer agent-relativity (second-personality) on the basis of agent-neutrality (the fact that an agent is liable to being held accountable by any member of the moral community). Moreover, the argument has the form of an inference to the best explanation,¹³⁹ so Darwall's Mill's-Point argument claims that we can best explain the agent-neutrality of accountability by way of the agent-relativity (second-personality) of obligation. And, put in those terms, it is clearly dubious that such an argument will succeed. Rather, a much better explanation of the agent-neutrality of accountability is surely the *agent-neutrality* of obligation. That is to say, the symmetry of the Realist account makes it a much better candidate for the explanation of Mill's Point than is Darwall's Second-Personal account. As such, I argue that Darwall's Mill's-Point argument for second-personality fails.

morally valuable states of affairs). The Second-Personal account, on the other hand, holds that moral reasons are given by a second person to a first (see Ch1, s.1.2), so this account posits no more than moral agents – the existence of which we were always committed to anyway. So it would appear that the Second-Personal account in fact posits no new entities at all, making it as parsimonious as can be, and indeed more parsimonious than the Realist account.

I take it that Darwall does view his account much like this: as the claim that all we need to posit in order to account for moral reasons is just the moral agent. However, it has been my objective in this project to show that that simple argument does not work. In one version of this persistent complaint: the fact that a second person has the authority to hold a first accountable does not entail that that first person has reciprocal authority to hold the second accountable (Ch.2, s.1.), and so the kind of reasons generated are not universal and are not moral (Ch.1, s.1.3). We need more than just the moral agent in order to account for moral reasons: before it is plausible that a second person might literally *give* a first a moral reason, we need to know that that second person's command is *valid* (Ch.2.3). However, as I have argued, the most plausible account of validity is the *Realist* one (Ch.2, s.2.3.1, and s.1.1, above). Since the Second-Personal account cannot claim that all that is necessary to account for moral reasons is the moral agent, but must give an account of the validity of a second person's commands, it is far from obvious to me that the Second-Personal account is indeed as parsimonious as it may at first have appeared to be.

However, I set these questions aside since my present point (above) is that the argument currently being advocated is *not* one of parsimony.

¹³⁷ See ss.1.1 and 1.2, above.

¹³⁸ See my Ch.3.

¹³⁹ Darwall 2006, p.76, fn.26, and 2010c, p.32, as cited in my Ch.3, introduction.

2. Arguments against Realism

In section 1 of this chapter, I presented the main reasons to favour a Realist account over a Second-Personal one, namely, that Realism is at least as good an explanation as – if not an even better explanation than – the Second-Personal account of the validity of moral reasons,¹⁴⁰ Mill’s Point and all its broader aspects,¹⁴¹ Pufendorf’s Point,¹⁴² as well as the agent’s special authority to forgive a wrongdoer¹⁴³ – and, not only does Realism have greater explanatory power than the Second-Personal account, I argued that the *kind* of explanation proffered by the Realist is preferable in that it is symmetrical, whereas the Second-Personal account is not.¹⁴⁴

In this section 2, I consider two objections to the Realist account suggested above. The first objection follows straightforwardly from the argument of section 1.3, above. Whereas I have argued that the kind of symmetrical account offered by the Realist is preferable to that offered by the Second-Personal theorist,¹⁴⁵ this first objection is straightforwardly that I am simply mistaken, and the kind of account offered by the Realist is one that presents precisely the *wrong* kind of reason for moral action. This is an objection to Realism presented by Darwall in *The Second-Person Standpoint*. The second objection that will be considered is not one that appears in *The Second-Person Standpoint*, but is essentially one put forward by Korsgaard in her own critique of Realism. Whereas I have argued that Realism provides a better explanation of Mill’s-Point than does Darwall’s own account,¹⁴⁶ this second objection is that Realism actually presents *no* “explanation” of any moral phenomenon at all.

I present both of these objections, below, and then go on to respond to them.

2.1. Reasons of the wrong kind, reasons of the right kind

Even if I have been successful in demonstrating that Realism is at least as good an – if not a better – explanation of Mill’s Point than Darwall’s own Second-Personal account, I expect that Darwall would object to this Realist attempt. I suspect that he would say that, even if Realism can account for Mill’s Point, it does so in the *wrong way*. (Compare this to Korsgaard’s complaint that, even if the Analogical argument can demonstrate that I have duties concerning you, it would do so “in the

¹⁴⁰ See s.1.1, above.

¹⁴¹ See s.1.2.2, above.

¹⁴² See s.1.2.3, above.

¹⁴³ See s.1.2.4, above.

¹⁴⁴ See s.1.3, above.

¹⁴⁵ See s.1.3, above.

¹⁴⁶ See ss.1.2.2 and 1.3, above.

wrong way”¹⁴⁷). This anticipated objection to Realism follows on the basis of what Darwall calls “Strawson’s Point”:

Strawson argued that social desirability is not a reason of “the right *sort*” for practices of moral responsibility “as we understand them” [...]. When we seek to hold people accountable, what matters is not whether doing so is desirable, either in a particular case or in general, but whether the person’s conduct is culpable and we have the authority to bring him to account. *Desirability is a reason of the wrong kind to warrant the attitudes and actions in which holding someone responsible consists in their own terms.*¹⁴⁸

For the purposes of this argument, I will accept Strawson’s Point that social desirability is “a reason of the wrong kind” for holding agents accountable. I will also accept Darwall’s claim that Strawson’s Point gives rise to Pufendorf’s Point.¹⁴⁹ Pufendorf’s Point, recall, is that there is an important distinction to be drawn between genuine moral accountability, on the one hand, and mere behavioural modification by negative reinforcement, on the other.¹⁵⁰ If, as Strawson claims, social desirability is “a reason of the wrong kind” for moral accountability then it follows that behavioural modification is not genuine moral accountability – that is, Pufendorf’s Point follows. To, say, tell me off for breaking promises *in order to* improve my future behaviour by decreasing the likelihood that I will break a promise again (i.e. in order to modify my behaviour) is to tell me off in order to make my future behaviour more *socially desirable*, and thus, it is not to hold me genuinely morally accountable. This is why, as Pufendorf observed, behavioural modification by negative reinforcement is not genuine moral accountability: behavioural modification by negative reinforcement is supported by “a reason of the wrong kind”: the *social desirability* of the modified behaviour. Rather, the only reason for holding an agent genuinely morally accountable is *because the agent deserves it* – because, when “the person’s conduct is culpable”, it is a matter of *justice* that he/she is held accountable.

¹⁴⁷ Korsgaard 1996, p.134, as cited in my Ch.1, s.2.2.

¹⁴⁸ Darwall 2006, p.15, citing Strawson 1962/1968, pp.72 and 74. See also Darwall 2006, pp.65-66.

¹⁴⁹ Of course, Strawson’s Point would not have been available to Pufendorf himself, writing as he did a couple of hundred years before Strawson was even born. But, it may be argued – and, I think, Darwall would argue – that, even though Pufendorf obviously wouldn’t have been familiar with “Strawson’s Point” *by name*, he might nevertheless have been aware (perhaps only tacitly) of the general idea of Strawson’s Point, and it was this general idea that led Pufendorf to make “Pufendorf’s Point”. However, it’s not clear to me that this is the case, that Pufendorf relies on anything like Strawson’s Point in making Pufendorf’s Point. I don’t see much evidence in Pufendorf (1672/1934) for this Strawsonian reconstruction, and I think it is more likely supported by the fact (if it is a fact) that God holds us accountable only at the end of our lives, or even at the end of all time (as I argued in Ch.1, s.6.1.2), or by an appeal to epistemic authority (as I argued in Ch.2, s.1.4). Yaffe (2010, p.252) complains similarly that in Darwall’s historical approach

lies the danger of taking affinity for identity, of ascribing a view to a figure that is, rather, the view of someone who thinks similarly to, but also differently from, the way that figure actually thinks. There is a hint of this in Darwall’s interpretation of Reid’s critique of Hume and I am inclined to argue the same of Darwall’s interpretation of Pufendorf. Nevertheless, I set these concerns aside because my aim in this section is to show that the Realist proposal is threatened by neither Pufendorf’s Point nor Strawson’s Point, nor the two Points working together.

¹⁵⁰ Darwall 2006, mainly pp.111-112, referencing Pufendorf 1672/1934, p.91, as cited in my Ch.1, s.6.1.2.

According to this objection, the Realist proposal is incompatible with genuine moral accountability, because it is committed to “a reason of the wrong kind” for holding agents accountable: desirable states of the world – namely, agents’ states of conviction. Just as, on the Realist account, the foot-treader has a reason to remove his foot from on top of yours when he is in a position to bring about a desirable state of the world – namely, the relief of your pain – I have argued that an individual has the authority to hold an agent accountable when that individual is in a position to bring about a desirable state of the world – namely, the agent’s conviction.¹⁵¹ So, on the Realist account, an individual’s holding an agent accountable is a means to a socially desirable end – the agent’s accountability – just as, say, my smacking the dog is a means to the desirable end of the dog’s no longer barking. But, Strawson’s Point is that social desirability is “a reason of the wrong kind”, so, just as my smacking the dog is not a case of genuine moral accountability, the individual’s telling the agent off on this Realist proposal is not a case of genuine moral accountability either. Thus, the Realist proposal misconstrues the nature of accountability. Darwall, of course, does not explicitly and deliberately respond to the Realist objection I have presented in this chapter, but I take it that this Strawson’s-Point objection would be his stance, as evidenced, for instance, by this quotation:

[H]ow can the authority of someone to demand something derive simply from the desirability of his being able to do so? [...] The most that it can ground is a desire to be able to demand something, not any demand, or practice of accountability, itself.¹⁵²

I have argued that the Realist proposal of this chapter is able to account for Mill’s Point, including all its broader aspects, and that, as such, Darwall’s Second-Personal account is not the best explanation of Mill’s Point, and so his Mill’s-Point argument for second-personality fails.¹⁵³ However, if Realism misconstrues the nature of moral accountability by trading on a “reason of the wrong kind”, then of course it is false that the Realist proposal provides an explanation of Mill’s Point at least as good as Darwall’s Second-Personal account: rather, Realism provides no explanation at all of Mill’s Point – which is a point about genuine moral accountability. If this Strawson’s-Point objection to the Realist proposal is correct, then Darwall’s Mill’s-Point argument for second-personality stands.

I will argue, however, that Realism is *not* committed to a “reason of the wrong kind” for accountability and therefore does not misconstrue the nature of accountability. As such, Realism continues to present a serious threat to Darwall’s Mill’s-Point argument.

¹⁵¹ See s.1.2.2, above.

¹⁵² Darwall 2006, pp.311-312.

¹⁵³ See s.1, above.

2.1.1. Response

As Nagel writes with regards to his defence of Realism: “In arguing for this claim, I am somewhat handicapped by the fact that I find it self-evident.”¹⁵⁴ Likewise, I find it fairly obvious how Realism is able to escape the charge that it misconstrues the nature of accountability. As such, the argument of this section will be relatively brief. I hope, however, that its relative brevity does not undermine its importance.

The Realist response to the Strawson’s-Point argument is just this: nothing in the Realist account commits it to social desirability as the reason for which agents are held accountable. (As one indication of this point, consider how the term “social desirability” (or equivalent) hadn’t arisen even once throughout the duration of the preceding discussion of Realist accountability, until the Strawson’s-Point charge was laid.) Of course, some versions of Realism *may* ground accountability in social desirability; for instance, Utilitarianism could allow that an agent ought to be held accountable when the action of holding that agent accountable produces a greater net utility, such as a deterrent effect. Holding an agent accountable because of the social utility of deterrence would indeed be a reason of the wrong kind for accountability, according to Strawson. But, notice that Realism generally is not committed to this – nor even is Consequentialist Realism committed to this. As Darwall concedes in a footnote, a Consequentialist Realist theory that located moral value not in social desirability but elsewhere would avoid the “reason of the wrong kind” charge:

Of course, [...] an [A]ct- or [R]ule-[C]onsequentialist theory of right might avoid this problem [of “a reason of the wrong kind”] if it could be adequately grounded in an equal claim to happiness, welfare, or something similar.¹⁵⁵

Realism, being a world-regarding metaethic, is necessarily committed to the claim that moral value is located in the world – but, of course, there are many more kinds of states of affairs in the world than just the socially desirable. Any metaethic that locates moral value in any worldly state of affairs will count as Realist (though, of course, some states of affairs will be more plausible candidates than others). A Realist may locate moral value in pleasure or happiness or friendship or justice, *etc.* Nothing intrinsic to Realism commits it to the idea that social desirability is morally valuable, and thus Realism is not necessarily committed to providing a “reason of the wrong kind” for

¹⁵⁴ Nagel 1986, pp.159-160.

To be completely accurate, Nagel actually makes this remark with regard to the fact (if it is a fact) that pain and pleasure are ‘agent-neutral’ reasons. (Korsgaard (1996, p.41) also cites this quotation as describing Nagel’s approach to Realism generally.) But, the quotation does capture the general attitude expressed in Nagel’s book as a whole, as, for instance, also expressed in: “[I]n general, there is no way to prove the possibility of [R]ealism; one can only refute impossibility arguments, and then more often one does this the more confidence one may have in the [R]ealist alternative.” (Nagel 1986, p.144).

¹⁵⁵ Darwall 2006, p.104, fn.27.

accountability. Those Realisms that reject the idea that what is morally valuable is socially desirable, therefore, will not violate Strawson's Point.

And, the kind of Realism I have been advocating is one according to which the world is such that obligated agents are agents who are liable to being held accountable, just as the world is such that electrons are negatively charged.¹⁵⁶ Electrons are not negatively charged because it's socially desirable, and neither are obligated agents liable to being held accountable because it's socially desirable. Rather, an agent is liable to being held accountable just because he/she *had reason* to perform an action which he/she failed to perform, and an individual is in a position to hold that agent accountable when the individual knows all this and has the means of communicating it to the agent.¹⁵⁷ That is, an individual who is in a position to hold an agent accountable, on the proposed Realist account, need only be in a position to bring the agent to a state of conviction; the individual needn't also be in a position to deter future, similar behaviour by that agent. If it were an additional criterion that an individual be in a position to bring about more socially desirable behaviour in the accountable agent, then of course this Realist account would start looking a lot more like an account of behavioural modification (rather than genuine moral accountability) – but it's *not* a criterion of the Realist account I have advocated. Note that the end of such Realist accountability is just the agent's *conviction* – that is, just the *belief* – that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it (bearing in mind, furthermore, that this is an account of accountability I distilled out of Darwall's own text).¹⁵⁸ In lucky cases, it may indeed follow as the result of such conviction that the agent's future behaviour will be socially more desirable – but that's not the *end* of accountability, as I have presented it; the end is just the agent's conviction, for it is this internal state (rather than any future behaviour) that distinguishes genuine moral accountability from mere behavioural modification.¹⁵⁹ I argue, then, that Realism – at least the kind of Realism I have advocated in this section – does not fall foul of Strawson's Point. The charge that Realism is committed to a reason of the wrong kind for accountability simply doesn't stick. Furthermore, I go on to argue that the Realist account I have advocated actually presents a reason of precisely the *right* kind for accountability.

If social desirability is a reason of the wrong kind for accountability, we should ask: What is a reason of the *right* kind? In a more recent paper, Darwall explains that:

To be a reason of the right kind, a consideration must justify the relevant attitude in its own terms. It must be a fact about or feature of some object, appropriate consideration of which

¹⁵⁶ See s.1.2, above.

¹⁵⁷ See s.1.2.2, above.

¹⁵⁸ See s.1.2.1, above.

¹⁵⁹ See s.1.2.1, above.

could provide someone's reason for a warranted attitude of that kind toward it. It must be something on the basis of which, or as we also say, "*for which*," someone could (and appropriately would) come to hold the attitude as a conclusion of a process of considering (deliberating about) *whether* to do so.¹⁶⁰

This explains why social desirability is a reason of the wrong kind for accountability: as Darwall was cited above, "The most that it can ground is a desire to be able to demand something, not any demand, or practice of accountability, itself."¹⁶¹ That is, the only "attitude" that social *desirability* justifies "in its own terms" is just the *desire* to hold an agent accountable, rather than the actual practice of holding accountable. So, what, then, justifies an agent's liability to be held accountable? It is hardly a ground-breaking observation that, where desirability justifies desire, *accountability* justifies liability to being held accountable. That is, the reason of the *right* kind for holding an agent accountable is when that agent is indeed *accountable*, or, as Darwall put it in his original statement of Strawson's Point, 'culpable':

When we seek to hold people accountable, what matters is not whether doing so is desirable, either in a particular case or in general, but whether the person's conduct is culpable and we have the authority to bring him to account.¹⁶²

Yet *none* of this presents an obstacle to the Realist proposal. I have argued that, on the Realist account, an individual has the authority to hold an agent accountable when that individual is "in a position" to do so, and that an individual is in a position to hold an agent accountable when 1. the individual knows the facts of wrongdoing, namely, what the agent was obligated to do, and whether or not the agent respected this obligation (and possibly also knowledge of any mitigating circumstances), and 2. the individual has the ability to communicate this knowledge to the agent.¹⁶³ That is, on this Realist proposal an individual holds an agent accountable because the individual *knows that the agent is culpable* (and possesses the necessary means of telling the agent this). Far from being a reason of the wrong kind for accountability, this is precisely a reason of the *right* kind: on this Realist proposal, individuals hold agents accountable *because those agents are accountable*.

Interestingly, more recently Darwall's claim has transformed from the original claim of Strawson's Point, that with a reason of the right kind "what matters is [...] whether the person's conduct is culpable and we have the authority to bring him to account",¹⁶⁴ to the much narrower claim that: "Reasons of the right kind [...] are irreducibly second personal."¹⁶⁵ Of course, if it *is* the case that reasons of the right kind are second-personal, then my Realist account of accountability will

¹⁶⁰ Darwall 2010a, p.264.

¹⁶¹ Darwall 2006, pp.311-312, as cited above.

¹⁶² Darwall 2006, p.15, as cited in my s.2.1, above.

¹⁶³ See s.1.2.2, above.

¹⁶⁴ Darwall 2006, p.15, as cited above.

¹⁶⁵ Darwall 2010a, p.260.

necessarily present a reason of the wrong kind for accountability, since Real reasons are obviously not second-personal. But, why should we accept that the only reasons of the right kind for accountability are second-personal? As I have shown, above, Real reasons meet Darwall's criterion of "justify[ing] the relevant attitude in its own terms",¹⁶⁶ so it is presumptuous to assert that the only reasons of the right kind for accountability are second-personal. I agree with Brewer, therefore, that "Darwall has operated with too restrictive an idea of what counts as the 'right kind of reason'"¹⁶⁷ in failing to consider the Realist alternative. Second-personality may well meet Darwall's criterion for a reason of the right kind in "justify[ing] the relevant attitude in its own terms"¹⁶⁸ – but, I have argued, it is not the only metaethic that can do so: Realism can too.

Finally, I raise some concerns as to whether it is the case at all that Darwall's own Second-Personal account meets the criterion of "justify[ing] the relevant attitude in its own terms"¹⁶⁹ by providing a reason of the right kind for holding agents accountable. The Realist account, I have argued (above), does meet this standard because the "reason" such an account yields for holding agents accountable is that those agents are indeed culpable, and the individuals holding those agents accountable know so and have the means of communicating this knowledge of culpability¹⁷⁰ – and this, I have argued, is (unlike social desirability) precisely a reason of the *right* kind as required by Darwall. That is, on the Realist account, the reason that individuals hold agents accountable is that they possess *epistemic authority* in their knowledge of the facts of wrongdoing. Darwall, of course, rejects such epistemic authority, claiming instead that the requisite authority must be "second-personal".¹⁷¹ This is the upshot of Darwall's ground-breaking Second-Personal account, which construes obligation not as a (Real) matter of the way the world is, but as a matter of a *relationship* between an agent and another individual. It follows on the Second-Personal account, therefore, that an individual must stand in the right kind of relationship to the agent, namely, one in which the individual has the second-personal authority to hold the agent accountable. The "reason" for which individuals hold agents accountable on the Second-Personal account, therefore, is that those individuals possess the *second-personal authority* to do so.

But, is second-personal authority a reason of the right kind? I briefly raise two considerations in support of the suspicion that it is not. Firstly, as we have seen, (social) *desirability* is a reason of the

¹⁶⁶ Darwall 2010a, p.264, as cited above.

¹⁶⁷ Brewer 2008, p.167.

¹⁶⁸ Darwall 2010a, p.264, as cited above.

¹⁶⁹ Darwall 2010a, p.264, as cited above.

¹⁷⁰ See s.1.2.2, above.

¹⁷¹ See, for instance, Darwall 2006, p.12. See also my Ch.2, s.1.4.

right kind only for *desiring*;¹⁷² *accountability*, I have argued, is a reason of the right kind for *holding accountable*. For what is second-personal *authority* a reason of the right kind? If the basic grammatical pattern is anything to go by, the answer appears to be: *authorizing* – but that is not the answer Darwall desires, which is, again: *holding accountable*. The first reason for suspecting that Darwall’s Second-Personal account does not itself provide a reason of the right kind for holding agents accountable, therefore, is that it’s not quite clear – at least, not by the obvious pattern followed by *desirability* and *desiring* – that second-personal authority is a “reason for” holding agents accountable at all. As Darwall stipulates, “To be a reason of the right kind, a consideration must justify the relevant attitude in its own terms”,¹⁷³ and it’s just not clear that second-personal *authority* justifies *holding accountable* “in its own terms”, in that way that *desirability* justifies *desiring*.

Moreover, I think there is reason to be suspicious of second-personal authority as a “reason for” holding agents accountable because of the way such a “reason” shifts focus away from the agent-being-held-accountable, and onto the powers of the individual-holding-accountable: namely, the individual’s second-personal authority. This is counterintuitive. Surely the only “reason for” holding an agent accountable is a reason *about the agent* – not about the individual holding the agent accountable. Foremost amongst such reasons about the agent for which an agent may be held accountable is, of course, the fact that the agent is *culpable* – and this is precisely what the Realist account captures (above). The Second-Personal account, on the other hand, suggests that the “reason for” holding agents accountable is second-personal authority, that is, the authority of the second person over the first – but this is a fact about the second person, the individual-holding-the-agent-accountable, rather than about the first-person agent-who-is-to-be-held-accountable. The reason proffered by the Second-Personal account for holding agents accountable is, counterintuitively, one about the individuals who hold agents accountable rather than about the agents themselves who are to be held accountable, and this is grounds for suspicion that this is not a reason of the right kind at all.

Therefore, where I have argued (above) that the Realist account can indeed provide a reason of the right kind for holding agents accountable, it is unclear to me that the Second-Personal account can achieve even this. Below, I continue the theme of argument that Realism actually outperforms the Second-Personal account in those arenas where Darwall has touted its success.

¹⁷² Darwall 2006, pp.311-312, as cited above.

¹⁷³ Darwall 2010a, p.264, as cited above.

2.2. Realist explanation

I have argued against Darwall's Second-Personal account on the grounds that his Mill's-Point argument for it fails, leaving us no reason to think that there are such things as second-personal reasons in the first place. And, I have (amongst many other arguments presented over the course of the preceding two chapters) argued that Darwall's Mill's-Point argument fails on the grounds that the argument has the form of an inference to the best explanation – yet Realism provides at least as good an explanation of Mill's Point as second-personality does.

However, one might wonder whether the suggestions of this chapter do indeed constitute a Realist *explanation* of the various moral phenomena under discussion. Why *is it*, on this Realist proposal, that (part of) what it is to be morally obligated is to be liable to being held accountable? If the answer is "Because the world *just is* such that (part of) what is to be morally obligated is to be liable to being held accountable",¹⁷⁴ then that doesn't seem anything like a satisfactory *explanation* of Mill's Point at all. Rather, it's simply a restatement of the same point.

Indeed, this complaint harkens back to Korsgaard's major objection to Realism, raised already in Chapter 1. In that context, the complaint was that Realism is unable to answer Korsgaard's "normative question", that is, why an agent ought to do what is morally required of him/her. In Korsgaard's words:

You ask the normative question: you want to know whether this terrible claim on you is justified. Is it really true that this is what you *must* do? The [R]ealist's answer to this question is simply 'Yes'. That is, *all* he can say is that it is *true* that this is what you ought to do.¹⁷⁵

And, as we noted there, this is no answer at all. Similarly, one might complain that the Realist "explanation" of Mill's Point – namely, that the world just is such that what is to be morally obligated is to be liable to being held accountable – is really no explanation at all.

I defend the Realist proposal against this possible objection, below.

2.2.1. Response

I will defend the Realist proposal against the objection that it provides no real, thorough explanation of Mill's Point¹⁷⁶ on three counts: suggesting that the (apparent) lack of Realist explanation is not necessarily a problem, proposing that, despite initial appearances, the Realist can indeed supply the requisite explanation, and arguing that Realism fares no worse than the Second-Personal account in this regard.

¹⁷⁴ See s.1.2, above.

¹⁷⁵ Korsgaard 1996, p.38, as cited in my Ch.1, s.2.2.

¹⁷⁶ See s.2.2, above.

As a first possible response, the Realist may simply shrug and retort that all explanations have to come to an end somewhere, and Realism's comes to an end at the point of the way the world is – where the world just is such that we are liable to being held accountable for acting in accordance with our moral obligations. This explanation may not go very deep – but that doesn't necessarily imply that it is *wrong*. Above, I compared Realist accounts of morality and accompanying phenomena to other facts about the world such as the negative charge of the electron.¹⁷⁷ If this comparison is correct, then asking why it is (according to the Realist) that (part of) what it is to be morally obligated is to be liable to being held accountable is like asking why it is that (part of) what it is to be an electron is to be negatively charged. That *just is* (part of) what it is to be an electron! Since nobody seems to be bothered by the lack of explanation as to why electrons are negatively charged, the Realist explanation that moral obligations *just are* those for which we are liable to being held accountable is, the Realist may argue, likewise unproblematic.

This retort, that all explanations must come to an end somewhere, may not be very compelling, however. But, I think the Realist case can be bolstered in at least two ways. Firstly, as I go on to argue, some deeper explanation than this is indeed available on certain Realist accounts. And, secondly, I will argue that the Second-Personal account does not fare much better than the bare assertion that “That's just the way things are”.

Firstly, what deeper explanation might the Realist offer of Mill's Point? Thaddeus Metz has recently put forward a Realist account of normativity¹⁷⁸ that might plausibly be taken to rebut Korsgaard's complaint that “*all* [the Realist] can say is that it is *true* that this is what you ought to do”¹⁷⁹ rather than actually *explaining* why we ought to do what we have moral reason to do. I will suggest that this account might easily be adapted to respond to the present, closely related criticism that Realism provides no genuine explanation of Mill's Point.

Metz draws his inspiration from the so-called ‘Cornell Realists’,¹⁸⁰ who draw a close analogy between Scientific Realism and Moral Realism. Consider, for instance, the identity claim that “Water is H₂O.” How might the Scientific Realist account for the universal, objective, and necessary truth of such a claim? One compelling answer appeals to synthetic, *a posteriori* necessities and a

¹⁷⁷ See s.1.2, above.

¹⁷⁸ Metz 2013, pp.91-93.

¹⁷⁹ Korsgaard 1996, p.38, as cited in s.2.2, above.

¹⁸⁰ Boyd 1988, Sturgeon 1988, Brink 1989, Gilbert 1990, Miller 1992, as referenced by Metz 2013, p.91, fn.10. Note that these Cornell Realists will generally be Consequentialists. (See my Ch.1, fn.32.) Metz, however, is unusual in that he is a Non-Consequentialist Realist. Nevertheless, Metz's argument, drawn as it is from Cornell Realists, will be applicable to my present point, which, following Darwall, concerns predominantly Consequentialist Realists (see my Ch.1, s.2.).

Kripkean/Putnamian causal theory of reference. That is, to claim that “Water is H₂O” is to claim that the two terms “water” and “H₂O” happen to pick out the very same thing in the world. This is an *a posteriori* claim, because it is one the truth of which we have discovered over time, through scientific investigation. It is also a necessary truth, because, once a “dubbing process” has taken place, the two terms are taken to “rigidly designate” the very same thing. And, since the sense of “water” (being a colourless, odourless, drinkable liquid, for instance) is different to that of “H₂O” (a particular molecular compound), “Water is H₂O” is a synthetic truth rather than an analytic one. Likewise, claim the Cornell Realists, with regards to moral truths. The claim that, say, “Wrongness is maximization of harm” is a synthetic, *a posteriori* necessity: to make such a statement is to claim that the two terms “wrongness” and “maximization of harm” pick out the self-same behaviour, and that this is something we have discovered by the *a posteriori* method of our metaethical theorizing; the claim is necessary once the two terms have come to pick out the very same thing as a result of a dubbing process; and the claim is synthetic, since we might just as well have dubbed “wrongness” to mean “degradation of persons”, for instance.¹⁸¹

Metz employs this same strategy to account for the normativity of Real moral reasons, claiming that “wrongness” and “overriding, categorical reason not to perform actions” have likewise been dubbed to pick out the self-same entity in the world. Just as the Scientific Realist is able to offer an explanation of why water is H₂O, therefore, the Moral Realist may likewise be able to explain to Korsgaard’s amoralist why he/she ought to do what he/she is morally obligated to do, that is, why moral obligations are normative.¹⁸²

And, of greater importance to our present purposes, the Moral Realist may likewise also be able to explain *Mill’s Point* with some plausibility: why it is true that “A moral obligation is one to which an obligated agent is liable to being held accountable.” That is, I suggest that the Moral Realist might, following Metz’s lead, extend the Cornell Realist strategy to the case of Mill’s Point, claiming plausibly that “moral obligation” and “that for which an agent is liable to being held accountable” co-refer in the same way that “water” and “H₂O” do. That is, the claim “Moral obligations are those

¹⁸¹ Metz 2013, pp.91-92.

¹⁸² I must add that I unfortunately do not think that Metz’s (2013) argument here is a successful response to Korsgaard’s (1996) criticism of Realism. Metz may well be able to account for normativity by claiming that “wrongness” and “overriding, categorical reason not to perform” co-refer, but it’s unclear to me that this is a successful response to the amoralist. Even if those two terms do co-refer, why should the amoralist accept in the first place that there are such things as overriding, categorical reasons not to perform certain actions? Likewise, the claim that “Unicorn horn is XYZ” may well ‘explain’ “unicorn horn” insofar as it informs us of the chemical composition thereof, but it still gives us no reason to accept that there are indeed such things as unicorn horns. Nevertheless, insofar as the present objection (see s.2.2, above) is not Korsgaard’s criticism *per se*, but rather the distinct (albeit related) question of whether Realism can explain *Mill’s Point*, this Cornell Realist/Metizian strategy does appear promising.

for which an agent is liable to being held accountable” might be held to be an *a posteriori* claim, one discovered over the course of many years of metaethical theorizing. And, following a dubbing ceremony, it is a necessary truth, since “moral obligation” and “that for which an agent is liable to being held accountable” rigidly designate the very same entities. Lastly, it is a synthetic truth that the two terms co-refer:¹⁸³ we might instead have taken “moral obligation” to pick out categorical norms,¹⁸⁴ or have taken it to pick out supreme, overriding reasons for action.¹⁸⁵

Therefore, following Metz’s lead, the Realist might plausibly proffer a competing Cornell Realist-style explanation of Mill’s Point than just the bare assertion, or apparent non-explanation, that the world simply is such that moral obligations are those for which agents are liable to being held accountable. Furthermore, I will now go on to argue, this is a depth of explanation of Mill’s Point unavailable to the Second-Personal theorist himself.

As the present objection questions whether Realism is able to genuinely *explain* Mill’s Point, we should ask whether Darwall’s Second-Personal account is able to do the same. Why is it, on the Second-Personal account, that moral obligations are those for which we are liable to being held accountable? Mill’s Point follows, on Darwall’s Second-Personal account, because of the nature of moral reasons, as second-personal.¹⁸⁶ A second-personal reason is one given to a first person by a morally authoritative second person.¹⁸⁷ That is, what it is to be subject to moral obligation, on such

¹⁸³ One possible objection at this point is that, unlike the claim “Water is H₂O”, “Moral obligation is that for which an agent is liable to being held accountable” is not synthetic; rather, it is just part of the *concept* of moral obligation that agents are liable to being held accountable for them (as the claim “All bachelors are unmarried men” is analytic, rather than synthetic, since it is just part of the concept of *bachelor* that such individuals are unmarried men).

I must admit that I am sympathetic to this objection. It does, too, appear to be Metz’s position, writing as he does that “The term ‘wrongness’ [...] is [...] associated with actions that *pro tanto* warrant guilt or blame” (2013, p.92). And, if it is correct that “Moral obligation is that for which an agent is liable to being held accountable” is a conceptual, analytic truth, then the analogy with the Scientific Realist’s account of “Water is H₂O” breaks down.

However, if “Moral obligation is that for which an agent is liable to being held accountable” is a conceptual, analytic truth, then the dissolution of the analogy with the water/ H₂O case is loss no for Moral Realism’s cause. The objection presently under consideration, recall, is that Moral Realism is unable to account for Mill’s Point, namely, that we are liable to being held accountable for our moral obligations (see s.2.2, above). But, if it turns out that “Moral obligation is that for which an agent is liable to being held accountable” is just a conceptual truth, then of course Moral Realism cannot be blamed for its failure to explain such a truth. As a conceptual truth, there *is* no explanation of it available, other than to simply point to the concepts involved. (Likewise, can any theory be blamed for failing to account for the truth of “All bachelors are unmarried men”?) In other words, if the suggested Realist account of Mill’s Point (above) is wrong, because “Moral obligation is that for which an agent is liable to being held accountable” is actually a conceptual, analytic truth, then the anti-Realist objection of s.2.2, above, should not arise in the first place.

¹⁸⁴ Kant 1785/1996b, as referenced by Darwall 2006, p.26. See the introduction to my Ch.3.

¹⁸⁵ Scheffler 1992, and Brink 1997, p.255, as referenced by Darwall 2006, p.26. See the introduction to my Ch.3.

¹⁸⁶ See my Ch.3.

¹⁸⁷ See Ch.1, s.1.2.

an account, is to be under the authority of a second person. What authority is this? The authority to hold agents accountable. A second-personal reason is a reason given to a first person by a second person with the moral authority to hold the first person accountable. It follows, therefore, on this Second-Personal analysis, that a moral obligation is one for which agents are liable to being held accountable.

But, is this Second-Personal account of Mill's Point really any better an *explanation* than the Realist alternative? On the purportedly unacceptably shallow Realist account of Mill's Point,¹⁸⁸ the Realist looks to his/her account of the nature of moral reasons in order to explain Mill's Point: that is, locating Real moral value in the world, the Realist explains Mill's Point by looking to the world, claiming that the world is such that moral obligation is that for which agents are liable to being held accountable. Yet, the form of this Realist response is identical to that of the Second-Personal theorist. The Second-Personal theorist explains (purports to explain) Mill's Point by looking to his/her account of the nature of moral reasons, as second-personal, and claiming that second-personal authority is such that we may justifiably hold others accountable to the obligations they owe to us. If the Realist 'explanation' of Mill's Point is unacceptable, therefore, surely the Second-Personal correlate is too. What, we should ask, does this Second-Personal account really *explain*? It aims to explain Mill's Point by looking to its account of the nature of moral reasons, as second-personal, where second-personal authority is such that we may justifiably hold others accountable to the obligations they owe to us – but why, as I have been at pains to argue throughout this project, should we accept that there is indeed such a thing as "second-personal authority"?

Perhaps this point is better made by way of the following contrast. On the Realist proposal I have presented in this chapter, an individual has the authority to hold an agent accountable when that individual possesses the requisite *epistemic authority*, that is, when that individual has knowledge of the facts of wrongdoing (and the means of communicating them to the agent).¹⁸⁹ The Realist explanation of an individual's authority to hold an agent accountable, therefore, is that that individual knows what the agent did, and is in a position to bring the agent to a state of conviction. Now, contrast this with the Second-Personal account of an individual's authority to hold an agent accountable. Darwall vehemently denies that such authority is epistemic, insisting instead that it is "second-personal".¹⁹⁰ But, what is this *second-personal authority*? As Darwall writes with reference to human rights:

¹⁸⁸ See s.2.2, above.

¹⁸⁹ See s.1.2.2, above.

¹⁹⁰ See, for instance, Darwall 2006, p.12. See also my Ch.2, s.1.4.

It includes a second-personal authority to resist, complain, remonstrate, and perhaps use coercive measures of other kinds, including, perhaps, to gain compensation if the right is violated.¹⁹¹

And, note that “to resist, complain, remonstrate,” etc., is just to hold an agent accountable, in various ways. So, second-personal authority is just the authority to hold agents accountable. That is, whereas the Realist explanation of an individual’s authority to hold an agent accountable is that that individual has epistemic authority – that is, knows what the agent did, and is in a position to bring the agent to a state of conviction – the Second-Personal explanation of an individual’s authority to hold an agent accountable is that that individual has second-personal authority – *which is just the authority to hold agents accountable*. On the Second-Personal account, therefore, an individual has the authority to hold others accountable when he/she has the authority to hold agents accountable. This surely falls short of being a robust *explanation* of the order that has been demanded of the Realist. In another context, Brewer puts this point plainly, claiming that Darwall’s Second-Personal account

boil[s] down to the unhelpful insistence that moral obligations are required because people can legitimately require one to perform them, and that one must act morally because other people can legitimately demand that one do so.¹⁹²

This is, of course, just a by-product of Darwall’s founding his Second-Personal account, as he explicitly does, on a “circle of irreducibly second-personal concepts”,¹⁹³ but this circle arguably provides no deeper explanation than the Realist alternative. If Realism is subject to the criticism that it fails to provide a genuine explanation of Mill’s Point, therefore, the Second-Personal account is too.

I conclude, therefore, that, once again, Realism is in fact in at least as good as a – if not a better – position than Darwall’s Second-Personal account. Whereas the initial criticism was that Realism is unable to offer a proper, robust explanation of Mill’s Point,¹⁹⁴ it turns out that Realism is indeed able to offer a detailed explanation of the co-reference of “moral obligation” and “that for which agents are liable to being held accountable” in the style of the Cornell Realists. Furthermore, it turns out that the Second-Personal account is itself guilty of the charge initially laid against the Realist, since the Second-Personal account of an individual’s authority to hold agents account, in Brewer’s words, “boil[s] down to the unhelpful insistence” that an individual has the second-personal authority to hold agents accountable.¹⁹⁵

¹⁹¹ Darwall 2006, p.18.

¹⁹² Brewer 2008, p.167.

¹⁹³ Darwall 2006, p.11. See my Ch.2, s.2.3.2.

¹⁹⁴ See s.2.2, above.

¹⁹⁵ Brewer 2008, p.167, as cited above.

3. Conclusion

Mill's Point is that there is a conceptual connection between obligation and accountability.

According to Darwall, this accountability-obligation connection is best explained by the second-personal nature of obligation: we are held accountable *by* others because we are obligated *to* those others.¹⁹⁶ This demonstrates, therefore, that there are such things as second-personal reasons. I have argued, however, that Darwall's Second-Personal explanation of Mill's Point struggles to account for its broader aspects, namely, that we are liable to being held accountable not only by the victims of our misdeeds but also by ourselves, as well as by members of the general moral community.¹⁹⁷ If moral reasons exist in the relationship between a first and second person, as the Second-Personal account holds, it will have trouble explaining why a first person or a third person may equally have the authority to hold the first person accountable to the reason the second person gave him/her.

Realism, on the other hand, is easily able to account for Mill's Point, including all its broader aspects: an individual – *any* individual, be it the victim, or the agent him-/herself, or a member of the general moral community – has the authority to hold an agent accountable when that individual is in a position to do so.¹⁹⁸ Since Realism can explain Mill's Point just as well as – if not better than – the Second-Personal account, Darwall Mill's-Point argument for second-personality, as an inference to the best explanation, fails. Again, we are left without any grounds to accept that such things as second-personal reasons exist in the first place.

Moreover, Realism is able to account for various other aspects of morality, many of which – like Mill's Point – Darwall frequently draws upon as support for his Second-Personal account. That is, Realism is able to offer an account of the validity of moral demands,¹⁹⁹ of Pufendorf's Point,²⁰⁰ and of the victim's special authority to forgive,²⁰¹ and is able to provide a reason of the right kind for holding agents accountable,²⁰² as well as providing a thorough, deep account of why moral obligations are those for which agents are liable to being held accountable.²⁰³ This leaves Darwall's Second-Personal account in a very weak position.

¹⁹⁶ See the introduction to Ch.3.

¹⁹⁷ See my Ch.3 and 4.

¹⁹⁸ See s.1.2.2, above.

¹⁹⁹ See s.1.1, above.

²⁰⁰ See s.1.2.3, above.

²⁰¹ See s.1.2.4, above.

²⁰² See s.2.1, above.

²⁰³ See s.2.2.1, above.

4. Summary

Darwall's Mill's-Point argument infers from the nature of accountability to the nature of reasons as second-personal. I have argued that this inference fails. Interestingly, in a more recent work, Darwall reverses the direction of this original Mill's-Point argument, arguing from second-personal authority to the second-personality of accountability:

If one person has practical authority with respect to another, then this would seem to mean not just that the latter has a reason of whatever priority or weight for acting as the former directs but also that the latter has some responsibility *to* the former for doing so, that the latter is, in some way or other, answerable to the former.²⁰⁴

But, this new direction of argumentation is no more promising than the former. This argument takes for granted that there is such a thing as second-personal authority – that is, that a second person has the power to obligate a first just in virtue of commanding the first person to take some action – but we have as yet no grounds for supposing that a second person does indeed have such power over the first.²⁰⁵ I'm sure that Darwall thinks that the direction of the Mill's-Point argument is reversible in this way because it forms part of the "interdefinable circle"²⁰⁶ of Second-Personal concepts.²⁰⁷ But, again, we have as yet been given no way to break into this circle – at least, no way of doing so that does not in one way or another reduce second-personality to the non-second-personal.

Do second-personal reasons exist? Recall that, unless we have grounds for believing that there are such things as second-personal reasons, Darwall's argument for Morality as Equal Accountability does not go through. That argument, recall, is that

any second-personal claim or "summons" (*Aufforderung*) presupposes a common competence, authority, and, therefore, responsibility as free and rational, a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally.²⁰⁸

In order for this Fichtean argument to produce its conclusion, it needs to first establish that there is indeed such a thing as a second-personal claim or "summons".²⁰⁹ To that end, I have examined two possible arguments for the existence of second-personal reasons. The first was the argument from intuition,²¹⁰ which established that there is some intuitive support for second-personal reasons – although it is limited by the intuitive requirement that any given reason must be a *valid* one. However, the most plausible, non-circular account Darwall could provide of validity (*viz.* grounding in

²⁰⁴ Darwall 2010a, p.266.

²⁰⁵ See Ch.2, s.2.3.

²⁰⁶ Darwall 2006, p.12.

²⁰⁷ See my Ch.2, s.2.3.2.

²⁰⁸ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000, as cited in my Ch.1, s.6.2.1.

²⁰⁹ See Ch.2, s.2.1.

²¹⁰ See Ch.2, s.2.2.

the Second-Personal standpoint, understood in Contractualist terms²¹¹) either bottomed out as non-second-personal (first-personal)²¹² or proved to be circular, by presupposing equal dignity.²¹³ Contractualist equal dignity can't be presupposed because equal dignity is precisely what the argument for Morality as Equal Accountability is supposed to *prove*. So, the intuitive argument for second-personal reasons is inconclusive.

The second argument examined in support of the existence of second-personal reasons was Darwall's Mill's-Point argument, according to which the second-personal nature of moral reasons can be inferred on the basis of the fact that we are held accountable by others. However, as I hope to have shown over the course of this present chapter and the preceding two,²¹⁴ Darwall's Mill's-Point argument fails too. Realism provides a better explanation of Mill's Point and all its broader aspects at least as good as – if not better than – Darwall's Second-Personal account; as such, Darwall's inference to second-personality is unfounded.²¹⁵

As such, we are left without any reason to suppose that there are indeed such things as second-personal reasons, and so we have no reason to conclude equal dignity. We are back at square-one with the argument for Morality as Equal Accountability: it doesn't even get off the ground. And, without Morality as Equal Accountability's conclusion of equal dignity, moral reasons on this Second-Personal account will not be *universal*: it will not necessarily follow that you owe the same obligation to me not to, for instance, cause me undue harm as I owe it to you. So, the Second-Personal account fails as an account of moral reasons, which have universal normative force.²¹⁶

Moreover, even if we were able to establish that second-personal reasons do actually exist, recall that I have also argued that Darwall's argument for Morality as Equal Accountability is a *non-sequitur*.²¹⁷ That is, all that is entailed by Pufendorf's Point is that the first and second persons have authority *over the first*, and not that that first person enjoys equal, reciprocal authority over the second. So, even if Darwall were able to prove that second-personal reasons do indeed exist, he would still face the battle of subsequently demonstrating that such reasons are universal and that all agents enjoy equal, second-personal authority over all other agents.

In the final chapter, I go on to examine one last argument for the existence of second-personal reasons. If successful this argument will bypass the need for the argument for Morality as Equal

²¹¹ See Ch.2, s.2.3.3.1.

²¹² See Ch.2, s.2.3.3.2.1.

²¹³ See Ch.2, s.2.3.3.2.2.

²¹⁴ See Ch.3 and 4.

²¹⁵ See s.3, above.

²¹⁶ See Ch.1, s.1.3.

²¹⁷ See Ch.2, s.1.

Accountability in order to establish the universality of moral reasons. First, I lay the groundwork for this more radical argument for second-personality, which is the task of the following chapter.

6.

Does Moral Philosophy rest on a category-mistake?

This chapter and the next will represent a fairly radical shift from preceding chapters. In particular, the objective of this chapter and the next – unlike that of the previous four chapters – will be to *defend* (something like) Darwall’s Second-Personal account. And, the means to this end will constitute a further difference about this chapter, which is that its content will draw heavily on the Philosophy of Mind, rather than just on Metaethics. The aim of this chapter is to lay the groundwork for the defence of a particular reading of *The Second-Person Standpoint*, what I will call the ‘Ryle-inspired’ reading thereof. This Ryle-inspired reading of *The Second-Person Standpoint* will be the subject of the following chapter, and the groundwork for it laid down in this chapter will be modelled on Gilbert Ryle’s argument for his own philosophy of mind. As Ryle’s argument for his philosophy of mind followed on the heels of his famous critique of Cartesian Dualism, so will my argument for the Ryle-inspired reading of *The Second-Person Standpoint* follow on the heels of a critique of Realism. The bulk of this chapter will be spent in defence of the claim that Realism is unacceptable insofar as it commits what Ryle has called a “category-mistake”.

However, I have, of course, just spent a full chapter arguing *for* Realism, insofar as Real, world-regarding moral facts provide a better explanation of various moral phenomena (in particular, Mill’s Point) than do the bipolar, person-regarding reasons posited by Darwall’s Second-Personal account.¹ How, then, could I now go on to argue *against* Realism? What I aim to do over the course of this chapter and the next is to make a case for the claim that the strategy of the preceding chapters depends on a particular reading of *The Second-Person Standpoint*, which – although it is the natural, textually well-supported reading – may not be the most charitable reading. That is, the strategy of the preceding chapters depends on reading Darwall as presenting the Second-Personal account as *an alternative to* Realism, as one metaethical amongst others, such that, if it were to be shown that Realism provides a better account of the nature of moral reasons than does Darwall’s Second-Personal account (as I have tried to show previously²), then Darwall’s Second-Personal account must be rejected in favour of Realism. Indeed, this is precisely the reading of *The Second-Person*

¹ See Ch.5.

² See Ch.5.

Standpoint encouraged by Darwall's own set-up of the project in its very first pages, where he presents the reader with two possible responses to the foot-treader: one in which you draw his attention to a Real, world-regarding reason to remove his foot from on top of your, and one in which you give him a second-personal, person-regarding reason to do so.³

Despite this being the way Darwall sets up his own Second-Personal account, as an alternative to Realism, I think that there is a another reading of it available, a reading which draws its inspiration from Ryle. Ryle famously rejected Cartesian Dualism – but not on the grounds that it identifies the nature of mental states incorrectly, namely, as modes of mental substance. Rather, Ryle rejected Cartesian Dualism – as well as Idealism and Materialism⁴ – on the grounds that it tries to identify mental states as any kind of substance at all – be it mental or physical substance. That is, Ryle's criticism of Cartesian Dualism is not that it identifies mental states as modes of *mental* substance, but is the deeper objection that it identifies mental states as modes of any *substance* at all. Ryle doesn't reject Cartesian Dualism on the grounds that it's the less preferable option to an alternative account of the nature of mental states (such as Physicalism, for example), but on the grounds that such approaches – ones that aim to determine the true nature of mental states – are fundamentally malconceived; they are category-mistaken. Consequently, Ryle's own philosophy of mind is to be preferred not because it presents the better alternative to Cartesian Dualism, in the sense that it presents the better account of the nature of mental states; rather, it is to be preferred because it presents the category-*unmistaken* conception of the mind.

Likewise, I will argue that the Second-Personal account is best read not as presenting an alternative account of the nature of moral reasons to that presented by Realism. (Indeed, when we read *The Second-Person Standpoint* that way, as we have done for the preceding five chapters, then, as I have argued,⁵ *Realism* is in fact the preferable metaethic.) Instead, I will argue with reference to Realism, such approaches are category-mistaken. Rather, on a Ryle-inspired reading of *The Second-Person Standpoint*, the Second-Personal account is to be preferred because it presents the category-*unmistaken* conception of morality. (What exactly this "category-*unmistaken*" conception of morality might be will be the subject of the following chapter.)

On this reading of *The Second-Person Standpoint*, the Second-Personal account is not just one metaethic amongst others, such as Realism (and Neo-Kantianism, and Hobbesian Egoism, and Divine Command theory, as I originally presented it⁶). The Second-Personal account is not an alternative to

³ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁴ Ryle 1949/2009, p.12.

⁵ See Ch.5.

⁶ See Ch.1.

all those other metaethics in the sense that it offers a rival account of the nature of moral reasons. Instead, on this Ryle-inspired reading, the Second-Personal account contests those other metaethics' very conception of the subject matter, offering a deep, radical challenge to traditional, category-mistaken Metaethics. This gives substance to the high praise Darwall has received for *The Second-Person Standpoint*, such as that by Fleischacker:

[*The Second-Person Standpoint*] offers a thorough, intriguing and deeply thoughtful reconception of the entire sphere of morality. And what it has to say is quite new. More than any recent book, [...] Darwall's book challenges the [metaethical] presuppositions of practically every going moral theory.⁷

My claim of this chapter and the next will be that, on a Ryle-inspired reading, "Darwall's book challenges the [metaethical] presuppositions of practically every going moral theory" in that it challenges their *category-mistaken* foundations.

This is not, however, a claim that receives overwhelming support from Darwall's own text; Darwall certainly doesn't explicitly charge Realism (let alone "practically every going moral theory"⁸) with a category-mistake, as I will do. As such, I can't be sure if Darwall himself would consent to this Ryle-inspired presentation of the Second-Personal approach. As I will go on to describe in the following chapter, some portions of his text do support the more radical, Ryle-inspired approach that I will be taking in this chapter and the next. On the other hand, many other portions of *The Second-Person Standpoint* – such as Darwall's own presentation of it in the opening pages of the book as an alternative to Realism⁹ – support the reading I have assumed in preceding chapters. In many ways, I think *The Second-Person Standpoint* is something like a snapshot of a butterfly emerging from a chrysalis: halfway to being something new and different, but still with a foot in the old way – not quite fully worked-out. Lavin sees something of a similar struggle in *The Second-Person Standpoint*; as he writes: "in this new book Darwall turns against his Kantian roots," but "[o]f course, one can never entirely break with one's past: it hangs on in Darwall's sense of the questions and methods of moral philosophy".¹⁰ This chapter, I hope, will help work out the Second-Personal approach all the way to full emergence – but it will require breaking those few old strands that hold it back. I am not sure, however, whether Darwall will appreciate this new approach (or the butterfly metaphor).

Whether or not this is a reading of *The Second-Person Standpoint* that Darwall would accept, I am at least not the only one to whom such ideas have occurred. At least one reviewer of Darwall, Brewer, reads *The Second-Person Standpoint* as responding to a category-mistake, writing: "Darwall

⁷ Fleischacker 2009, p.117, as cited in the introduction to this dissertation.

⁸ Fleischacker 2009, p.117, as cited above.

⁹ Darwall 2006, pp.5-7, as referenced above.

¹⁰ Lavin 2008.

maintains, it is a category mistake to” suppose that “[a] moral obligation is [...] merely an act that it would be a very good idea to do”, when, rather “it is an act which one owes it to someone or other to do”.¹¹ However, Brewer makes this remark only in passing, and unfortunately offers no substantive argumentation for the claim. I will try to supply that argumentation in this chapter and the next.

Similar ideas seem also to have motivated H.A. Prichard. In his well-known paper, ‘Does Moral Philosophy rest on a mistake?’, Prichard wishes to mount a deep, fundamental challenge similar to the one I do, in arguing that, yes, Metaethics in general does “rest on a mistake”.¹² Even though Prichard would not have had the terminology of either the ‘category-mistake’ or ‘second-personality’ available to him, he did accuse Moral Philosophy of “consist[ing] in the attempt to answer an improper question”¹³ – where, interestingly, the charge of the “improper question” is one generally attributed to Ryle¹⁴ (who published more than thirty years after Prichard).

Below, I contrast my present approach with Prichard’s argument for the claim that Moral Philosophy “rest[s] on a mistake”.

1. ‘Does Moral Philosophy rest on a mistake?’

The title of this chapter is adapted from that of H.A. Prichard’s influential paper, ‘Does Moral Philosophy rest on a mistake?’¹⁵ In this paper, Prichard argued that the two old rivals, Utilitarianism, on the one hand, and Kantianism, on the other, each suffer such serious flaws that it points to a common, fundamental error – leading Prichard to conclude that Moral Philosophy does indeed “rest on a mistake”. In suggesting that a Ryle-inspired reading of the Second-Personal account is to be preferred because such an account would avoid the error of the rest of Metaethics, namely, a category-mistaken conception of morality,¹⁶ I will be arguing something similar to Prichard. Interestingly, in an interview conducted after the publication of *The Second-Personal Standpoint*, Darwall reveals that he sees the same affinity I do between Prichard’s paper and the evolution of Darwall’s own view:

What I came to see was that there is a whole class of reasons that are second-personal [...]. I’ve come to think that the old point that Prichard made in “Does Moral Philosophy Rest on a Mistake?” is just an instance of [this] point. Prichard said it’s a mistake to try to derive reasons for moral obligations or reasons that could establish moral obligation from self-

¹¹ Brewer 2008, p.163.

¹² Prichard 1912.

¹³ Prichard 1912/1949, p.2.

¹⁴ See, for instance, Ryle 1949/2009, p.12.

¹⁵ Prichard 1912.

¹⁶ See the introduction, above.

interested considerations. [...] I think that's right, but I think that the reason that's right is because moral obligations are themselves irreducibly second-personal.¹⁷

Despite these similarities, however, there will also be notable differences between my argument and Prichard's, which I detail, below.

One major difference between my argument and Prichard's will be that, on the basis of his paper, 'Does Moral Philosophy rest on a mistake?', Prichard concludes that (a version of) Realism is the correct metaethic. I will be arguing, on the other hand, for a Ryle-inspired reading of the Second-Personal account and *against* Realism. This difference in our respective conclusions is due to the difference in our respective diagnoses of the particular error involved. Whereas I will be arguing that the relevant error is a category-mistake¹⁸ (hence, my slight alteration of Prichard's original title, to read: Does Moral Philosophy rest on a *category*-mistake?), according to Prichard, the mistake upon which Moral Philosophy rests is one of derivation: trying to ground moral reasons in something else, such as the Utilitarian's trying to ground moral reasons in 'utility', and the Kantian's trying to ground them in duty.¹⁹ Given this diagnosis of Moral Philosophy's error, Prichard goes on to conclude non-Utilitarian, non-naturalist Realism: since the failures of Utilitarianism and Kantianism demonstrate that it is a mistake to try to ground moral reasons in something else, the correct metaethic must be one that does *not* attempt to ground moral reasons in anything and instead simply accepts that such things exist in and of themselves – that is, according to Prichard, the correct metaethic must be (non-Utilitarian) Realism. This is interesting, considering that Darwall's prime adversary in *The Second-Person Standpoint* is Moore,²⁰ a non-naturalist Realist like Prichard. Unlike Prichard, and like Darwall, I will in this chapter be arguing against Realism, and I will do so on the basis of the charge not of illicitly attempting to ground moral reasons in something extraneous, but of committing a category-mistake. Even if Prichardian Realism avoids the error of attempting to ground moral reasons in something extraneous, I will argue that it nevertheless succumbs to the Ryle-inspired charge of category-mistake.

This is, however, one further respect in which this chapter will – unfortunately – resemble Prichard's paper, and that is in the speculative nature of its conclusion. As I have noted, Prichard critiques only Utilitarianism and Kantianism, yet he goes on to draw the very broad conclusion that all of "Moral Philosophy rest[s] on a mistake".²¹ On the basis of his explicit criticisms of Utilitarianism and Kantianism, the reader can project the nature of Prichard's unspoken criticisms of all the other

¹⁷ Darwall 2009, pp.122-123. Brewer (2008, p.167) suggests a similar reading of Darwall.

¹⁸ See the introduction, above.

¹⁹ Prichard 1912/1949, pp.3-4.

²⁰ Darwall 2006, throughout: pp.5, 28, 37, 158 (fn.17), 217 (fn.5), 221, 226–227, 231, 280, 282–283, and 286.

²¹ Prichard 1912.

metaethics that would comprise traditional Moral Philosophy, but the paper nevertheless falls short of actually demonstrating that those remaining metaethics likewise “rest on a mistake”. Unfortunately, this present chapter will also fall short in this way: I will argue explicitly for the category-mistakenness of Realism only, although I would wish to claim that all of traditional Moral Philosophy is guilty of a category mistake. Here I again follow Ryle: he asserts that all traditional philosophies of mind – Cartesian Dualism, Idealism and Materialism included²² – commit a category-mistake, although he offers explicit argumentation for the category-mistakenness of Cartesian Dualism only. So, the arguments of this chapter unfortunately will not amount to a comprehensive demonstration of the category-mistakenness of all traditional Moral Philosophy, but at least here I am in good company. Although the argumentation of this chapter will be less than conclusive, I trust that, like my predecessors’, it will at least be novel and provocative. Furthermore, I remind the reader that the aim of this chapter is not so much to critique Realism (nor traditional Moral Philosophy generally), but instead to lay the foundation for an alternative, positive account of the Second-Personal approach, which will be the subject of the following chapter.

The rest of this chapter will be concerned with defending the claim that Realism is guilty of a category-mistake. We turn, first, to Ryle’s account of what a category-mistake is.

2. Category-mistakes

It strikes me that there are several points of analogy between the Philosophy of Mind and Metaethics, or, as Prichard calls it, Moral Philosophy. The respective debates are not exactly isomorphic, but there are sufficiently many similarities to suggest that the one can learn from the other – in particular, that Metaethics can learn from the Philosophy of Mind.²³ On the one hand, this is unsurprising: if the Philosophy of Mind is concerned with the nature of mental states, and if moral reasons are mind-dependent, then of course the Philosophy of Mind will be relevant to Metaethics. But, notice that this explanation of the connection precludes mind-*independent* reasons – that is, world-regarding, or Real, reasons. Perhaps surprisingly, then, my argument will be that the metaethic of Realism is analogous to the philosophy of mind that is Cartesian Dualism. They are analogous in their dualisms, their epistemologies, and their category-mistakenness. As Ryle critiques Cartesian Dualism of committing a category-mistake, so will I critique Realism. Furthermore, as Ryle uses the charge of Cartesian Dualism’s category-mistakenness to build his own, radically different philosophy of mind, so I will use my charge of Realism’s category-mistakenness to build support for

²² Ryle 1949/2009, p.12.

²³ The metaethicist Russ Shafer-Landau (2003, p.106ff) makes use of a similar strategy in employing Jaegwon Kim’s (1993, 1996, 1998) argument against non-reductive Physicalism in the Philosophy of Mind.

the Ryle-inspired reading of *The Second-Person Standpoint*. This is my primary objective here: not so much to critique Realism, but to lay the groundwork for a radically different, category-unmistaken reading of *The Second-Person Standpoint* (which will be the subject of the following chapter).

So, firstly, what is a category-mistake? According to Ryle, a category-mistake is made when one treats something “as if [it] belonged to one logical type or category (or range of types or categories), when [it] actually belong[s] to another.”²⁴ Ryle gives a whole host of illustrations of category-mistakes, some of which are more useful than others, including the examples involving: the university,²⁵ the marching division,²⁶ team-spirit,²⁷ the British Constitution,²⁸ the average taxpayer,²⁹ gloves,³⁰ and jokes.³¹ For the purposes of illustration, I will recap just one of Ryle’s examples of a category-mistake here – my favourite:

A foreigner watching his first game of cricket learns what are the functions of the bowlers, the batsmen, the fielders, the umpires and the scorers. He then says 'But there is no one left on the field to contribute the famous element of team-spirit. I see who does the bowling, the batting and the wicket-keeping; but I do not see whose role it is to exercise *esprit de corps*.'³²

The foreigner from Ryle’s illustration has committed a category-mistake in that, in asking who on the field is responsible for team-spirit, he has mistakenly assumed team-spirit to be of the same logical type as batting, bowling, wicket-keeping, *etc.*, when of course it is not; team-spirit is a different kind of thing altogether. Team-spirit is not an activity performed by a player or players on the field (such as batting or bowling); rather, it is just the perfectly ordinary tendency to “[perform] a task keenly”³³ – that is, it is a *way of* performing cricketing activities, such as batting or bowling. It is not another entry in the list of activities performed by cricketers, and to think so is to commit a category-mistake. Furthermore, when we don’t find team-spirit where we were mistakenly looking for it – on the field – the temptation is to conclude that it must, therefore, be a *special* kind of activity: not a physical, bodily activity like batting or bowling, but an activity of the ‘spirit’ – team-*spirit*. Rather, Ryle contends, the proper response is to realize that team-spirit is not an activity at all – neither a mundane one nor a special one – but just a way of performing cricketing activities.

²⁴ Ryle 1949/2009, p.6.

²⁵ Ryle 1949/2009, p.6.

²⁶ Ryle 1949/2009, p.6.

²⁷ Ryle 1949/2009, pp.6-7.

²⁸ Ryle 1949/2009, p.7.

²⁹ Ryle 1949/2009, p.7.

³⁰ Ryle 1949/2009, p.11.

³¹ Ryle 1949/2009, p.11.

³² Ryle 1949/2009, pp.6-7.

³³ Ryle 1949/2009, p.7.

Dr Jack Ritchie has pointed out to me that this is a particularly weak definition of ‘team-spirit’. I agree. However, it is good enough to do the job of distinguishing team-spirit as a way of performing an activity rather than as another kind of activity.

Ryle thinks that the Cartesian Dualist, like the foreigner from the team-spirit example, commits a category-mistake: as the foreigner makes a category-mistake in thinking that team-spirit is one of the same kind of thing as bowling and batting, the Cartesian Dualist makes a category-mistake in thinking that the mind is one of the same kind of thing as the body, namely, another kind of *substance*. And, when the Cartesian Dualist discovers that he cannot locate the mind as he can the body, he concludes that the mind must be of a special kind of substance, one not located in space: *mental* substance. As Ryle describes the Cartesian Dualist's category-mistake:

The differences between the physical and the mental were thus represented as differences inside the common framework of the categories of 'thing,' 'stuff,' 'attribute,' 'state,' 'process,' 'change,' 'cause' and 'effect'. Minds are things, but different sorts of things from bodies; mental processes are causes and effects, but different sorts of causes and effects from bodily movements. And so on.³⁴

The Cartesian Dualist's mistake, then, is to treat the mind as a *thing*, a substance – and a *causal* substance at that – when the mind is actually neither thing, substance, nor cause. Likewise, I will try to argue that the mistake on which traditional Moral Philosophy rests is representing moral reasons “inside the common framework of the categories of 'thing,' 'stuff,' 'attribute,' 'state,' 'process,' 'change,' 'cause' and 'effect'”, when really they are “different sorts of things” altogether.

Why it is that Ryle thinks Cartesian Dualism is guilty of a category-mistake is a question I will return to shortly. First, I try to establish the analogy between Cartesian Dualism and Realism. Then, once I explain why Ryle thinks that Cartesian Dualism is guilty of a category-mistake, the same reasoning should carry across to Realism.

3. Realism and Cartesian Dualism

The task of this section is to establish an analogy between Cartesian Dualism and Realism. On some scores, the analogy is most easily drawn between Cartesian Dualism and those Realists of the non-naturalist, intuitionist variety in particular. As I have noted, above, Prichard is one such Realist,³⁵ as is Darwall's prime adversary,³⁶ Moore.³⁷ I am of the opinion, though, that the analogy extends to other forms of Realism too. Like J.L. Mackie, I take it to be true that

the central thesis of intuitionism is one to which any objectivist [i.e. Realist] view of values is in the end committed: intuitionism merely makes unpalatably plain what other forms of objectivism wrap up.³⁸

³⁴ Ryle 1949/2009, p.9.

³⁵ See s.1, above.

³⁶ Moore 1903, as referenced by Darwall 2006, throughout: pp.5, 28, 37, 158 (fn.17), 217 (fn.5), 221, 226–227, 231, 280, 282–283, and 286, as referenced in my Ch.1, s.2.

³⁷ See Moore (1903, Preface) for an explanation of his particular variety of 'Intuitionism'.

³⁸ Mackie 1977, p.38.

To this end, I take care to also appeal to those Realists of the naturalist, non-intuitionist bent where appropriate. Even if the broader analogy between Cartesian Dualism and Realism is unsuccessful, however, this will not undermine the charge that Realism commits a category-mistake. The form of that argument (for the conclusion that Realism commits a category-mistake) is not an argument by analogy, and so its success does not depend on the success of the analogy between Cartesian Dualism and Realism. The purpose of the analogy to follow is not so much to build a case for the category-mistakenness of Realism, but instead to assist in making clear what is otherwise a difficult, subtle point. The actual argument for Realism's category-mistakenness occurs in section 4.2, below. First, I offer the reader an easier way into the argument, by analogy with Ryle's better-known critique of Cartesian Dualism.

3.1. Dualisms

The first similarity to note between Realism and Cartesian Dualism is that both posit metaphysical dualisms. This is obvious in the case of Cartesian Dualism, according to which there are two kinds of substances: mind, and matter. That Realism also posits a dualism may be less clear – however, I think Realism may posit dualism even in more than one way.

The first kind of Realist dualism is clearly a supposition of those non-naturalists such as Prichard and Moore. According to such theorists, there exist *sui generis* moral facts, facts about the world – like facts about sticks and stones – but special facts that possess normative force in and of themselves. That is, such theorists posit a dualism of facts: ordinary, descriptive facts on the one hand, and *sui generis*, moral, normative facts on the other.

Realists of the naturalist variety, however, will of course deny that there are any such things as *sui generis* moral facts. But, I argue, they are nevertheless committed to essentially the same dualism: on the one hand, non-moral facts like facts about stick and stones, and on the other, facts which are not in and of themselves normative, but which are of moral import – like facts about the maximization of welfare, for instance. If this naturalistic dualism of facts does not convince, however, naturalist (as well as non-naturalist) Realists will nevertheless be committed to the following dualism, a dualism of reasons.

Whereas the above dualism contrasts moral reasons with non-*reasons* (descriptive facts about the world), this second dualism contrasts moral reasons with non-*moral* reasons. Both moral and non-moral reasons are reasons in the sense that they tell the agent how he/she ought to behave, but their respective sources are very different on a Realist view. This is effectively the major argument

in favour of Realism that we considered at the beginning of this project:³⁹ moral reasons are universal; non-moral reasons are not; so, moral reasons must be a different kind of thing to non-moral reasons. Where non-moral reasons, such as the reason to complete my dissertation, are person-regarding, moral reasons must be world-regarding. Since the world is objective, and a world we all share, it follows, therefore, that moral reasons will be universal. Moral reasons are, like non-moral reasons, *reasons* nonetheless, but they are a different kind of reason – just as the Cartesian Dualist posits that the mind is a substance, only a different kind of substance to the body. In this way, the Realist posits a dualism of reasons.

This second Realist dualism is perhaps best made clear by way of a contrast with an alternative account of moral reasons. On the Neo-Kantian account, for instance, moral reasons are of the same order as non-moral reasons. The difference between moral and non-moral reasons, on this Neo-Kantian account, isn't quite one of degree, but it is not a difference in kind either. The difference is perhaps best described as one of directness. On this Neo-Kantian account, moral reasons follow directly from an individual's nature as a (reasoning) person: it is in virtue merely of the fact that I am a person that you ought not to cause me undue harm. Non-moral reasons, on the other hand, follow less directly: I have a reason to complete a dissertation not merely in virtue of the fact that I am a person, but in virtue of the fact that I am a person-who-wants-to-obtain-her-PhD. So, on this Neo-Kantian account, moral and non-moral reasons are essentially the same kind of thing, differing only in how directly they stem from an individual's personhood.⁴⁰

Contrast this Neo-Kantian account with the Realist approach, according to which moral and non-moral reasons differ not in how directly or indirectly they are derived from the same source, but in their very *directedness*: moral reasons, on the Realist account, are *world-regarding*, rather than person-regarding – they don't even 'look' in the same direction as non-moral reasons, let alone doing so more directly. Moral reasons, on either a naturalist or a non-naturalist Realism, are different kinds of things to non-moral reasons. Thus, Realism posits a dualism of reasons, as Cartesian Dualism posits a dualism of substances.

This is the first point of analogy between Realism and Cartesian Dualism, and it gives rise to the second, as follows.

3.2. Epistemological model

Realism and Cartesian Dualism's common metaphysical dualism leads both to also share an epistemological model. In both cases, knowledge of the subject matter is a matter of *perceiving*

³⁹ See Ch.1, s.2.

⁴⁰ See Ch.1, s.5.

what is the case. This is a natural implication of conceiving of the relevant subject matter as a thing that exists in the world – for, how do we know about such things? – we see, touch or otherwise perceive them, of course.

On the one hand, the Cartesian Dualist posits that mental states are known by *introspection*: that is, the perception of the contents of one's own mind. As Ryle describes it, a person "can take a (non-optical) 'look' at what is passing in his mind",⁴¹ and this is how he/she knows what his/her mental states are.

The Realist epistemological model, on the other hand, is likewise perceptual – only, where the Cartesian Dualist looks inwards, the Realist looks outwards and regards the world. On the Realist account, coming to have moral knowledge is a matter of *discovery*, discovering the way the world is. (Again, construct this with the Neo-Kantian account, according to which I come to have knowledge of moral reasons not by looking to the world but by reflecting on my own nature as a person.⁴²) As we have already discussed, Moral Realists are often fond of drawing a close analogy between their domain and Scientific Realism.⁴³ On such Realist accounts, as we come to discover other facts about the world, like the fact that electrons are negatively charged, so we discover the nature of right and wrong.

This perceptual approach is particularly stark in the case of Realist intuitionists, such as Prichard,⁴⁴ according to whom we literally *intuit* those moral reasons existing in and of themselves. But, even non-intuitionist Realist Nagel endorses such a perceptual account. For instance, Nagel writes: "The method [for gaining knowledge of moral facts] is to begin with the reasons that *appear* to obtain [...]",⁴⁵ and he writes this in the aptly named book *The View From Nowhere* – suggesting that reasons are known by our *viewing* them. Interestingly, Nagel draws the same analogy between Metaethics and Philosophy of Mind that I have drawn, comparing Realism to non-Physicalism/Cartesian Dualism. In contrasting his Realism with Moral Nihilism, he writes: "I believe that this [Nihilist] conclusion is the result of a mistake comparable to the one that leads to [P]hysicalism, with its attendant reductionist elaborations."⁴⁶ He draws this comparison in the same passage, quoted above, in which he explicitly characterizes his account as perceptual (although there are many such passages),

⁴¹ Ryle 1949/2009, p.4.

⁴² See Ch.1, s.5.

⁴³ See Ch.5, s.2.2.1.

⁴⁴ Prichard 1912,

⁴⁵ Nagel 1986, p.141 (emphasis added).

⁴⁶ Nagel 1986, p.141.

suggesting that we can know that Nihilism is wrong and that there really are such things as moral reasons because we can *perceive* them.⁴⁷

So, both Realism and Cartesian Dualism share an epistemological model, according to which knowledge of the relevant subject matter is gained by something like perception. This follows on the basis of both theories' understanding their subject matters in metaphysical terms. And, both of these points have led to criticisms of both Realism and Cartesian Dualism, which, as I will go on to show in the following section, are again analogous.

3.3. Analogous criticisms

In this present section, I limit myself to discussion of just two analogous criticisms levelled at Cartesian Dualism, on the one hand, and Realism, on the other. Others will arise later in the chapter.

Firstly, Cartesian Dualism is often criticized as positing something “spooky”⁴⁸ or “occult”⁴⁹ in the form of mental substance. Likewise, Realism is famously criticized by Mackie as positing “queer” entities, “of a very strange sort, utterly different from anything else in the universe”.⁵⁰ This strikes me as being roughly the same criticism in both cases; namely, that there's something *weird* about the idea of mental substance as there is about Real moral facts. Neither is the kind of stuff posited by our best science.

Non-naturalists clearly posit such weird entities in the form of their *sui generis* facts, but it may appear that naturalists, such as Utilitarians, escape this charge. Hedonistic Utilitarians, for instance,

⁴⁷ Korsgaard (1996, p.41, fn.68), likewise, describes Nagel's metaethic as one according to which “our relation to reasons is one of *seeing* that they are there [...]”

Korsgaard would complain, however, that Nagel's argument is no response to Nihilism at all. It does not address the amoralist's challenge (see my Ch.1, s.2.2.). Telling the amoralist that there are such things as moral reasons because he can “see” that there are does not address his worry, which is: “It *seems* like there are such things as moral reasons, but are there *really* such things?” The Realist response that one can “see” that there are reasons is like telling the external world sceptic that of course there's an external world because he can see that there is, which is to miss the point of the worry altogether.

However, this isn't *quite* Nagel's Realist response. His argument, rather, is that, if it appears that there is a reason from many perspectives different to your own, then you can be reasonably certain that there is indeed a reason. (If several other people at the party also report seeing a pink elephant, as well as the pizza delivery boy who has not partaken of the punch, then you can be reasonably certain that you are not mistaken and there is indeed a pink elephant in the room.) To quote in full that quotation provided in part, above: “The method [for gaining knowledge of moral facts] is to begin with the reasons that *appear* to obtain from my own point of view and those of other individuals; and ask what the best perspectiveless account of those reasons is.” (Nagel 1986, p.141). However, the idea that others' points of view are relevant in determining what our reasons are leads Korsgaard to conclude that Nagel is a *Constructivist* (probably of the Contractualist variety) and not a Realist at all: taking others' points of view into account isn't a useful heuristic for discovering what is really out there, but a means of determining – in the sense of *constructing* – what our reasons are. (See Korsgaard 1993, and 1996, p.41, fn.68.)

⁴⁸ See Gertler 2008, p. 295.

⁴⁹ See Ryle 1949.

⁵⁰ Mackie 1977, p.38.

identify morally relevant facts as states of pleasure – and what could be more natural, and less ‘weird’, than pleasure? Mackie argues, however, that all versions of Realism – or ‘objectivism’, to use his term – posit queer entities:

the central thesis of intuitionism is one to which any objectivist view of values is in the end committed: intuitionism merely makes unpalatably plain what other forms of objectivism wrap up.⁵¹

Mackie’s criticism of the queerness of Realism is tied to his objections concerning the Realist epistemological model, which I discuss as part of my second point, below.

The second criticism common to both Cartesian Dualism and Realism is that they both give rise to scepticism about their respective subject matters. Mackie’s criticism of the queer metaphysical nature of Real moral entities (above) is tied to his criticism of the Realist’s epistemological model, in particular that of a Realist intuitionist, such as Prichard.⁵² According to Mackie, the suggestion that we ‘intuit’ moral truths is nothing short of “a travesty of actual moral thinking”.⁵³ As Cartesian Dualism posits that we have knowledge of our mental states by way of a special perceptual sense – namely, introspection⁵⁴ – so must intuitionists posit that we have a “special faculty of moral perception or intuition”.⁵⁵ The problem with this, however, is that it is simply *false* that we possess such a faculty of moral sense. This leads Mackie to embrace the sceptical position of Error theory, asserting that all moral claims are false and that we therefore have no moral knowledge. Insofar as Realism’s metaphysical stance commits it to the claim that moral knowledge is had by moral intuition, since we possess no such faculty of moral intuition, scepticism follows.

Despite the analogies between Realism and Cartesian Dualism, however, it may seem that Cartesian Dualism can avoid this sceptical outcome. This is because, according to Cartesian Dualism, mental states are transparent and introspection is incorrigible. Therefore, even though our knowledge of morality is subject to scepticism, our knowledge of our mental states is apparently indubitable and not a matter of scepticism. If Mackie is correct, scepticism may well be the implication of a Prichardian Realism – but it needn’t be that of Cartesian Dualism. I can, Cartesian Dualism asserts, have knowledge of my own mind.

⁵¹ Mackie 1977, p.38.

⁵² See s.3.2, above.

⁵³ Mackie 1977, p.38.

⁵⁴ See s.3.2, above.

⁵⁵ Mackie 1977, p.38.

We do have several reasons for doubting these Cartesian claims to transparency and incorrigibility⁵⁶ – but, even if they are true, they pertain only to one’s knowledge of one’s *own* mind. When it comes to knowledge of *others’* minds, Cartesian Dualism suffers the same sceptical fate as Mackie claims of Realism. That is, Cartesian Dualism suffers the ‘zombie problem’ (otherwise known, much less excitingly, as the ‘problem of other minds’): even though you behave in all the same, intelligent ways that I do, because I have no access to *your* mental states, I cannot know that you have a mental life as I do; you may be nothing more than a ‘zombie’ going through the motions. My faculty of introspection allows me to know my own mental states, but I possess no faculty of “extrospection” that would allow me to know yours. Thus, even though Cartesian Dualism might avoid it in the case of self-knowledge, the theory falls prey to scepticism when it comes to knowledge of others’ minds. On the Cartesian Dualist account, we cannot know others’ mental states because we do not possess the requisite faculty of extrospection; likewise, on the Realist account, according to Mackie we cannot know moral facts because we do not possess the requisite faculty of moral sense. That is, on both the Cartesian Dualist and Realist intuitionist accounts, scepticism looms large.

But, one may object, not all Realists are intuitionists. (Not even all non-naturalists are intuitionists.) Not all Realists are committed to the idea that we possess a special moral sense by which we discover moral facts. Many Realists will have a much more plausible epistemological account, according to which we come to have moral knowledge not by direct intuition but by *inference* (which is, after all, how we come to have most of our knowledge about the world). I turn to these more plausible Realist accounts of moral knowledge, below.

3.4. Inferential knowledge

The sceptical arguments canvassed above⁵⁷ presuppose that the immediacy of perception is the only way to gain knowledge of things. And, of course this isn’t the case. Since we are limited, finite beings, very little of our knowledge is actually the result of direct perception. Rather, much – if not most – of it is the result of *inference*.

So, we might argue in response to the Cartesian Dualist’s zombie problem that we *can* have knowledge of others’ mental states – by inference. Because, in my own case, I know that it is my desire to obtain my PhD that causes me to work tirelessly on my dissertation, when I observe another poor soul engaged in similar labour, I can justifiably infer that it is because he/she too

⁵⁶ Problems with Cartesian introspection were noted as early on as David Hume (1739, Bk.1, pt.IV, s.6). They are seriously challenged by Sigmund Freud’s theory of the Unconscious. And, more recently, Cartesianism’s “perceptual model” is criticized by Sydney Shoemaker (1994a and 1994b) and Richard Moran (2001).

⁵⁷ See s.3.3, above.

desires to obtain his/her PhD. Thus, we might argue that we can indeed have knowledge of others' mental states, by inference.

And, we might argue similarly with regard to Realism that we can have knowledge of moral reasons, by inference. Others' intelligent behaviour is caused by their mental states, thus I can infer on the basis of your intelligent behaviour that you have a mental states like mine; moral behaviour is caused by moral reasons, therefore I can infer on the basis of your (or my) moral behaviour that there are Real, moral reasons; that is, I can infer Realism. We don't need Mackie's special faculty of moral sense, because we can infer the existence of Real, moral reasons without it. For instance, Realist Richard Werner argues that:

the most reasonable explanation of reports of moral observations includes the positing of the existence of the moral entities mentioned in the reports.⁵⁸

In this paper, Werner draws a close analogy between Moral Realism and Scientific Realism.⁵⁹ So, for instance, we know that electrons must exist because they explain our scientific observations, such as a burning light bulb; likewise, we know that moral reasons must exist because they explain our moral observations.⁶⁰ A Realist may, therefore, reject intuitionism – probably for those good reasons outlined by Mackie⁶¹ – and nevertheless hold that we do have knowledge of moral reasons: inferential knowledge, that is. As Werner holds: “Both moral and scientific entities are posits which provide *the inference to the best explanation* given what we do observe”,⁶² as do many other Realists, some of whom I discuss in the following section.

Thus, a further analogy between the two domains is that Realism, like Cartesian Dualism, can avoid the threat of scepticism by positing inferential knowledge of the relevant subject matter's existence.

3.5. Origins

Above, I considered the suggestion that Cartesian Dualism can claim that we do have knowledge of others' minds, where that knowledge is inferential knowledge.⁶³ This claim that we can infer the existence of others' mental states is what Ryle calls the “origin” of Cartesian Dualism.⁶⁴ According to Ryle, the origin of Cartesian Dualism is the natural observation that there is a “difference between

⁵⁸ Werner 1983, p.653.

⁵⁹ See my Ch.5, s.2.2.1.

⁶⁰ We might construe Nagel's argument for Realism (see s.3.2, above) along similar lines; that is, as an inference to the best explanation. (See fn.47, above, for Korsgaard's alternative interpretation of Nagel.) Nagel, however, takes the explanandum to be not moral actions – as my example suggests – but moral points of view, opinions on what is or isn't morally right. If a sufficiently large group of us – especially a sufficiently large group of us with diverse points of view – all agree that ϕ ing would be wrong, then the best explanation of this 'moral observation' of our convergence is that it is indeed a moral fact that ϕ ing is wrong.

⁶¹ Mackie 1977, p.38, as cited in s.3.3, above.

⁶² Werner 1983, p.667 (emphasis added).

⁶³ See s.3.4, above.

⁶⁴ Ryle 1949/2009, p.8.

human behaviours which we describe as intelligent” and “those which we describe as unintelligent”, such as the behaviours of animals. The Cartesian Dualist, according to Ryle, explains this difference as follows:

The difference between the human behaviours which we describe as intelligent and those which we describe as unintelligent must be a difference in their causation; so, while some movements of human tongues and limbs are the effects of mechanical causes, others must be the effects of non-mechanical causes, i.e. some issue from movements of particles of matter, others from workings of the mind.⁶⁵

That is, the difference between intelligent behaviour and unintelligent behaviour is taken by the Cartesian Dualist to be a difference in *causes*: when behaviour is intelligent it's because it has a special kind of cause – namely, a *mental* cause: thought – as opposed to the standard kind of cause, a physical cause. This is the “origin” of Cartesian Dualism, therefore: there is observed to be a difference between intelligent and unintelligent behaviours, and we explain this observed difference by inferring a special cause in the former case. This special cause is the mental state. Therefore, the Cartesian Dualist concludes that mental substance exists – that is, the Cartesian concludes that there is a dualism of substances: physical, as well as mental.⁶⁶

Realism arguably shares the same origin as Cartesian Dualism. As the Cartesian Dualist posits mental substance in order to explain the observed difference between intelligent and unintelligent behaviour, the Realist posits moral facts in order to explain the difference between moral and non-moral (amoral or immoral behaviour). As Realist Shafer-Landau puts it:

Moral facts are routinely invoked in causal explanations of events in the world. An officer's moral corruption explains his having taken the bribes; the goodness of the benefactor caused her to donate as she did; the stranger's moral probity caused him to return the found money to its rightful owner.⁶⁷

That is, as the Cartesian Dualist posits mental substance in order to explain the occurrence of intelligent behaviour, Realists likewise posit moral facts in order to explain the occurrence of moral behaviour.

Following Shafer-Landau's lead, I present my own examples of cases in which moral facts are employed in order to explain the observed difference between moral and non-moral behaviour (where “non-moral behaviour” covers both amoral behaviour as well as positively immoral behaviour). Consider three characters: the generous man, who gives a handful of coins to a beggar; the scatter-brained man, who accidentally drops a few coins into a beggar's lap when they fall through a hole in his pocket; and the vainglorious man, who tosses his change into a beggar's lap to

⁶⁵ Ryle 1949/2009, p.9.

⁶⁶ See s.3.1, above.

⁶⁷ Shafer-Landau 2003, p.100.

impress his girlfriend. Intuitively, the generous man's action is moral, while the scatter-brained man's and vainglorious man's actions are non-moral. Plausibly, what sets the generous man's action apart as *moral* is that it was caused by (his appreciation of⁶⁸) “the Good”, whereas neither the scatter-brained nor the vainglorious men's actions were so caused. The generous man's action is moral because he is driven to act by something greater than himself, something beyond him: *Real moral facts*, like the fact that it is good to give to those in need. The vainglorious man, on the other hand, is led to action by something else: by his own selfish desire to impress. That is, the difference between moral and non-moral action is easily accounted for by Realism as follows: amoral actions, such as the scatter-brained man's, are 'uncaused' and accidental; immoral actions are caused by the agent's own selfish desires – desires internal to the self; and moral actions are caused by something greater than the agent, something transcending the agent and therefore external to the agent: Real moral facts.

For a clearer view of Realism's explanatory power here, contrast Realism's analysis of the generous man's action with that of Neo-Kantianism. On the Neo-Kantian account, what is moral is a *person-regarding* matter.⁶⁹ As such, on the Neo-Kantian account, there's no difference *in kind* between the actions of the vainglorious man and those of the generous man: *both* issue from something internal to the person, be it the agent's selfish desire to impress or the 'impulse of duty'. Because, unlike Realism, Neo-Kantianism doesn't posit Real, world-regarding moral norms, Neo-Kantianism isn't able to account for the difference between moral and non-moral action as Realism is: on the Neo-Kantian account, both moral and immoral action is caused by something internal to the agent. Realism, on the other hand, is able to account for the difference between moral and non-moral action quite straightforwardly: moral action has a special cause – something not internal but *external* to the agent, namely, world-regarding, Real moral norms.⁷⁰ This is the origin of Realism.

As Cartesian Dualism explains the difference between intelligent and unintelligent behaviour by positing the mental cause of the former, Realism explains the difference between moral and non-moral actions by positing the cause of the former as Real moral facts. In the literature, the Realism debate has indeed revolved around the question of the causal efficacy of moral properties, with

⁶⁸ I examine this line in s.4.2.2, below.

⁶⁹ See Ch.1, s.5.

⁷⁰ Interestingly, here we observe a reversal of the Cartesian Dualist's position. Where the Cartesian Dualist reasons that

[t]o salve our right to employ appraisal-concepts, the field of their proper application had to be shown to lie somewhere else than this external world, and an internal world of unmeasurable but purposeful forces was thought to do the trick⁷⁰

the Realist reasons that, in order to “salve our right to employ appraisal-concepts”, an *external* world of moral facts will “do the trick”.

those theorists advocating Realism arguing precisely that moral properties are causally efficacious, and those opposing Realism arguing that they are not. For instance, Shafer-Landau argues that supervenient moral properties will “inherit” the “causal potency of their physical realizers. And so we can keep them in the ontology.”⁷¹ That is, we can be sure that Real moral properties do exist, and we can therefore “keep them in the ontology”, *because* they are causally potent.⁷² Shafer-Landau also cites a long list of philosophers who have entered the Realism debate (either advocating Realism or opposing it) on precisely the point of the causal efficacy of moral properties: Robert Audi, David O. Brink, Simon Blackburn, Ronald Dworkin, Allan Gibbard, Gilbert Harman, Colin McGinn, Nicholas Sturgeon, Judith Jarvis Thomson, and Crispin Wright, amongst others.⁷³ Realism and Cartesian Dualism, therefore, share the same origin insofar as both seek to explain an observed difference between two phenomena by positing the special cause of one.

In this section 3 I have argued that there are various points of analogy between Cartesian Dualism and Moral Realism. In particular, both originate in the attempt to explain the observed difference between two phenomena by positing the special cause of one (above). And, in both cases, this leads to the supposition of a metaphysical dualism.⁷⁴ Both also posit similar epistemological models insofar as knowledge of the particular subject matter is taken to consist in a kind of perception.⁷⁵ This in turn produces the further point of analogy that both are similarly accused of leading to scepticism, because we simply do not possess the requisite faculties to perceive what is posited to exist.⁷⁶ Yet, both theories respond to such criticisms by allowing inferential knowledge of the subject matter, which is a further point of analogy between them.⁷⁷ In the following section, I argue that there is yet another point of analogy between Realism and Cartesian Dualism: both are guilty of a category-mistake.

⁷¹ Shafer-Landau 2003, p.109.

⁷² To be fair, however, Shafer-Landau (2003, p.112) does assert that “Nothing I have said here is intended to represent an argument for moral [R]ealism. I do not think that considerations of causal or explanatory efficacy can do that.” However, the causal “considerations” he presents do *look* a lot like an argument, and at least one reviewer, Miller (2006), also reads Shafer-Landau like this.

⁷³ Audi 1997, pp.115-121, Brink 1989, pp.182-197, Blackburn 1991, Dworkin 1996, Gibbard 1990, Ch.6, Griffin 1996, pp.61-64, Harman 1977 and 1986, McGinn 1997, Ch.2 and 3, Quinn 1986, Sturgeon 1984, 1986, 1991 and 1995, Thomson (in Harman & Thomson) 1996, pp.74-94, and Wright 1992, Ch.5, all as referenced by Shafer-Landau 2003, pp.99-100.

⁷⁴ See s.3.1, above.

⁷⁵ See s.3.2, above.

⁷⁶ See s.3.3, above.

⁷⁷ See s.3.4, above.

4. Diagnosing a category-mistake

I have argued that Realism and Cartesian Dualism share the same origin insofar as both seek to explain an observed difference between two phenomena by positing the special cause of one: Cartesian Dualism, in positing mental substance in order to explain intelligent (as opposed to unintelligent) behaviour, and Realism, in positing Real, moral reasons in order to explain moral (as opposed to non-moral) behaviour.⁷⁸

However, according to Ryle, this origin of Cartesian Dualism is precisely its problem. It goes wrong right at the very beginning; it is conceptually confused and category-mistaken. Cartesian Dualism's (biggest) mistake, claims Ryle, is precisely its conceiving of the mind as the *cause* of intelligent behaviour in the first place.⁷⁹ To quote Ryle again:

The differences between the physical and the mental were thus represented as differences inside the common framework of the categories of 'thing,' 'stuff,' 'attribute,' 'state,' 'process,' 'change,' 'cause' and 'effect'. Minds are things, but different sorts of things from bodies; mental processes are causes and effects, but different sorts of causes and effects from bodily movements. And so on.⁸⁰

Therefore, if Realism shares the same origin as Cartesian Dualism in positing a special kind of cause in order to explain an observed difference between phenomena, then Realism is likely guilty of the same category-mistake.

In the following section, I present Ryle's justification for his diagnosis of a category-mistake with reference to Cartesian Dualism. Thereafter, I go on to mount essentially the same argument with reference to Realism, which will support an analogous diagnosis.

4.1. Cartesian Dualism's category-mistake

Ryle's support for the charge of category-mistake is that Cartesian Dualism suffers some serious problems – and that these are problems of the Cartesian's own category-mistaken making. Ryle's contention is not with these problems themselves, however, but is rather the claim that Cartesian Dualism suffers such problems *because* it commits a category-mistake.

Take, for instance, the Cartesian Dualist's problem of mental causation.⁸¹ As is well known, this is the problem that, if the mind is a separate substance to the body and not extended in space, as the Cartesian Dualist takes it to be, then it's utterly mysterious how the mind could enter into causal interactions with the body and, for instance, "get my finger to squeeze the trigger".⁸² Quoting Ryle:

⁷⁸ See s.3.5, above.

⁷⁹ See s.2, above.

⁸⁰ Ryle 1949/2009, p.9, as cited in s.2 above.

⁸¹ Ryle 1949/2009, pp.2-3.

⁸² Ryle 1949/2009, p.68.

[...] since minds, not being in space, could not be described as being spatially inside anything else, or as having things going on spatially inside themselves [...] theorists are found speculating how stimuli, the physical sources of which are yards or miles outside a person's skin, can generate mental responses inside his skull, or how decisions framed inside his cranium can set going movements of his extremities.

Even when 'inner' and 'outer' are construed as metaphors, the problem [of] how a person's mind and body influence one another is notoriously charged with theoretical difficulties.

What the mind wills, the legs, arms and the tongue execute; what affects the ear and the eye has something to do with what the mind perceives; grimaces and smiles betray the mind's moods [...]. But the actual transactions between the episodes of the private history and those of the public history remain mysterious, since by definition they can belong to neither series.⁸³

That is, the Cartesian Dualist's own dualism of mental and physical substances⁸⁴ means that the mind (on the Cartesian conception thereof) can never causally interact with the body.

However, recall the origin of Cartesian Dualism:⁸⁵ the mind/body substance-division was originally brought in by the Cartesian Dualist in order to explain the difference between intelligent behaviour and unintelligent behaviour: the former was postulated to be the effect of a special kind of *cause*, thought.⁸⁶ Yet the Cartesian construal of the mind as the opposite substance to that of the body has the peculiar effect of rendering the mind *unable* to enter into causal interactions with the body, thus *precluding* the (Cartesian) possibility of intelligent behaviour. The Cartesian binds his own hands with mental substance. What was postulated in order to explain minded behaviour ends up precluding its very possibility. This awful irony functions something like a *reduction ad absurdum*: since the mental can't possibly both be and not-be the cause of intelligent behaviour, it must have been a mistake – a category-mistake – to think of the mental as the *cause* of such behaviour in the first place.

Thus, Ryle's charge of category-mistake is something like a diagnosis: the symptoms are the various well-known problems faced by Cartesian Dualism, such as the problem of mental causation, and the underlying condition producing those symptoms is the category-mistake. Further support for this diagnosis is gained by the fact that, when we adopt Ryle's alternative Dispositional account of the mental, the problem of mental causation disappears. Unfortunately, it is beyond the scope of this present paper to examine Ryle's positive account of the mind, but what I will say is that it manages to avoid the problems faced by the category-mistaken Cartesian Dualist because it conceives of the mental in a radically different manner to that in which the Cartesian Dualist conceives of it: namely, as a (set of) *disposition(s)*, rather than as a (collection of) *state(s)*. The following quotation from Ryle

⁸³ Ryle 1949/2009, p.2.

⁸⁴ See s.3.1, above.

⁸⁵ See s.3.5, above.

⁸⁶ See s.3.5, above.

is brief and not contextualized, but hopefully serves to draw the distinction between the category-mistaken conception of the mental and Ryle's radically different Dispositional one:

But the reason why the skill exercised in a performance cannot be separately recorded by a camera is not that it is an occult or ghostly happening, but that it is not a happening at all. It is a disposition, or complex of dispositions, and a disposition is a factor of the wrong logical type to be seen or unseen, recorded or unrecorded.⁸⁷

Unfortunately, I am not here in a position to explain Ryle's Dispositional account, nor how it avoids the problem of mental causation, but I do hope to do more than this in the Metaethical realm that has been the preoccupation of this project. In particular, I hope to show that, as Ryle's Dispositional account is a radically different conception of the mental to that of the category-mistaken Cartesian Dualist – “not [...] an occult or ghostly happening, but [...] not a happening at all” – so is the Ryle-inspired Second-Personal approach a radically different conception of morality to that of Realism. So, if Realism suffers from a category-mistake, what are its symptoms?

4.2. Realism's category-mistake

The symptom of Cartesian Dualism's category-mistake is the problem of mental causation. Likewise, I will argue, the symptom of Realism's category-mistake is what I will call the problem of moral responsibility.

4.2.1. The problem of moral responsibility

The “origin” of Realism, I have argued, is the explanation of the observed difference between moral and non-moral behaviour, where moral behaviour is that with a special cause: the Real, moral fact. For instance, the generous man's action is moral, whereas the vainglorious man's action is not, because the generous man's action is caused by the transcendent, moral reason to help those in need, whereas the vainglorious man's action is caused by his selfish desire to impress his girlfriend.⁸⁸

The problem with this Realist conception of moral action arises, however, when we consider that moral action is action for which we are morally responsible. This is essentially Mill's Point.⁸⁹ In the preceding chapter, I argued that Realism is indeed capable of providing a reason “of the right kind” for holding agents morally accountable, where “what matters” is “whether the person's conduct is culpable”⁹⁰ (and not whether holding that agent accountable is socially desirable).⁹¹ But, we must ask, when is an agent culpable? And, an agent is surely only culpable for his/her actions when that agent acted *freely*. Moral action, therefore – action for which we are responsible – must be free

⁸⁷ Ryle 1949/2009, pp.21-22.

⁸⁸ See s.3.5, above.

⁸⁹ See Ch.3.

⁹⁰ Darwall 2006, p.15, as cited in my Ch.5, s.2.1.

⁹¹ See Ch.5, s.2.1.1.

action. The problem of moral responsibility, as I will go on to explain, is that Realism's own account of the nature of moral reasons bars it from giving an account of moral action as free action.

There are, of course, various accounts of freedom available. Some theorists – incompatibilists and libertarians – hold that free actions are those which are entirely uncaused by prior events.⁹² Such incompatibilist accounts, however, are understandably unpopular. As the compatibilist Walter Stace has argued, *contra* incompatibilism:

[M]oral responsibility is not only consistent with determinism, but requires it. The assumption on which punishment is based is that human behaviour is causally determined. [...] If human actions and volitions were uncaused, it would be useless either to punish or reward [...]. Thus moral responsibility would entirely disappear. [...] And this is in itself a strong argument against the common view of philosophers that free will means being undetermined by causes.⁹³

That is, Stace argues that, far from it being the case (as the incompatibilist claims) that free and morally culpable behaviour is behaviour "undetermined by causes", free and morally culpable behaviour is precisely "human behaviour [that] is causally determined". This is because behaviour that is uncaused by prior events would be capricious and random – not the behaviour of an agent who acts for reasons, and therefore not morally culpable behaviour. I accept as correct this argument by Stace for compatibilism. Free and morally culpable behaviour must be behaviour that is causally determined.

Free and morally culpable behaviour, however, must be behaviour determined by the *right* kind of cause. What constitutes the right kind of cause? To this end, Stace presents the reader with a series of pairs of examples:

Gandhi fasting because he wanted to free India.	The man fasting in the desert because there was no food.
Stealing bread because one is hungry.	Stealing because one's employer threatened to beat one.
Signing a confession because one wanted to tell the truth.	Signing because the police beat one.
Leaving the office because one wanted one's lunch.	Leaving because forcibly removed. ⁹⁴

Stace notes that all eight of these actions are causally determined – however, those four acts in the left-hand column are acts we usually describe as free. What is special about the causes of the acts in the left-hand column that makes them free acts, and those in the right-hand column not?

⁹² See, for instance, Chisholm 1976 and 1978.

⁹³ Stace 1952/2008, p.424.

⁹⁴ Stace 1952/2008, p.422.

The criterion of a free cause, Stace theorizes, is simply that it is *internal* to the agent. For instance, Gandhi's fasting is a free action, and a praiseworthy one too, because it is caused by his *desire* to free India, whereas the man lost in the desert fasts for no such desire and his fasting is therefore not praiseworthy. As Stace captures this compatibilist thesis:

[F]ree acts are all caused by desires, or motives, or by some sort of internal psychological states of the agent's mind. [...] [U]nfree acts, on the other hand, are all caused by physical forces or physical conditions outside the agent.⁹⁵

And, this compatibilist position, according to which agents act freely when they are caused to act by "internal psychological states of mind" rather than by external causes, appears to be Darwall's position too. As Darwall writes:

In [...] holding [an agent] responsible, we would have to assume that he had *within him* a source of motivation to do as he was morally obligated.⁹⁶

This internality of the agent's reason for action, Darwall explains, "is the difference between a kind of coercion, on the one hand, and free self-determination [...], on the other".⁹⁷

Above, I presented three cases: the generous man, the scatter-brained man, and the vainglorious man.⁹⁸ Now consider a fourth case: a man has a gun pointed at him by a crazed egalitarian, who orders the man to give the beggar all of his change, and, naturally, the man complies. Call this character the coerced man. Is the coerced man morally responsible for his action, is he to be praised for bestowing such a handsome sum upon the beggar? Of course not. And the compatibilist account of freedom (above) easily explains why not: the cause of the coerced man's action is not internal to himself; instead, he is moved to action by the threat of physical harm. Here we might usefully borrow Pufendorf's expression: the coerced man's action is not free, because his will has been shaken "with an external force", namely, the "object of terror" that is the crazed egalitarian's threat.⁹⁹ Since the cause of the coerced man's action is not internal to himself, his action is not one for which he is morally responsible; it is a non-moral action.¹⁰⁰

⁹⁵ Stace 1952/2008, p.423.

⁹⁶ Darwall 2006, p.34 (emphasis added).

⁹⁷ Darwall 2006, p.23.

⁹⁸ See s.3.5, above.

⁹⁹ Pufendorf 1672/1934, p.91, as cited by Darwall 2006, p.111, as cited in my Ch.1, s.6.1.2.

¹⁰⁰ There are, of course, other compatibilist accounts available than that of Stace, in particular, those that provide more sophisticated accounts of what counts as an "internal" cause than Stace's simple criterion of mentality.

For instance, on Kwame Anthony Appiah's (2003) compatibilism, "[w]hat makes me responsible, in short, is that I acted on my reasons", where this is a compatibilist account in that "[i]t is simply irrelevant whether those reasons were the result of inexorable causal processes" (p.376). On the face of it, this version of compatibilism may appear congenial to the Realist account of moral reasons. However, note that Appiah defines a "reason" as follows: "Let us call *my understanding of the facts and my aims*, taken together, "my reasons"" (p.376; emphases added). That is, a reason, for Appiah, is already an internal, mental state. In

On this compatibilist account of free and morally culpable action, Realism's problem of moral responsibility comes into focus. As the origin of Cartesian Dualism is the supposition of mental substance in order to explain the observed difference between intelligent and unintelligent behaviour, the origin of Realism is the supposition of Real, moral facts in order to explain the observed difference between moral and non-moral behaviour. In the Cartesian Dualist case, intelligent behaviour is to be distinguished from unintelligent behaviour because the former has a special kind of cause: thought. In the Realist case, moral behaviour is to be distinguished from non-moral behaviour because the former has a special kind of cause, one that transcends selfish, human desires: the Real, moral fact.¹⁰¹ Only actions resulting from this special kind of objective cause are moral actions and therefore actions for which agents are responsible.

However, note that on a Realist metaethic, Real moral facts will be *external* to the agent. Realism is a world-regarding metaethic.¹⁰² It posits a dualism of reasons, such that moral reasons are different kinds of things to non-moral reasons – precisely in that they are *not* products of the agent's own desires, but instead come from outside him/her, from the world.¹⁰³ Unlike Neo-Kantian reasons, which are known by reflecting on the nature of oneself as a person, Real moral reasons are known by discovery, by looking beyond oneself rather than deeper into oneself.¹⁰⁴ On a Realist metaethic, therefore, moral reasons are external to agents. As Kevin DeLapp puts this point, there is "[t]oo deep of a rift between actual moral agents and [the] free-floating disembodied values" posited by Realism,¹⁰⁵ such that those Real moral reasons are external to agents.

claiming that an agent is responsible when he/she acts on his/her reasons, therefore, Appiah is, like Stace, claiming that free behaviour is that cause by internal, psychological states. Likewise, Frankfurt (1971) provides a more sophisticated compatibilism than Stace's, but it essentially depends on the same free-internal/unfree-external division as Stace's. According to Frankfurt, an agent acts freely when he/she *identifies* with the desire that gives rise to his/her behaviour; this leaves room for the possibility of behaviour that is unfree, even though it is caused by the agent's own mental states, because it is caused by a desire with which the agent does not identify. Here a key example is Frankfurt's "unwilling addict", who takes drugs because he wants to – but who (contrary to a less sophisticated account such as Stace's) does so *unfreely*, because the unwilling addict does not identify with his desire to use drugs. I do think that a more sophisticated compatibilist account – particularly such as Frankfurt's – is more plausible. However, it depends on essentially the same internal/external division as Stace's account – only with a different criterion for what counts as an internal, and therefore free, cause. Since the basic idea of Frankfurt's compatibilism is essentially the same as that of Stace's, the possibility of this more sophisticated version of compatibilism will not affect my argument to follow. In fact, as I will go on to explain in fn.122 (below), this more plausible, Frankfurtian compatibilism actually presents a *harder* case for the Realist. As such, I stick to a Stacean compatibilism in the body of the argument, out of charity to the Realist.

¹⁰¹ See s.3.5, above.

¹⁰² See Ch.1, s.2.

¹⁰³ See s.3.1, above.

¹⁰⁴ See s.3.2, above.

¹⁰⁵ DeLapp 2009, p.2.

And, the upshot of this Realist account of the nature of moral reasons is that, if they lead agents to action, they cannot do so *freely*. On the most plausible, compatibilist account of freedom, “free acts are all caused by desires, or motives, or by some sort of internal psychological states of the agent’s mind”,¹⁰⁶ yet Real moral reasons are, by definition, *not* “internal psychological states of the agent’s mind”. On a Realist account, therefore, moral action is unfree action, because it is action brought about by an *external* cause: the Real moral reason. But unfree action is action for which agents are not responsible. So, on the Realist account, moral action is, absurdly, action for which agents are not morally responsible. This is the Realist’s problem of moral responsibility.

The Cartesian Dualist posits mental substance in order to explain the difference between intelligent and unintelligent behaviour, where the former is caused by thought and the latter is not. However, in positing that thought is of an entirely different substance to the body – an immaterial substance unable to causally interact with physical substance (or, indeed, with anything at all) – the Cartesian Dualist effectively precludes the possibility of the very phenomenon he/she originally set out to explain: intelligent behaviour. Likewise, the Realist posits Real, moral facts in order to explain the difference between moral and non-moral behaviour, where the former is caused by the Good and the latter is not. However, in positing that Real, moral reasons are world-regarding, existing externally to agents, the Realist effectively precludes the possibility of the very phenomenon he/she originally set out to explain: moral behaviour, behaviour for which we are morally responsible when we act freely.

Below, I consider a popular objection to my claim that Realism suffers the problem of moral responsibility. But, I go on to reject this objection. Having defended the claim that Realism suffers the absurd problem of moral responsibility – as Cartesian Dualism suffers the absurd problem of mental causation – I go on to claim that Realism is, like Cartesian Dualism, guilty of a category-mistake.

4.2.2. The internalization defence

I have argued that Realism suffers the problem of moral responsibility, in that the Realist account of moral reasons implies that no moral action can be performed freely.¹⁰⁷ As DeLapp puts the point,

To be precise, DeLapp thinks that this “rift” is posited by *Objectivism*, which he idiosyncratically takes to be a distinct metaethical to Realism, although the two are commonly conflated. What I (and many others) have described as ‘Realism’ is what DeLapp thinks is more properly called ‘Objectivism’.

¹⁰⁶ Stace 1952/2008, p.423, as cited above.

¹⁰⁷ See s.4.2.1, above.

there is “[t]oo deep of a rift between actual moral agents and [the] free-floating disembodied values” posited by Realism,¹⁰⁸ such that those Real moral reasons are always external to agents.

One common response to this charge is as follows: all the Realist needs to do is close the “rift” between agents and moral facts. And, this is easily done by claiming that agents *aren’t* directly caused by Real, moral reasons to behave morally; rather, their moral behaviour is the result of their *appreciation* of the moral reasons. (Likewise, the tub of ice-cream doesn’t force itself down my throat; rather, I recognize the worldly fact of the ice-cream’s proximity and availability, and then my *desire* to eat it – rather than the tub itself – leads me to do so).

And, the argument goes, most plausible Realisms will indeed posit “closed rifts” in their accounts of moral actions. For instance, Shafer-Landau takes something like this line.¹⁰⁹ His Realism distinguishes between “objective practical reasons” and “[m]otivating or subjective reasons”. Whereas Shafer-Landau’s “objective practical reasons” “are objective facts or standards that are not constrained by an agent’s motivational capacities”, his “[m]otivating or subjective reasons” “are the reasons that an agent *takes herself* to have for performing an action, and they are what motivate that performance.”¹¹⁰ That is, on Shafer-Landau’s Realism, agents are caused to act not by external, “objective practical reasons”, but by the agents’ *appreciations* thereof, by “the reasons that an agent *takes herself* to have”.

The benefit of a Realist account such as the above is clear: an agent’s appreciation of moral reasons (as opposed to the moral reason *per se*) will be “some sort of internal psychological [state] of the agent’s mind”¹¹¹ – precisely the kind of state that compatibilists have required as the causes of free action. And, if it’s the agent’s appreciation of moral reasons, rather than the moral reasons themselves, that cause moral behaviour, such behaviour will indeed be free action, and therefore action for which agents are morally responsible, just as it should be.

Call this the ‘internalization defence’ of the Realist account of moral action, since it works (tries to work) by internalizing the cause of moral action – that is, by construing the cause of moral action not as the Real, moral reason itself, but as the mental appreciation thereof. If the cause of moral action is, as this internalization defence claims, an internal, mental state, subsequent moral action will indeed be free action.

¹⁰⁸ DeLapp 2009, p.2, as cited in s.4.2.1, above.

¹⁰⁹ Shafer-Landau 2003.

¹¹⁰ Miller 2006, p.326, referencing Shafer-Landau 2003.

¹¹¹ Stace 1952/2008, p.423, as cited in s.4.2.1, above.

However, this internalization defence of Realism is ultimately unacceptable. As I explain, below, it pushes the Realist into a dilemma.

4.2.3. The Realist dilemma

Realism, I have claimed, posits Real moral facts in order to explain the observed distinction between moral and non-moral behaviour: moral behaviour is that behaviour caused by Real, moral facts, whereas non-moral behaviour is that which is not.¹¹² The internalization defence, on the other hand, argues that it is not Real, moral facts themselves that cause behaviour, but instead it is the agent's appreciation of those facts that result in action.¹¹³ Since such action is caused by internal, mental states (namely, the state of appreciation of a fact), such action will be free action on a compatibilist model – and will, therefore, be action for which agents are morally responsible, contrary to my claims regarding the Realist's problem of moral responsibility.¹¹⁴

The trouble with the Realist's internalization defence is straightforward, however. If the Realist wishes to argue that moral behaviour is not caused by Real, moral reasons, but by the agent's mental states about those Real, moral reasons, the Realist will indeed have thereby provided an account of moral behaviour according to which moral agents act freely. However, in so doing the Realist will also have undermined his/her own Realism. This is because the origin of Realism, recall, is the explanation of the observed difference between moral and non-moral behaviours, according to which the former has a special cause – and, Realism originates in the supposition that what is special about the cause of moral behaviour is that it is world-regarding, rather than person-regarding: not, say, the selfish desire to impress, but the transcendent, objective, *Real, moral reason*.¹¹⁵ The vainglorious man's behaviour is caused by a mental state, and his behaviour is not moral; on the other hand, the generous man's behaviour has an entirely different kind of cause: not a mental state, but the Real, moral reason that we ought to help others in need, and this is what sets the generous man's behaviour apart as moral behaviour.¹¹⁶ What's special about the cause of moral behaviour on this account – and what gives rise to Realism – is just the idea that the cause of moral behaviour is *not* a mental state but a different kind of cause altogether: *the Real, moral reason*. Yet, this is precisely what the internalization defence *denies*, by insisting that the cause of moral behaviour is not the Real, moral reason, but rather the agent's mental states about Real, moral reasons.

¹¹² See s.3.5, above.

¹¹³ See s.4.2.2, above.

¹¹⁴ See s.4.2.1, above.

¹¹⁵ See ss.3.1 and 3.5, above.

¹¹⁶ See s.3.5, above.

If the Realist advocates the internalization defence, therefore, his/her position flies in the face of the very origin of Realism, thereby undermining his/her own account. And, if we deny the origin of Realism, we have no good reason to posit that there are indeed such things as Real, moral reasons in the first place – that is, we are left without reason to be Realists at all. Christian Miller makes essentially the same point in his compelling criticism of Shafer-Landau’s Realism:¹¹⁷ if Shafer-Landau distinguishes as he does between “objective practical reasons” and “[m]otivating or subjective reasons”, where the latter “are the reasons that an agent takes herself to have for performing an action, and [only] they are what motivate that performance”,¹¹⁸ then it follows that agents are never caused to act by objective practical reasons – *Real reasons*. In Miller’s words, this is the strangely Anti-Realist conclusion that “[w]e can never be motivated by what we ought, objectively, to do.”¹¹⁹ And, this conclusion is a problem “especially for a [M]oral [R]ealist”,¹²⁰ the origin of whose metaethic is just that it is Real, moral reasons that motivate moral action. Likewise, DeLapp makes essentially the same observation of Realists who internalize the cause of moral action; in his words, Realists, “who maintain that moral values are co-extensional with [...] psychological properties threaten to render moral values not quite [R]ealist enough.”¹²¹ In the effort to defend Realism against the problem of moral responsibility, by internalizing the cause of moral action, therefore, such a Realist effectively undermines his/her own Realism.¹²²

¹¹⁷ See s.4.2.2, above.

¹¹⁸ Miller 2006, p.326 (emphasis removed), referencing Shafer-Landau 2003, as cited in s.4.2.2 above.

¹¹⁹ Miller 2006, p.326.

¹²⁰ Miller 2006, p.327.

¹²¹ DeLapp 2009, pp.5-6.

¹²² What if we accept a more sophisticated compatibilism, such as Frankfurt’s (1971), according to which the cause of free behaviour is not only a mental state, but a mental state with which the agent *identifies*? (See fn.100, above.) If this is the correct account of free action, then, in order for an agent’s moral behaviour to be free behaviour and behaviour for which he/she is morally responsible, it must be the case that the agent is led to action not only by his/her appreciation of the moral reasons (as the internalization defence suggests), but by his/her *identification with* his/her appreciation of the moral facts. That is, if we accept this Frankfurtian compatibilism, the criteria which a Realist account of moral action must meet in order for it to count as free action are even more demanding.

And, we have no good grounds for confidence that any agent will indeed identify with his/her mental states about moral reasons. An agent may well have veridical mental states about moral reasons, and yet simply not identify with them. Frankfurt allows that any agent may identify with, or – as he puts it in more recent work – “wholeheartedly love what is evaluatively nondescript, or what is bad, or what is evil” (2004, p.98). And, if an agent were to so identify, any subsequent moral action by that agent would, again, absurdly, be unfree and action for which the agent is not responsible. (This is essentially the same problem for Realism as that already examined in the very first chapter of this project. See Ch1, ss.2.2 and 4.)

It is for this reason, as I commented in fn.100, above, that I have focused instead on a Stacean compatibilism. Even though I think that a Stacean compatibilism, crude as it is, is less plausible than a Frankfurtian compatibilism, it actually presents the more charitable case for a Realist account of free, moral action; such a Stacean account does not meet the Frankfurtian difficulty that agents may simply fail to identify with their mental states about moral reasons.

My argument here draws its inspiration from one originally proposed by the Anti-Realist Gilbert Harman.¹²³ Harman targets the Realist's popular analogy with Scientific Realism.¹²⁴ He argues that, in the case of a scientific observation, like "There goes a proton", "you need to make assumptions about certain physical facts to explain the occurrence of the observations that support a scientific theory"; on the other hand, in the case of a moral observation, such as that it is wrong to set a cat on fire, "you do not seem to need to make assumptions about any moral facts to explain the occurrence of the so-called moral observations".¹²⁵ In order to explain the moral observation, Harman argues, "all we need assume is that you have certain more or less well articulated moral principles that are reflected in the [judgements] you make,"¹²⁶ that "you have acquired perhaps as a result of your upbringing".¹²⁷ Harman's suggestion is that the best explanation of my judgement that, for instance, the action of setting a cat on fire is wrong is just that I was brought up to believe that igniting live animals is wrong (as I indeed was brought up to believe). And, since explanations of moral observations can, according to Harman, "proceed in entirely psychological terms", "we do not need to assume that moral facts or entities exist in order to explain our moral beliefs and motives".¹²⁸ And, if we do not need to posit moral facts in order to explain an individual's moral judgements, then we have no good reason to believe that there are such things as moral facts at all, and Anti-Realism follows. This argument led to the well-known Realist/Anti-Realist debate between Sturgeon¹²⁹ and Harman (later to be joined by Simon Blackburn¹³⁰), the victor of which is still inconclusive.

My present argument clearly shares some affinities with Harman's original argument, in particular, the claim that, if we can explain our observations of moral actions by reference to their mental causes only, then we have no good reason to posit that there are such things as Real, moral facts – that is, we have no good reason to accept Realism. There are, however, some important differences between Harman's original argument and my present argument. Firstly, whereas Harman's argument threatens not only Realism but Moral Cognitivism generally,¹³¹ my present argument targets only Realism. This is because of the second, following difference between our arguments.

¹²³ Harman 1977. See especially Ch.1, pp.3-10.

¹²⁴ See my Ch.5, s.2.2.1.

¹²⁵ Harman 1977, p.6.

¹²⁶ Harman 1977, p.7.

¹²⁷ Harman 1977, p.4.

¹²⁸ Korsgaard 1996, p.45, referencing Harman 1977.

¹²⁹ See Sturgeon 1985.

¹³⁰ Blackburn 1990.

¹³¹ Joyce 2015; Korsgaard 1996, pp.45-47.

This second difference is that Harman is committed to the claim that purely psychological explanations of our observations of moral actions are indeed the best explanations thereof; indeed, his claim is as strong as that “an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the [judgement] you make.”¹³² And, this is precisely the point on which the Realist Surgeon has challenged Harman, arguing that some moral phenomena can only be explained by reference to moral facts.¹³³ This is not a debate I need here to enter into, however – because my present claim is, unlike Harman’s, not that the best explanation of our observations of moral behaviour proceeds in entirely psychological terms. Instead, I make only the conditional claim that, if moral behaviour is caused not by Real, moral reasons but by the agent’s mental states, then we are left without any good reason to infer that there are indeed such things as Real, moral reasons.

And, the rest of my argument is completed by the Realist him-/herself. That is, if the Realist plays the internalization defence against my problem of moral responsibility,¹³⁴ then the Realist him-/herself supplies the antecedent to the above conditional: moral behaviour is caused not by Real, moral reasons, according to the internalization defence, but by the agent’s mental states about moral reasons. And, if the Realist’s own explanation of moral behaviour is that it is caused not by Real, moral reasons but by the agent’s mental states about moral reasons, then, following Harman, I conclude that we have no good reason to believe that there are such things as Real, moral reasons – that is to say, we have no good reason to accept Realism in the first place. (This explains the first difference I described between my argument and Harman’s (above), which is that, while his targets Cognitivism generally, mine targets only Realism. This follows, of course, because I have argued only with reference to those Realists who claim internalization that the antecedent obtains, that is, that moral behaviour is not caused by Real, moral reasons.)

This ushers in a third major difference between my argument and Harman’s, which is that, unlike Harman, I argue that the Realist faces a *dilemma*. If the Realist accepts that moral behaviour is caused by Real, moral reasons, then the Realist gives us a good reason to believe that there are indeed such things as Real, moral reasons, because they are posited by the best explanation of our moral behaviour.¹³⁵ However, if the Realist takes this first option, then the Realist encounters the problem of moral responsibility: if moral action is caused by Real, moral reasons, which are world-regarding and external to the agent, then moral action is not free action and is therefore not action

¹³² Harman 1977, p.7.

¹³³ Surgeon 1985.

¹³⁴ See s.4.2.2, above.

¹³⁵ See s.3.5, above.

for which agents are morally responsible.¹³⁶ On the other hand, if the Realist takes the internalization route and insists that moral behaviour is actually caused by agents' appreciation of Real, moral reasons¹³⁷ – that is, by agents' mental states, which are internal and not external causes of behaviour – then the Realist undermines his/her own position. If moral behaviour is caused by agents' mental states, then moral behaviour will indeed be free, but this explanation of moral behaviour leaves us without good reason to posit that there are indeed such things as Real, moral reasons, since, as Harman noted, all the causal-explanatory work is done by the agent's mental states. And, if we have no good reason to believe that there are such things as Real, moral reasons, then we should reject Realism. This is the Realist dilemma.

As one, final Realist response, it might be argued that the Real, moral reason *does* have a causal role on this picture: moral action is caused (freely) by mental states about moral facts – and those mental states are, in turn, caused by the Real, moral reasons themselves. That is, Real, moral reasons do have a causal role on the internalization picture: causing the mental state of appreciation of moral reasons (which, on turn, causes moral behaviour).

However, we have little compelling, independent reason to accept this response. Causing the appreciation of moral reasons would indeed constitute a causal role, and would salvage the Real, moral reason from causal irrelevance. However, why should we accept the picture proposed above? In order to salvage Realism from the dilemma, the Realist must show not only that it is possible that Real, moral reasons might cause mental states about moral reasons; the Realist must also show that this is *probable*, that it is highly unlikely that we would have any mental states about moral reasons in the absence of causation by Real, moral reasons themselves. But, the Realist cannot convincingly meet this brief. The existence of Real, moral reasons is not the only possible explanation of mental states about moral reasons. It is a well-established fact that not all mental states have veridical causes: consider hallucinations, imaginings, and dreams, *etc.* A hallucination of a pink elephant gives us no good reason to believe that there is indeed a pink elephant present. And, if mental states about moral reasons are of the same category as hallucinations – and why couldn't they be? – then mental states about moral reasons will, likewise, give us no good reason to believe that there are indeed such things as Real, moral reasons. Indeed, here we might draw yet another point of analogy between Realism and Cartesian Dualism:¹³⁸ as the Cartesian Dualist suffers the problem of external world scepticism, so will a Realist who plays the internalization defence suffer the problem of what we might call “external reason scepticism”. Just our having mental states about moral reasons does

¹³⁶ See s.4.2.1, above.

¹³⁷ See s.4.2.2, above.

¹³⁸ See s.3, above.

not entail that there are indeed such things as Real, moral reasons that cause those mental states about moral reasons. Those mental states about moral reasons could very well be caused instead – as Harman has pointed out – by our upbringings.¹³⁹

We have already encountered Mackie’s objections to intuitionist Realism.¹⁴⁰ We simply do not possess the requisite sense to “intuit” Real, moral reasons. As such, we have no direct evidence of their existence. The only evidence we could have of their existence is inferential: inferring their existence as the best causal explanation of our observations of moral action.¹⁴¹ This is just the origin of Realism, to borrow Ryle’s phrase again.¹⁴² But, if the Realist suggests, as per the internalization defence,¹⁴³ that moral action is not caused by Real, moral reasons themselves, but by mental states about moral reasons, and if we have no good reason to believe that those mental states about moral reasons could only have come about as the causal effects of Real, moral reasons themselves, then we have no good reason to posit the existence of Real, moral reasons in the first place, and we ought therefore to reject Realism.

On the other horn of the dilemma, if the Realist maintains that moral action is caused by Real, moral reasons, then – Real, moral reasons being external to the agent – we end up with the absurd conclusion that moral action is action for which agents are not responsible. This is the Realist’s problem of moral responsibility,¹⁴⁴ and it is a problem of the Realist’s own making. It is, I will argue, a symptom of a category-mistake.

4.2.4. Diagnosis

I have argued that there are various points of analogy between Cartesian Dualism and Realism.¹⁴⁵ I now argue that there is yet a further point of analogy between this philosophy of mind and this metaethic, in that both are also category-mistaken.

The Cartesian Dualist posits mental substance in order to account for the distinction between intelligent and unintelligent behaviour, only the former of which is caused by thought.¹⁴⁶ However, this ‘solution’ only causes bigger problems for the Cartesian Dualist, because the characterization of thought as mental substance absurdly precludes any possibility of mental causation.¹⁴⁷ Likewise, the Realist posits Real, moral reasons in order to account for the distinction between moral and non-

¹³⁹ Harman 1977, p.4, as cited above.

¹⁴⁰ Mackie 1977, p.38, as referenced in s.3.3, above.

¹⁴¹ See s.3.4, above.

¹⁴² See s.3.5, above.

¹⁴³ See s.4.2.2, above.

¹⁴⁴ See s.4.2.1, above.

¹⁴⁵ See s.3, above, as well as s.4.2.3.

¹⁴⁶ See s.3.5, above.

¹⁴⁷ See s.4.1, above.

moral action, only the former of which is caused by Real, moral reasons – but the nature of those Real, moral reasons as world-regarding and external to the agent absurdly precludes the possibility of morally responsible moral action – that is, it precludes the possibility of moral action at all.¹⁴⁸ (And, the most promising solution to the Realist’s problem of moral responsibility, the internalization defence,¹⁴⁹ is unacceptable because it ultimately undermines its own Realism.¹⁵⁰) As Ryle puts a similar point: “The curious conclusion results that though [Real, moral reasons] were called in to explain our appraisals of actions, this explanation is just what they fail to provide.”¹⁵¹ And, as Ryle argues that Cartesian Dualism’s “curious” problem of mental causation is a symptom of a category-mistake, so I argue that Realism’s “curious” problem of moral responsibility is a symptom of a category-mistake.

To charge a theory with a category-mistake, recall, is not just to criticize a theory for its failure to explain some phenomenon; it is, rather, to make the radical criticism that the theory fails as it does *because* it fundamentally misconceives the phenomenon under explanation.¹⁵² That is, in the Cartesian Dualist’s case, Ryle’s charge of a category-mistake is not just the claim that the Dualist is unable to explain how mental states might cause intelligent behaviour; it is the claim that the Dualist is unable to do so because the Dualist goes wrong right at the beginning, in the presumption that intelligent behaviour is to be differentiated from unintelligent behaviour by the former’s *cause*. The Cartesian Dualist’s category-mistake, according to Ryle, is that

The differences between the physical and the mental were [...] represented as differences inside the common framework of the categories of ‘thing,’ ‘stuff,’ ‘attribute,’ ‘state,’ ‘process,’ ‘change,’ ‘cause’ and ‘effect’. Minds are things, but different sorts of things from bodies; mental processes are causes and effects, but different sorts of causes and effects from bodily movements. And so on.¹⁵³

That is, the Cartesian Dualism’s category-mistake is not positing mental states as causes of intelligent behaviour of the wrong sort of substance (mental substance) – but positing them as *causes* at all.

The upshot of this is that a similar criticism will apply to other philosophies of mind too, even those that – unlike Cartesian Dualism – do not posit mental substance. For instance, since Physicalism also treats the mind as if it belongs to the category “of ‘thing,’ ‘stuff,’ ‘attribute,’ ‘state,’ ‘process,’ ‘change,’ ‘cause’ and ‘effect’”¹⁵⁴ – where the mind is, quite literally, the sort of ‘thing’ that is a *brain* – Ryle also charges Physicalism with a category-mistake, even though Physicalism construes mental

¹⁴⁸ See s.4.2.1, above.

¹⁴⁹ See s.4.2.2, above.

¹⁵⁰ See s.4.2.3, above.

¹⁵¹ Ryle 1949/2009, p.53. Ryle actually makes these remarks with regard to volition.

¹⁵² See s.4, above.

¹⁵³ Ryle 1949/2009, p.9, as cited in ss.2 and 4 above.

¹⁵⁴ Ryle 1949/2009, p.9, as cited above.

states as physical states.¹⁵⁵ And, Ryle levels the same criticism at Idealism too, for the same reason: even though Idealism posits that the mind is of the same substance as everything else in existence – namely, mental substance – it too construes the mind as belonging to the category “of ‘thing,’ ‘stuff,’ ‘attribute,’ ‘state,’ ‘process,’ ‘change,’ ‘cause’ and ‘effect’”, and that is Idealism’s category-mistake.¹⁵⁶ If Ryle’s diagnosis of the category-mistake is correct, therefore, then it follows that it is not only Cartesian Dualism, but all philosophies of mind that construe mental states as cause of intelligent behaviour – which will comprise almost all of the traditional Philosophy of Mind¹⁵⁷ – that are guilty of this category-mistake.

So, as Ryle claims that Cartesian Dualism’s – and, indeed, all of traditional Philosophy of Mind’s – category-mistake is to think of mindedness in terms “of ‘thing,’ ‘stuff,’ ‘attribute,’ ‘state,’ ‘process,’ ‘change,’ ‘cause’ and ‘effect’”,¹⁵⁸ DeLapp notes that “the default language” of the Realist is – strikingly similarly – that “of ‘furniture,’ ‘structural elements,’ ‘objects,’ and entities similar to the entities studied by physicists”.¹⁵⁹ Therefore, just as Ryle claims that the Cartesian Dualist’s category-mistake is to conceive of mindedness as the effect of a special kind of causal thing – the mental state – I argue that the Realist’s category-mistake is to conceive of moral action as the effect of a special kind of causal thing – *the moral reason*. Julia Tanney, a Ryle scholar, hints at this application of Ryle’s category-mistake argument to the subject matter of Metaethics. As she writes:

Compare: an agent acts for, or because of, certain reasons rather than others when those reasons *cause* her action, where reasons are now construed as mental states and triggering events thought to be (or to be realized or instantiated in, or dependent upon) some physical (presumably neurophysiological) states or other.¹⁶⁰

So, too, would it be a category[-]mistake to imagine that reasons [...] play a role in the explanation of action on (almost) the same level as the internal processes that have a role to play in the explanation of the body’s motions [...].¹⁶¹

Like Tanney, therefore, I argue that Realism commits a category-mistake in “imagin[ing] that reasons [...] play a role in the explanation of action” “when those reasons *cause* [an agent’s] action”. And, my justification for this is that the Realist suffers the symptom of the problem of moral

¹⁵⁵ Ryle 1949/2009, p.12.

¹⁵⁶ Ryle 1949/2009, p.12.

¹⁵⁷ In her Foreword to the 60th anniversary edition of Ryle’s book, Julia Tanney (2009, pp.xx-xxi) charges even Functionalism with having committed a category-mistake.

¹⁵⁸ Ryle 1949/2009, p.9, as cited above.

¹⁵⁹ DeLapp 2009, p.5, citing Mackie 1977, pp.17 and 23, and Brink 1989, p.6.

¹⁶⁰ Tanney 2009, pp.xlii-xliii, fn.73, referring the reader to Davidson (1963) as a prime example of category-mistakenly construing reasons as causes.

¹⁶¹ Tanney 2009, p.xlv.

responsibility¹⁶² (and that the most convincing Realist defence against that criticism lands the Realist in a self-defeating dilemma¹⁶³).

There may, however, be Realist responses to my allegation of the problem of moral responsibility, besides the internalization defence, which I have not considered. (Likewise, there are versions of Dualism – several, in fact – that purport to solve the Dualist’s problem of mental causation while nevertheless retaining the core Dualist insight that mental states are of a different substance to physical states. Yet Ryle considers none of these.) If there are such further responses available, they are certainly worthy of consideration. However, I will not take them up here. This is because my present interest, recall,¹⁶⁴ is not so much to criticize Realism at all. Instead, it is to lay the groundwork, by way of drawing a contrast with Realism, for the claim that, on a Ryle-inspired reading thereof, Darwall’s Second-Personal approach constitutes a radically different account of moral action, one that does not attempt to give an account thereof as the effect of a special kind of causal thing, where that causal thing is the moral reason. To this end, I am satisfied if I have managed to at least make a case for serious consideration of the claim that Realism is category-mistaken.

If I am correct that it is a category-mistake to think of moral action as the causal effect of moral reasons, then it will follow that all metaethics conceiving of moral action in this way will be category-mistaken – regardless of the particular metaethic’s posited nature of the moral reason: Real, Egoistic, commanded by God, bipolar, *etc.* That is, as Ryle takes his criticism to apply not only to his explicit target, the Cartesian Dualist, but also to Physicalists as well as Idealists, I hope that my present criticism – that it is a category-mistake to conceive of moral action as the causal effect of moral reasons – is applicable not only to my explicit target of this chapter, the Realist, but also to all traditional Moral Philosophy (as reflected in the Prichardian title of this chapter¹⁶⁵). A Ryle-inspired reading of *The Second-Personal Standpoint* will escape this charge of category-mistakenness, I will argue, because it construes moral action not as the result of any special cause, but simply as a way of relating to people, as I begin to explain, below.

5. Ways of behaving

It is a category-mistake, Ryle claims, of the Cartesian Dualist to presume that the difference between intelligent and unintelligent behaviour is explained by the difference in their causes, where the

¹⁶² See s.4.2.1, above.

¹⁶³ See ss.4.2.2 and 4.2.3, above.

¹⁶⁴ See ss.1 and 2, above.

¹⁶⁵ See s.1, above.

former is the causal effect of thought, and the latter is not.¹⁶⁶ So, if this is the category-mistaken approach to mindedness, what is Ryle's category-*unmistaken* alternative? If the difference between intelligent and unintelligent behaviour isn't a difference of their causes, then what is the difference?

Take, for instance, my favourite of Ryle's various examples: team-spirit.¹⁶⁷ What is the difference between playing with team-spirit and playing without it? Naturally, the difference is that they are simply two different *ways of playing*. Likewise, Ryle claims with regard to intelligent behaviour, it is not behaviour with a special, mental cause; it is just a *way of behaving*: behaving intelligently, rather than unintelligently. Behaving intelligently differs from behaving unintelligently in the same way in which playing with team-spirit differs from playing without team-spirit – and that is not by a difference in their respective causes. In Ryle's own words:

when we characterise people by mental predicates, we are not [*contra* Cartesian Dualism] making untestable inferences to any ghostly processes occurring in streams of consciousness which we are debarred from visiting: we are describing the *ways in which these people conduct parts of their predominantly public behaviour* [...]¹⁶⁸

And:

when we speak of a person's mind, we are not [*contra* Cartesian Dualism] speaking of a second theatre of special-status incidents, but of certain *ways in which some of the incidents of his one life are ordered*.¹⁶⁹

It is a mistake – a category-mistake – of Cartesian Dualism to posit special, immaterial causes in order to explain the difference between intelligent and unintelligent behaviour, because the difference between intelligent and unintelligent behaviour is just a difference of ways of behaving. And, just as we can know in what way an agent behaves – for instance, with team-spirit, or without team-spirit – without having to make any postulations about causes, so too do we know the difference between behaving intelligently and behaving unintelligently, without making any postulations about the cause of either way of behaving. When we understand that intelligent behaviour is just a different way of behaving to behaving unintelligently, we will see that there is no need to infer to their respective causes in order to explain the difference between them; the difference between them is already captured by their being different ways of behaving.

Category-mistakenly, the Cartesian Dualist assumes that what sets intelligent behaviour apart from unintelligent behaviour is its special cause: thought – yet, absurdly, the Cartesian Dualist finds him-/herself unable to explain how thought might actually cause intelligent behaviour. The Cartesian Dualist's problem of mental causation is a symptom of a category-mistake because it is a problem of

¹⁶⁶ See s.4.2.4, above.

¹⁶⁷ Ryle 1949/2009, p.7, as cited in s.2 above.

¹⁶⁸ Ryle 1949/2009, p.50 (emphasis added).

¹⁶⁹ Ryle 1949/2009, p.149 (emphasis added).

the Dualist's own making. As Ryle points out, we *already knew* the difference between intelligent and unintelligent behaviour – long before theorizing about its (impossible) cause – just as we know the difference between other ways of behaving, such as playing with team-spirit or without:

The question, 'How do persons differ from machines?' arose just because everyone *already knew* how to apply mental-conduct concepts before the new causal hypothesis was introduced. This causal hypothesis could not therefore be the source of the criteria used in those applications.¹⁷⁰

This general Rylean approach is steadily gaining ground in the Philosophy of Mind. With reference to the problem of other minds, for instance, Anita Avramides argues that the problem of other minds would not even have occurred to us, did we not already have an affirmative answer to what she calls the "conceptual" problem of other minds: whether there are such things as other minds in the first place – that is, very much like Ryle, Avramides argues that we *already know* that there are other minds, before the problem of other minds ever occurs to us.¹⁷¹ In a similar manner, with reference to folk psychology, Shaun Gallagher has advocated Interaction theory as the proper solution to the deadlocked Simulation theory/Theory theory debate, arguing that we *already know* – indeed, that even neonates already know – how to discern the mental states of others, antecedently of either simulating or theorizing, also arguing that both Simulation theory and Theory theory actually *presuppose* his Interaction theory.¹⁷² This general Rylean approach, therefore, has been put to good work in other areas of the Philosophy of Mind.

And, it is interesting to note that Gallagher's Interactionist approach has been received in the literature as a *second-personal* approach to folk psychology.¹⁷³ Gallagher's Interactionism presents a radically different approach to the either first-personal Simulationism or third-personal Theory-theory: it is not a first-, not a third-, but a *second-personal* approach. Gallagher's Interactionism is a second-personal folk psychological account that takes the general Rylean approach, described above, by arguing that Simulation theory and Theory-theory overlook the fact that we already know how to discern other's mental states. This gives me hope that the same general Rylean approach might also be put to good work in the field not of Philosophy of Mind but of Metaethics – and, as Gallagher employs it in the Philosophy of Mind, to present a radically different, *second-personal* approach to Metaethics.¹⁷⁴ That is to say, the trend of application of Ryle's general strategy in

¹⁷⁰ Ryle 1949/2009, p.11 (emphasis added).

¹⁷¹ Avramides 2001.

¹⁷² For perhaps the fullest statement, see Gallagher's (2005) book, although he has produced a wealth of papers on the topic.

¹⁷³ See Dullstein 2012.

¹⁷⁴ At this point, I should briefly hint at one more criticism of Darwall, although I do not consider it very damaging. Darwall (2006) devotes a chapter of *The Second-Person Standpoint* to 'The psychology of the second person' (Ch.7, pp.151-180). One commentator – a professor of Law – cites with approval Darwall's "naturalistically sound moral psychology that is attentive to recent empirical developments" (Kar 2007, p.882),

various areas in the Philosophy of Mind ignites my optimism about a Ryle-inspired reading of *The Second-Person Standpoint*. This Ryle-inspired reading of *The Second-Person Standpoint* will be the occupation of the rest of this project.

As Ryle has argued that Cartesian Dualism is category-mistaken in positing a special cause – thought – of intelligent behaviour, when intelligent behaviour is just a way of behaving different to behaving unintelligently, so I argue that Realism is category-mistaken in positing a special cause – the Real, moral reason – of moral behaviour. Rather, I will argue, moral action is just *a way of relating to other people*, and, as such, we need not posit a special cause in order to explain how it differs from non-moral behaviour.

Interestingly, Korsgaard comes close to making a similar claim:

[A] basic feature of the [C]onsequentialist outlook still pervades and distorts our thinking: the view that the business of morality is to *bring something about*. [...] More importantly, moral philosophers have persistently assumed that the primal scene of morality is a scene in which someone does something *to or for* someone else. This is the same mistake that children make about another primal scene.¹⁷⁵ The primal scene of morality, I will argue, is not one in which I do something to you or you do something to me, but one in which we do something together. The subject matter of morality is not what we should bring about, but *how we should relate to one another*.¹⁷⁶

I argue something similar. Like Korsgaard, I think that morality is a way of relating to people, and that much of Metaethics – Realism, in particular – has neglected this point. However, unlike Korsgaard, I do not think that the problem is a spreading Consequentialist infection *per se*; rather, I think it is the deeper category-mistake of conceiving of moral action in terms of causation by moral reasons. This is a category-mistake that I think a Ryle-inspired reading of the Second-Personal approach avoids (comes closest to avoiding), which will be the subject of the following chapter.

but others find Darwall's "psychology of the second person" questionable. Lavin (2008) treats Darwall's appeal to empirical psychology as missing the point of his own project, and Fleischacker (2009, p.118), similarly, describes the digression into empirical psychology as part of *The Second-Person Standpoint's* "tak[ing] such a winding path to its conclusions". I am inclined to side with Lavin and Fleischacker on Darwall's inclusion of this Psychological material. Firstly, Darwall's frequent appeals to "empathy", apparently align him with a Simulationist approach. In the light of Gallagher's work, however, it seems clear that Interaction theory is a much better ally for Darwall's Second-Personal account than Simulation theory. Furthermore, if Darwall wishes to appeal simultaneously to Fichte's transcendental deduction of the intersubjective nature of the *I* (see below, and Ch.7), then his appeal to empirical data has no place in the argument whatsoever.

¹⁷⁵ Korsgaard's reference to the 'primal scene' is provocative. The term may connote a natural habitat: the area to which a phenomenon is native. However, 'primal scene' is also a Psychoanalytic term – and this is probably what Korsgaard is alluding to with "This is the same mistake that children make about another primal scene." In the Psychoanalytic sense, the primal scene is a child's first witnessing (or imagination) of the sex act (usually between the child's parents), often incorrectly and traumatically perceived as an act of violence ("What is Daddy doing to Mommy?"), rather than as a mutually consensual, cooperative engagement (*cf.* "What are Mommy and Daddy *doing together?*").

¹⁷⁶ Korsgaard 1993/1996, p.275; emphasis added.

Before we turn to the next chapter, however, let me draw one more point of analogy between Ryle and the Second-Personal approach. Above, I noted Ryle's claim that, *contra* Cartesian Dualism, we need not posit the special cause of intelligent behaviour because we already know what the difference is between behaving intelligently and behaving unintelligently. And, not only does Ryle claim that we already know the difference between behaving intelligently and behaving unintelligently, in even starker contrast to the Cartesian picture (and its supposition of the privacy of consciousness), Ryle also claims that we know how to recognize mindedness in *others* even before we recognize it in ourselves, that is, even before we are self-conscious. In his words:

[T]here is nothing mysterious or occult about the range of higher order acts and attitudes, which are apt to be inadequately covered by the umbrella-title 'self-consciousness'. They are the same in kind as the higher order acts and attitudes exhibited in the dealings of people with one other. Indeed the former are only a special application of the latter and are learned first from them [as follows:]¹⁷⁷

just as he had earlier acquired not only the ability, but also the inclination to direct higher order acts upon the acts of *others*, so he now becomes prone, as well as competent, to do the same upon his *own* behaviour; and just as he had earlier learned to cope not only with the particular performances of *others*, but also with their dispositions to conduct such performances, so he now becomes in some degree both able and ready to take steps, theoretical and practical, about his *own* habits, motives and abilities.¹⁷⁸

This Rylean claim, that we must be conscious of others' mindedness before we are conscious of our own, is strongly reminiscent of the claim by "one of Darwall's favo[u]rite sources", Fichte,¹⁷⁹ that:

Being an individual I, placing before oneself an end, is something a rational being must be educated to do, through the influence of another rational being. [...] Freedom is possible only through upbringing (*Erziehung*) through the influence of other free beings [...].¹⁸⁰

That is, interestingly,¹⁸¹ both Ryle and Fichte make the similar claims that an individual will develop into a full-blooded agent – self-conscious, free, and rational – “through the influence of” other

¹⁷⁷ Ryle 1949/2009, p.177.

¹⁷⁸ Ryle 1949/2009, pp.173-174 (emphases added).

¹⁷⁹ Fleischacker 2009, p.22.

¹⁸⁰ Wood 2006, p.74 (emphasis removed), referencing Fichte 1845/1970, vol.3, pp.39–40.

¹⁸¹ The parallel is “interesting” when one thinks of Fichte as a Continental, German Idealist and Ryle as British, Analytic, ‘ordinary language’ philosopher, since these are two schools of thought one would not expect to see much similarity between at all. However, it's not clear to me that this is the right way to understand Ryle – a philosopher I think is commonly misunderstood. He is definitely not the Philosophical Behaviourist he is commonly misunderstood to be (see Tanney 2009, pp.x and xxiv; Ayer 1970; and Ryle's own section titled ‘Behaviourism’, 1949/2009, pp.300-303). Tanney (2009, p.ix) describes Ryle as “on a different, somewhat idiosyncratic (and difficult to characterise), philosophical track”. Indeed, according to Wikipedia (2013), Ryle even said of his *Concept of Mind* that it “could be described as a sustained essay in phenomenology, if you are at home with that label” – suggesting that he comes much closer to the Continental tradition than he is commonly understood to do (although note, however, that this quotation is not referenced by Wikipedia). (Interestingly, as a further point of analogy between second-personal approaches in Metaethics and in the Philosophy of Mind, Gallagher (2005), likewise, grounds his Interaction theory in the work of Phenomenologists such as Merleau-Ponty (1962).)

agents. Insofar as Darwall's Second-Personal account depends on Fichte's Point,¹⁸² therefore, this is a further point of analogy between Ryle and Darwall.

Moreover, Fichte's argument for the above claim will mirror very closely the same argument I have provided in this chapter for the category-mistakenness of Realism. As I have argued that Realism is category-mistaken because it suffers the symptom of the absurd problem of moral responsibility – unable to account for moral action as free action¹⁸³ – Fichte will argue that a world-regarding Realism cannot account for our *freedom*; according to Darwall's reading of Fichte, only a Second-Personal account can do this. This claim will follow as a result of Fichte's transcendental deduction of the intersubjective nature of the *I*, which is the subject of the following chapter.

¹⁸² See my Ch.1, s.6.1.2.

¹⁸³ See s.4.2, above.

7.

A different approach

The main objection of this project has been that Darwall’s argument for universal, reciprocal Morality as Equal Accountability doesn’t even get started, because we do not have any good reasons for believing that there actually are such things as second-personal reasons. To this end, I examined possible accounts of the validity of second-personal commands, in the hope that it would vindicate the intuition that a second-person may give a first a reason – so long as that reason is *valid* – yet no acceptable account was uncovered.¹ I also examined Darwall’s Mill’s-Point argument, which infers the existence of second-personal reasons as the best explanation of our practices of holding each other accountable; however, I found that, where Darwall’s own Second-Personal account struggles to account for the broader aspects of Mill’s Point,² Realism fares much better,³ and as such Darwall’s Mill’s-Point argument for second-personality fails. As it stands, we have no good grounds on which to accept that there actually are such things as second-personal reasons, and so Morality as Equal Accountability does not follow.

In this final chapter, I will examine one last argument for the existence of second-personal reasons, which – like Fichte’s Point⁴ – comes from Fichte. The argument is the transcendental deduction that an agent’s freedom presupposes a “summons” (to use Fichte’s term) or a second-personal reason (to use Darwall’s term). As Darwall puts it:

Second-personal address both makes possible and simultaneously makes us aware of a species of practical freedom that lacks any analogue in the theoretical realm.⁵

That is, the possibility of a free agent presupposes that of second-personal reasons. So, if there are any free agents, it follows that second-personal reasons must exist – and, since there surely are free agents, there must be such things as second-personal reasons. Call this ‘Fichte’s Argument’.

Fichte’s Argument is to be distinguished from Darwall’s Fichte’s Point,⁶ the latter of which builds upon the conclusion of Fichte’s Argument, namely, that second-personal reasons exist. Once Fichte’s Argument establishes that there are such things as second-personal reasons, then Fichte’s

¹ See Ch.2, s.2.3.

² See Ch.3 and 4.

³ See Ch.5.

⁴ See Ch.1, s.6.2.1.

⁵ Darwall 2006, p.22.

⁶ See Ch.1, s.6.2.1.

Point should go through, and Morality as Equal Accountability should follow. Fichte's Point, recall, is that

any second-personal claim or "summons" (*Aufforderung*) presupposes a common competence, authority, and, therefore, responsibility as free and rational, a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally.⁷

That is, if Fichte's Argument is successful in establishing that there are such things as second-personal reasons, then it should follow that all agents share "a common competence, authority, and, therefore, responsibility", that is, Darwall's account of Morality as Equal Accountability should follow.

I take it that Fichte's Argument is crucial to Darwall's project: without it, I don't see why we should accept that second-personal reasons exist, nor, therefore, why we should accept Morality as Equal Accountability. Darwall, likewise, appears to regard Fichte's Argument as key to his own, and notes it as such in his initial setting-out of *The Second-Person Standpoint's* argument in Chapter 1 of that book.⁸ Furthermore, as Lavin notes, "in this new book [*The Second-Person Standpoint*] Darwall turns against his Kantian roots,"⁹ and it is by way of Kant's critic, Johann Gottlieb Fichte, that Darwall departs from Kant. However, even though Fichte is noted (by at least one reviewer) to be "one of Darwall's favo[u]rite sources",¹⁰ Fichte's Argument itself has received correspondingly little attention from reviewers. One notable exception to this general trend is Owen Ware,¹¹ whose excellent work on the nature of the Fichtean summons I will appeal to, below. Following Ware, I will argue that Fichte's Argument does not actually commit us to what Darwall thinks it commits us to – that is, I will argue that Fichte's Argument does not demonstrate the existence of second-personal reasons.

I begin this chapter with an examination of Fichte's Argument for the existence of second-personal reasons. I will contend, however, that, although Fichte's Argument demonstrates that there must be such a thing as the summons, a Fichtean summons is not identifiable with a Darwallian second-personal reason. As such, Fichte's Argument again leaves us without any good grounds on which to accept that there are indeed such things as second-personal reasons, and so Fichte's Point does not subsequently entail Morality as Equal Accountability.

I will go on to argue, however, that Fichte's Argument does entail something – which, as I will explain, is a certain *way of relating to people*. That is, Fichte's Argument entails nothing about the

⁷ Darwall 2006, p.21 (see also p.246), referencing Fichte 1791/2000; as cited in my Ch.2, s.8.1.

⁸ Darwall 2006, pp.20-23.

⁹ Lavin 2008, as cited in the introduction to my Ch.6.

¹⁰ Fleischacker 2009, p.22.

¹¹ Ware 2009.

nature of moral reasons as causal items – which, as I argued in the preceding chapter, is a category-mistaken approach to moral action.¹² Reading Fichte’s Argument as bringing into relief the way of relating to people constitutive of moral action will give me occasion to explore the Ryle-inspired reading of *The Second-Person Standpoint*, anticipated by the preceding chapter.¹³ And, as we will now see, Fichte’s Argument mirrors the Rylean argument I presented in that previous chapter: world-regarding reasons can’t account for our freedom; only Second-Personal reasons can – or, at least, only a Fichtean summons can. The remainder of this chapter will be devoted to exploring how *The Second-Person Standpoint* might have been read entirely differently from that reading assumed by the first five chapters of this project.

Responses to the claims of this chapter have been interestingly mixed. While some have balked at the suggestion that Darwall might be advocating anything as radical as the Rylean idea that moral action is just a way of relating to people, others have remarked that the suggestions to follow are pretty clearly an accurate description of Darwall’s project all along. I think the correct reading is somewhere in between these two extremes. As I have remarked before,¹⁴ I think that *The Second-Person Standpoint* represents something like a butterfly emerging from the chrysalis: halfway to being an entirely new metaethic, the likes of which we have never seen before, but also, as Lavin has put the point, not quite able to “entirely break with [its] past: it hangs on in Darwall’s sense of the questions and methods of moral philosophy”.¹⁵ Because it is apparently a controversial topic, I will motivate why I think Darwall almost presents a radically different, Ryle-inspired account of moral action – but not quite: what portions of the text support the Ryle-inspired reading, and which undermine it.

1. Fichte’s Argument

Fichte’s Argument is a transcendental deduction. It takes as its starting point the fact that the *I* is a free agent and is conscious of him-/herself as such – or, as Fichte puts it, “posits” him-/herself as such.¹⁶ We have already seen how Darwall tries to employ the Fichte’s-Point transcendental deduction, according to which the giving of a second-personal reason entails the reciprocal, universal authority of the first and second persons,¹⁷ but the starting point of that transcendental deduction, I complained, is far from “so basic that it is very difficult to subject it to s[c]eptical

¹² See Ch.6.

¹³ See Ch.6.

¹⁴ See the introduction to Ch.6.

¹⁵ Lavin 2008.

¹⁶ Fichte 1845/1970 vol.1, p.492, as cited by Wood 2006, p.70.

¹⁷ See Ch.1, s.6.2.1.

doubt”:¹⁸ why should we believe that there are any such things are second-personal reasons at all?¹⁹

The present Fichte’s-Argument transcendental deduction, on the other hand, fares much better: I surely am free, and I know it. What remains to be shown, then, is that it is a presupposition of my freedom that a second person gives me a second-personal reason.

According to Fichte, the free *I* not only exists, but exists as an active being. As such, the *I* must posit itself in opposition to that which is passively given to it: the “not-*I*”, namely, the world. The world is what the *I* strives against. As Allen Wood describes it, the *I*’s “willing takes the form of a striving against a material world on which it acts.”²⁰ So, in its consciousness of itself as a free agent, the *I* must first posit the not-*I*, the world.

However, the world is not a suitable candidate for a source of the free agent’s reasons. The explanation of this is essentially the same argument I have provided for Realism’s problem of moral responsibility:²¹ “the world [operates by] causal influence, through which one may be coerced, or manipulated, or conditioned to behave, but not enabled to act freely or autonomously”;²² the world presents “facts that causally necessitate what we do, restricting our freedom to do otherwise [...]”.²³ Because the world is what the agent strives *against* in acting, the world is not what gives the agent his/her self-conscious freedom. For that, we require another kind of not-*I*: the Other.

If I am to be conscious of myself as a free agent, I must be conscious of myself as *this* agent, *this I*. And, that will require the positing of the Other also. The Other is like this *I* in standing in opposition to the world, but the Other *I* is not *this I*, not *me* – and that is how I come to be conscious of myself as *this* agent, and not another one. So, Fichte concludes: “No thou, no I; No I, no thou”.²⁴ That the positing of the Other is a transcendental precondition of free agency is Fichte’s – and Darwall’s – basic point of departure from Kant. As Fichte writes in objection to Kant:

Nor can Kant understand by this pure apperception the consciousness of our individuality, or confuse the one with the other; for the consciousness of individuality is necessarily accompanied by another consciousness, that of a ‘thou’, and is possible only on this condition.²⁵

¹⁸ Wood, 2006, p.67, as cited in my Ch.2, s.2.1.

¹⁹ See Ch.2, s.2.1.

²⁰ Wood 2006, p.70, referencing Fichte 1845/1970, vol.1, pp.65, 261–271, and 285–291, and 1962, pt.4, vol.2, p.77.

²¹ See Ch.6, s.4.2.1.

²² Wood 2006, p.75.

²³ Wood 2006, p.74.

²⁴ Fichte 1845/1970, vol.1, p.189, as cited by Wood 2006, p.70.

²⁵ Fichte 1845/1970, pt.1, p.476, as cited by Wood 2006, p.71.

Fichte thinks that Kant's account is lacking in its failure to posit the Other as a transcendental precondition of the free agent.²⁶ (And, I add, it is because Kant fails to posit intersubjectivity as Fichte does that Neo-Kantian accounts are faced with the task of demonstrating that your personhood is valuable to me as my own personhood is valuable to me, by either the Generality or Analogical arguments.²⁷)

According to Fichte, to

be aware of my individual *I* is not so much activity in which I have already engaged as possible activity in which I might engage. For me, my individuality consists not merely, and not fundamentally, in facts that distinguish me from others but in possibilities of acting through which I actively determine who I am.²⁸

For Fichte, the possibility of a free agent's having done otherwise is given by the Other's *Aufforderung* or "summons".²⁹ In summoning me, the Other calls me to perform a certain action or be a certain way. These represent the possible ways I, as a free agent, might be or might have been. When the Other summons me, I as an agent have the freedom to accept or reject that summons:

Either I act in accordance with the summons or I do not act in accordance with it. If I have understood this summons, I can, of course, still decide not to act [as the summons represents]³⁰

And, therein consists my freedom as an agent: knowing that I could have done otherwise by way of the Others' summons presented to me, which I have the freedom to accept or reject. As Wood summarizes Fichte's argument:

The not-I (as the material world) resists our ends or may be brought into conformity with them. It may compel us to take one means to them rather than another, or it may make them impossible. But it cannot be the source from which we draw the concept of those ends. A summons, however, is precisely an object of consciousness which makes the concept of an end possible. It is, if you like, another species of not-I, a kind of not-I that does not resist our freedom, but on the contrary, is what makes freedom possible.³¹

That is to say, a summons is "what makes freedom possible". Since I surely am a free agent, then, it follows by Fichte's Argument that the summons – or, to use Darwall's term, the second-personal reason – does indeed exist.

²⁶ Wood (2006), interestingly, regards this feature of (Neo-)Kantianism as a *Cartesian* hangover, thereby construing Fichte as reacting to the Cartesian *cogito*. I take this as some affirmation of my suggestion that Darwall's Fichtean appeal constitutes a category-unmistaken alternative to traditional Metaethics, as Ryle's Dispositional account constitutes a category-unmistaken alternative to traditional, *Cartesian* Philosophy of Mind (see Ch.6).

²⁷ See Ch.1, s.5.

²⁸ Wood 2006, p.72, paraphrasing Fichte 1845/1970, vol.4, pp.218–219; italics added.

²⁹ Fichte 1962, pt.4, vol.2, p.179, as referenced by Wood 2006, p.72.

³⁰ Fichte 1962, pt.4, vol.2, p.179, as cited by Wood 2006, p.73.

³¹ Wood 2006, pp.73-74.

2. Towards a different approach

However, I question whether Darwall's second-personal reason is indeed the same thing as a Fichtean summons, and whether Darwall is correct that "[t]he most perspicuous interpretation of Fichte's idea is in terms of second-personal reasons."³² There seem to me to be significant differences between the two. Although Fichte's Argument establishes the existence of the Fichtean summons, I question whether it establishes the existence of the Darwallian second-personal reason.

2.1. The second-personal reason vs. the summons

If Fichte's Argument is successful, then it demonstrates that there are such things as Fichtean summonses. Since, Fichte argues, we are aware of our free agency only because of our freedom to accept or reject a summons – and since we surely are aware of our free agency – then it follows that we must have been "summoned".³³

But, what is a Fichtean summons? Fichte's original, German term for the summons is '*Aufforderung*'. Wood helpfully sheds light on this term and its proper translation:

In order to understand this argument [i.e. 'Fichte's Argument'³⁴], we need to understand more precisely what Fichte means by the word 'summons' (*Aufforderung*). This term in German is broader in its meaning than the English 'summons' – which perhaps suggests a command of some kind, such as a legal subpoena to appear in court. The term *Rechtsaufforderung* does cover that notion, but *Aufforderung* means something much less official and usually not in the least tinged with coercion. In many contexts, it is better to translate it as 'invitation', as in Carl Maria von Weber's *Aufforderung zum Tanz*, op. 65.³⁵ German dictionaries report that the meaning of *auffordern* ranges all the way from *bitten* (request or beg) to *verlangen* (demand or require).³⁶

That is, according to Wood, '*Aufforderung*' has a wide range of meanings, but it is best interpreted in "many contexts" as 'invitation' – and, as Wood goes on to explain, Fichte's Argument is one such context. Ware, likewise, interprets the term as meaning something more like 'invitation'.³⁷

Darwall, on the other hand, apparently interprets the term as falling much further towards the '*verlangen*' – 'demand' – end of Wood's "spectrum". According to Darwall, recall, to give a first person a second-personal reason is to "lay a claim or address a purportedly valid *demand*".³⁸ And, Darwall must think that second-personal reasons are identifiable with Fichtean summonses – otherwise he wouldn't employ Fichte's Argument as an argument for the conclusion that *second-*

³² Darwall 2006, p.256.

³³ See s.1, above.

³⁴ See s.1, above.

³⁵ This is a piano piece written by Carl Maria von Weber, the title of which is translated into English as: "Invitation to the Dance".

³⁶ Wood 2006, p.73, referencing Fichte 1962, pt.4, vol.2, p.179.

³⁷ Ware 2009.

³⁸ Darwall 2006, p.7, as cited in my Ch.1, s.1.2. (emphasis added).

personal reasons exist³⁹ – so, it follows that Darwall must, unlike Wood and Ware, be interpreting ‘*Aufforderung*’ as (something more like) ‘demand’.

However, Darwall’s interpretation is not well-motivated.⁴⁰ This is because Fichte’s Argument itself requires that ‘*Aufforderung*’ be interpreted as falling closer to the ‘invitation’ end of Wood’s spectrum. As the Wood excerpt, above, continues:

Fichte’s use here seems rather closer to the former [‘invitation’] end of the spectrum than to the latter [‘demand’ end of the spectrum]. For he specifically emphasizes that an *Aufforderung* (as he means it) leaves us free either to do or not do as we are summoned (or invited) to do.⁴¹

That is, Fichte’s Argument, as examined above,⁴² is that we are aware of our free agency only because of our freedom to accept or reject a summons. And, this clearly could not be the case unless we were free to accept or reject Fichtean summonses. While invitations are the sorts of ‘summonses’ we are free to accept or reject, demands apparently are not. As such, Fichte’s Argument itself requires that we interpret the ‘*Aufforderung*’ or ‘summons’ as more like an invitation than a demand, as indeed Wood and Ware do.

And, as Ware points out,⁴³ this kind of Fichtean summons – that is, as an invitation – is really nothing like the Darwallian second-personal reason. To get an idea of the major difference between the two, consider Darwall’s primary example of the second-personal reason, presented in Darwall’s foot-treader scenario.⁴⁴ According to Darwall, when you respond to the foot-treader second-personally, you “give [the foot-treader] a distinctive kind of (normative) reason for acting”⁴⁵ – namely, to get off your foot – when you “lay a claim” against him.⁴⁶ This second-personal reason you give the foot-treader is a moral reason. And, moral reasons are overriding;⁴⁷ as Darwall puts it in Mill’s Point: “When we are morally obligated, we are not morally free to act otherwise.”⁴⁸ It follows, therefore, that, when you give the foot-treader a second-personal (moral) reason to remove his foot from on

³⁹ See s.1. above.

⁴⁰ In his defence, although it does not make an appearance in *The Second-Person Standpoint*, Darwall does in an earlier work provide an argument for his stronger claim that, upon a summons, “one is not free *not* to recognize and respect” the Other (Darwall 2005, p.110, as cited by Ware 2009, p.272). However, Ware (2009) argues persuasively that “[t]he problem with Darwall’s analysis, then, is that it fails to distinguish between the two kinds of recognition in Fichte’s account” (p.274). For simplicity’s sake, I have in this chapter raised only Fichte’s passive notion of recognition – rather than the active notion that Darwall needs, but which he cannot simply help himself to. See Ware (2009) for an excellent discussion of this point.

⁴¹ Wood 2006, p.73, referencing Fichte 1962, pt.4, vol.2, p.179.

⁴² See s.1, above.

⁴³ Ware 2009, p.265.

⁴⁴ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

⁴⁵ Darwall 2006, p.4 (emphases removed).

⁴⁶ Darwall 2006, p.7.

⁴⁷ Darwall 2006, pp.26-27, as referenced in the introduction to my Ch.3.

⁴⁸ Darwall 2006, p.27, as cited in the introduction to my Ch.3.

top of yours, he is “not morally free to act otherwise.” So, where the Fichtean “*Aufforderung*” is – like an invitation – the kind of summons we are free to accept or reject at will, the Darwallian second-personal reason is *not*. As such, the Darwallian second-personal reason cannot be identified with the Fichtean summons, despite Darwall’s apparent assumptions to the contrary.

To be fair, Darwall does – intentionally, I assume – phrase Fichte’s Point carefully, as the claim that “any second-personal *claim* or “*summons*” (*Aufforderung*) presupposes” reciprocal, universal authority – rather than as the claim that any *second-personal reason* does so⁴⁹ – where this leaves open the correct interpretation of “*Aufforderung*” as a mere invitation rather than a demand. And, *The Second-Person Standpoint* is peppered with descriptions of the second person merely *attempting* to give a first a reason.⁵⁰ Such claims are more consistent with Fichte’s Argument and the correct interpretation of “*Aufforderung*”.

However, these more carefully worded claims do not support Darwall’s Second-Personal conclusions: it does not follow on the basis of a second person’s merely *attempting* to give a first a reason that those two persons share a mutual and reciprocal authority; indeed, the fact that the first may *reject* the second’s attempt seems to entail precisely the opposite: that the first has an authority the second lacks. I do not see, then, how this more accurate account of Fichte’s Argument could support Darwall’s desired conclusion that there exist second-personal reasons, where these are moral reasons according to which “we are not morally free to act otherwise”.⁵¹ What Fichte’s Argument supports is the existence of the *summons*, which is importantly different from a second-personal reason.⁵² Gideon Yaffe raises a similar complaint against Darwall:

Take, for instance, the case of a request stated as “Please pass the salt.” This is still a request [...], which is why this is a social act – but [...] he has not thereby generated any kind of reason for the adult to pass the salt. Not all social acts are reason-generating in the sense that matters to Darwall.⁵³

⁴⁹ Darwall 2006, p.21 (see also p.246) (emphasis added), referencing Fichte 1791/2000, as cited in the introduction to this chapter.

See also Darwall’s (2009, p.275) claim that “any second-personal address whatsoever” is sufficient as a starting point for Fichte’s Point – yet note that it occurs under the heading “*Addressing second-personal reasons*”.

⁵⁰ Darwall 2006. Such a description occurs as early as p.4.

See also Darwall’s (2010c, p.44) more recent claim that “Part of what it is to respect someone’s standing to make her own moral choices, even as one makes moral demands of her, is to leave it up to her how and whether to comply.”

⁵¹ Darwall 2006, p.27, as cited above.

⁵² I must note that Wood (2006, p.75), like Darwall, also takes Fichte’s Argument to establish that *reasons* are presupposed by free agency (although Wood, unlike Darwall, gives no indication as to what kind of reasons these might be – second-personal, world-regarding, etc.). Surprisingly, Wood offers this interpretation even in the light of urging the looser interpretation of Fichte’s *Aufforderung* (Wood 2006, p.73, as cited above). Here Wood differs from Ware (2009), whose interpretation I follow more closely.

⁵³ Yaffe 2010, p.251.

The summons that, according to Fichte, is presupposed by free agency is more like Yaffe's asking that you pass the salt than it is a *moral reason*.

So, again,⁵⁴ Darwall's argument for the existence of second-personal reasons fails. Fichte's Argument establishes the existence of the Fichtean summons, which we are free to accept or reject at will, but not of Darwallian second-personal, moral reasons, which we are not free to accept or reject at will. Since, after several attempts, we are still left without any good reason to accept that there are indeed such things as second-personal reasons, Darwall's argument for Morality as Equal Accountability doesn't even get off the ground.⁵⁵

2.2. Morality as a way of relating to people

Darwall's argument for Morality as Equal Accountability depends on the existence of second-personal reasons.⁵⁶ We have now examined several possible arguments for the claimed that second-personal reasons exist, but unfortunately none has been successful.⁵⁷ As such, Darwall's argument for Morality as Equal Accountability does not go through, and so we might justifiably conclude that the project of *The Second-Person Standpoint* fails.

However, I wish to suggest a different reading of *The Second-Person Standpoint*. I do not think it is the natural or most obvious reading thereof, but it is the next best, charitable move, if I am correct that the arguments previously considered fail to support Morality as Equal Accountability. This will be the Ryle-inspired reading of Darwall, the groundwork for which was laid in the preceding chapter. That is, I suggest that we read Darwall instead as offering an argument for moral action as a way of relating to people.⁵⁸

Let us return to Fichte's Argument.⁵⁹ According to this argument – and, likewise, according to the argument I mounted in the preceding chapter⁶⁰ – the world, or Real reasons, cannot give us our freedom, because the world “causally necessitate[s] what we do, restricting our freedom to do otherwise [...]”.⁶¹ Rather, argues Fichte, the agent is free in choosing whether or not to accept a summons (where a summons is correctly understood as more like an invitation and less like a demand⁶²) – and, importantly, this is a summons issued to the agent by an *Other*.

⁵⁴ See Ch.2, s.2, and Ch.3-5, as recapped in the introduction to this chapter.

⁵⁵ See Ch.2, s.1.

⁵⁶ See Ch.2.

⁵⁷ See Ch.2, s.2, Ch.3-4, and ss.1 and 2.1, above.

⁵⁸ See Ch.6, s.5.

⁵⁹ See s.1, above.

⁶⁰ See Ch.6, s.4.2.

⁶¹ Wood 2006, p.74, as cited in s.1 above.

⁶² See s.2.1, above.

This is Fichte's basic point of departure from Kant.⁶³ Whereas the Kantian agent constitutes him-/herself, independently and as a solo endeavour, the Fichtean agent's constitution is essentially a reaction to *Others*. That is, Fichte argues, *contra* Realism and *contra* Kant, that a person's knowledge of his/her own free agency is preceded by that of Others' – as, interestingly, we saw in the preceding chapter that Ryle argues, *contra* Descartes, that a person's knowledge of his/her own mental states is preceded by his/her knowledge of *others'* mental states.⁶⁴

According to Fichte, "consciousness begins with consciousness of a summons",⁶⁵ and, as Wood helpfully interprets it: "my own self-consciousness begins with my consciousness of another's consciousness as addressing me ("summoning" me)."⁶⁶ This leads Fichte to conclude, *contra* Kant, that the *I* is necessarily *intersubjective*.⁶⁷ Darwall does appear to come close to making the same claim when he writes: "the very concept of [a] person is itself a second-personal concept."⁶⁸ However, it is clear that what is involved in Fichtean (inter-) subjectivity is just reciprocal *consciousness*: my consciousness of you as consciously summoning me (where I am free to reject this summons). What Fichtean free agency presupposes is not, as Darwall has theorized, that the second-person Other gives the first-person *I* a *reason*, but just that the first person *recognizes the second as recognizing the first*.⁶⁹

I suggest, then, that moral action, category-unmistakenly understood, is not action caused by moral reasons (whatever their nature), but is, instead, a way of relating to people – where that way of relating to people is one in which the first person recognizes the second as recognizing the first.

Thus, although I do not think that Fichte's Argument is successful in establishing the existence of the second-personal reason,⁷⁰ I do think it is successful in establishing *something*: that "my own self-consciousness begins with my consciousness of another's consciousness as addressing me ("summoning" me)."⁷¹ If Fichte is correct that the very nature of the *I* is intersubjective, then it will follow that there are no agents who do not recognize Other agents as reciprocally recognizing them. And, if I am correct that moral action is a way of relating to people, where that way is one in which agents reciprocally recognize each other, then it will follow that morality is indeed *intrinsic to*

⁶³ See the introduction to this chapter.

⁶⁴ Ryle 1949/2009, pp.173-177, as cited in my Ch.6, s.5.

⁶⁵ Fichte 1962, pt.4, vol.2, p.189, as cited by Wood 2006, p.72.

⁶⁶ Wood 2006, p.72.

⁶⁷ Wood 2006.

⁶⁸ Darwall 2006, p.80 (emphasis removed).

⁶⁹ See Ch.4, ss.3.3 and 3.4.

⁷⁰ See s.2.1, above.

⁷¹ Wood 2006, p.72, as cited above.

persons.⁷² That is, if we read Darwall quite differently, as employing Fichte's Argument not to demonstrate that a special kind of reason – the second-personal reason – exists, but just as demonstrating that moral action is a way of relating to people, when those people reciprocally recognize each other, then we reach the positive outcome that morality is intrinsic to persons: moral action is not action caused by a moral reason, but is a way of relating to people; that particular way of relating to people is one in which parties reciprocally recognize each other; and, if Fichte is correct, reciprocal recognition is something any agent must partake in if she/she is to be an agent at all. That is, part of what it is to be an agent at all is to stand in moral relationships of reciprocal recognition to other agents. Note, though, that we reach this conclusion that morality is intrinsic to persons in an entirely different manner to that in which we (purportedly) reached it in this project's early consideration of Darwall's Second-Personal account.

That is, previously, the reasoning was that Darwall's Second-Personal account (purportedly) achieved an account of morality as intrinsic to persons *because of the nature of moral reasons as bipolar*. The Second-Personal account was taken to be so promising because it posited moral reasons not as world-regarding, but as person-regarding – and, importantly, not as first-personal but *second-personal*: as a function of the relationship between two persons.⁷³ To put it crudely, the idea was that moral reasons, on this Second-Personal account, will be intrinsic to persons, because moral reasons are the kinds of things created – or “given”⁷⁴ – when two or more people interact.

However, if my argument of the preceding chapter is correct, then this Second-Personal account, originally considered, is category-mistaken.⁷⁵ Moral reasons, I claimed in the previous chapter, are not *things*; they are not states with causal properties, and they certainly aren't states that cause moral action. Likewise, they are not the “function of the relationship between two persons” – because they aren't the *function* of anything. They aren't “given” by a second person to a first, because moral reasons they aren't of the category of things they may be given in the first place. I can imagine Ryle writing, in his characteristic style: “You can give a person a gift. You can even give a person a cold. But you cannot give a person a *reason*.” Instead, on the Ryle-inspired reading, moral action is a way of behaving, as performing with team-spirit is a way of performing:⁷⁶ it is a way of performing an action – in particular, one in which two agents reciprocally recognize one another – rather than the causal effect of any special cause. This holds even when that cause is posited to be – as Darwall apparently posits it to be – the second-personal reason with its special, bipolar nature.

⁷² See Ch.1, s.4.

⁷³ See Ch.1, s.6.2.2.

⁷⁴ See, for instance, Darwall 2006, p.4, as cited in Ch.1, s.1.2.

⁷⁵ See Ch.6.

⁷⁶ See Ch.6, s.5.

So, we might, as I try to motivate, read *The Second-Person Standpoint* in a very different way. Taking inspiration from Ryle, we might read Darwall not as positing second-personal reasons as reasons of a special nature, but instead as investigating moral action as a way of behaving – and finding, courtesy of Fichte, that it is a way of behaving premised upon reciprocal recognition. If this is the correct way to read *The Second-Person Standpoint*, then I agree with Lavin that the book represents “a significant philosophical achievement”⁷⁷ in its managing to shirk off the category-mistaken approach of traditional metaethics, such as Realism, and approach moral action instead in an entirely new, category-unmistaken way.

However, as I have noted, it is a controversial matter whether or not this is indeed the reading of Darwall.⁷⁸ Below, I explore how we might have read *The Second-Person Standpoint* quite differently, and in so doing examine how and why not all portions of Darwall’s text necessarily support this Ryle-inspired reading.

3. Darwall’s Second-Personal account

I turn now to Darwall’s Second-Personal account, where I hope to provide an interpretation according to which it presents a radically different, Ryle-inspired approach to that of a traditional metaethic such as Realism. In so doing, I will also explain how I think the Second-Personal account as it stands suffers some residual category-mistaken aspects. I begin where *The Second-Personal Standpoint* itself begins: with Darwall’s two possible ways of responding to the foot-treader.

3.1. The Second-Personal account as an alternative to Realism

I believe that the one of the primary mistakes of *The Second-Personal Standpoint* occurs in its opening pages already, in Darwall’s foot-treader scenario. There the reader is presented with two possible ways of responding to the foot-treader: one in which you, the victim of foot-treading, draw the foot-treader’s attention to a Real, pre-existing reason to reason his foot from on top of yours, and the other in which you demand that the foot-treader remove his foot from on top of yours, thereby giving him a different kind of reason – a second-personal reason – to do so.⁷⁹

The effect of pitting Realism and the Second-Personal account against each other in this way, right at the outset, is that the Second-Personal account is thereby presented as a traditional metaethic, like Realism – only proffering a different account of the nature of the moral reason: bipolar, rather than Real. As Darwall explicitly describes it, a second-personal reason is “a distinctive kind of (normative)

⁷⁷ Lavin 2008.

⁷⁸ See the introduction to this chapter.

⁷⁹ Darwall 2006, pp.5-7, as cited in my Ch.1, s.1.2.

reason for acting”.⁸⁰ However, on the Ryle-inspired reading of *The Second-Person Standpoint* that I am now concerned to defend, attempting to explain moral action by inferring to the special nature of its causative moral reason – be that nature Real, or bipolar – is a category-mistake. Rather, if *The Second-Person Standpoint* is the sort of “significant philosophical achievement”⁸¹ I gestured at, above,⁸² then it is not an *alternative* to Realism at all, in the sense that an “alternative” is a different answer but to the same (“improper”⁸³) question.

Jennifer Hornsby⁸⁴ has offered a compelling criticism of Daniel Dennett’s recent philosophy, which is analogous to my present criticism of Darwall. I hope that the reader will indulge a short digression into Hornsby, therefore, as an aid to my present point.

3.1.1. An analogy with Hornsby

To describe Hornsby’s argument, as I have done, as a “criticism” of Dennett is, perhaps, a too simplistic, flat description. This is because Hornsby argues for the complex position that Dennett gets a crucial point “exactly right”⁸⁵ in his earlier work, but that subsequent work of his goes on to destroy that fundamentally right point. This is not unlike my present position, which is that *The Second-Person Standpoint* almost gets something very important right, but that other portions of its text undermine it – where one key case in point is the initial set-up of the Second-Personal account as pitted against Realism.

Hornsby is concerned to defend a distinction Dennett draws in his 1969 book, *Content and Consciousness*. This is the distinction between the “personal” and the “sub-personal” levels of explanation. Personal-level explanations will be those that “proceed in terms of the needs, desires, intentions and beliefs of an *actor* in an environment”,⁸⁶ whereas sub-personal-level explanations will be those that proceed in terms of brain functions or other physical processes, for instance. And, according to the 1969 distinction that Hornsby wishes to defend, personal-level explanations are entirely autonomous of and not reducible to sub-personal-level explanations. Personal-level explanations are of a wholly different kind to those available at the sub-personal level of explanation. This is similar to my claim that, on a Ryle-inspired reading, *The Second-Person Standpoint* presents an entirely new approach to Metaethics, wholly different in kind to its

⁸⁰ Darwall 2006, p.4 (emphasis removed), as cited in my Ch.1, s.1.2.

⁸¹ Lavin 2008, as cited in s.2.2 above.

⁸² See s.2.2, above.

⁸³ Ryle 1949/2009, p.12, as cited in the introduction to my Ch.6. See also Prichard 1912/1949, p.2, as also cited in the introduction to my Ch.6.

⁸⁴ Hornsby 2000.

⁸⁵ Hornsby 2000, p.6, fn.2.

⁸⁶ Dennett 1969, p.164, as cited by Hornsby 2000, p.8.

predecessors, such as Realism, which attempt to explain moral action in terms of its special cause: the moral reason.

In subsequent work, Dennett introduces his famous “stances”: the physical stance, the design stance, and the intentional stance.⁸⁷ On the face of it, the stances will preserve the earlier personal/sub-personal distinction: the intentional stance corresponds to the personal level, and the physical and design stances correspond to the sub-personal level.⁸⁸ However, all the difference is made by Dennett’s new claim of the stances: that they are, to use Hornsby’s phrase, “continuous with” each other.⁸⁹ That is, Dennett claims that it is not the case that any one stance is appropriate to some kinds of systems but not others; instead, he claims that any stance may be taken up towards any kind of system – just as long as it works. Indeed, Dennett even allows that we may just as well take up the intentional stance to a thermostat as we do to other persons.⁹⁰

However, Hornsby takes serious issue with this later work of Dennett’s and how, she claims, it obliterates his earlier – correct – personal/sub-personal distinction. Even though it would appear that the personal level is preserved in the intentional stance, and the sub-personal level is preserved in the physical and design stances, Hornsby argues that this is not the case. This is because the stances are, as she puts it, “in competition with” each other.⁹¹ Now, on the face of it, it would appear that the claim that Dennett’s stances are “in competition with” each other is in direct contradiction of Hornsby’s earlier description of them as “continuous with” each other.⁹² However, the idea is that the stances are “continuous with” each other insofar as each stance could be used to explain the behaviour of the very same system – but, the stances are also “in competition with” each other in the sense that each stance vies to achieve the very same goal: explaining the behaviour of a particular system. To use a crude analogy, two sportspeople will only be “in competition with” each other if they play the same sport: if one plays tennis and the other golf – that is, if the same sport is not “continuous” between them – then they will never be “in competition with” each other. And, this is precisely the problem, according to Hornsby. If it is the case that Dennett’s personal level of explanation has been preserved in the intentional stance and that the sub-personal level of explanation has been preserved in the design and physical stances, then it will follow that the

⁸⁷ See, for instance, Dennett 1981.

Hornsby (2000, p.6, fn.2) argues, however, that Dennett “changed his mind” about the personal and sub-personal levels of explanation as early on as the 1970s.

⁸⁸ Hornsby 2000, p.9.

⁸⁹ Hornsby 2000, p.21, fn.27.

⁹⁰ Dennett 1981, p.65ff.

⁹¹ Hornsby 2000, p.20.

⁹² Hornsby 2000, p.21, fn.27, as cited above.

personal level of explanation and the sub-personal level of explanation are “continuous with” each other, vying to achieve the very same goal: explaining the behaviour of the self-same system.

However, that the personal and sub-personal levels of explanation are “continuous with” each other is precisely antithetical to Dennett’s original distinction. According to this original distinction, which Hornsby would want to preserve, personal-level explanation is a kind of explanation wholly different from, and entirely autonomous of and not reducible to, sub-personal level explanation. The trouble is that, in the move from his 1969 thesis to his later thesis, Dennett proffered the intentional stance as an *alternative to* the physical stance, as a way of achieving the same goal. When this is the case, the sense is lost in which personal-level explanations do something entirely different to sub-personal-level explanations; when they’re presented as being “in competition with”⁹³ each other, then one is just an alternative for doing exactly the same thing as the other. Hornsby, however, wishes to retain the sense in which personal-level explanations do something entirely different to sub-personal-level explanations.⁹⁴

Likewise, I argue, when Darwall sets his Second-Personal account up as an *alternative to* Realism in the very first pages of *The Second-Person Standpoint*, the stage is set for a metaethic that attempts to do exactly what Realism attempts to do (only, hopefully, better).

3.1.2. A Ryle-inspired reading of the foot-treader scenario

I have argued that it is a mistake to present the Second-Personal account as an alternative to Realism, because it encourages the category-mistaken understanding that the Second-Personal account differs from Realism in that it offers a different account of the nature of moral reasons: as bipolar, rather than as Real.⁹⁵ How else might we have read this initial set-up, in the foot-treader scenario?

If, as I have suggested, we read *The Second-Person Standpoint* as an exploration of the Ryle-inspired claim that moral action is a way of relating to people – in particular, a way of relating to people involving reciprocal recognition⁹⁶ – then we might read the foot-treader scenario quite differently. On this approach, the Realist response to the foot-treader will not count as a *moral* response at all. This is because merely drawing the foot-treader’s attention to a state of the world is not an action premised upon two persons’ reciprocal recognition of each other. On the Realist account, your role in the exchange with the foot-treader is merely – as I put it in a previous chapter – that of a *vessel*,

⁹³ Hornsby 2000, p.20, as cited above.

⁹⁴ It is interesting to note that much of Hornsby’s argument in defence of Dennett’s earlier (1969) distinction is motivated by none other than Ryle (1949). See Hornsby 2000, pp.7-14 and 23.

⁹⁵ See s.3.1.1, above.

⁹⁶ See s.2.2, above.

and, as such, any one of a number of suitable ‘vessels’ will do.⁹⁷ In the envisaged scenario you draw the foot-treader’s attention to the fact of your pain – but an observant third person might just as well have done the same, as might the foot-treading correlate of a seizure-response dog, or even a pressure-sensitive alarm system wired into one’s shoe. It is not merely that your role *qua* victim of foot-treading is incidental to the exchange – but that your role *qua agent* is incidental to the exchange: even an inanimate pressure-sensitive alarm system could have drawn the foot-treader’s attention to the fact of your pain. The way of relating to people captured by the Realist response to the foot-treader, then, is one in which you are merely incidental and irrelevant; it is not a way of relating to people premised upon reciprocal recognition – indeed it’s not even necessarily a way of relating *to people* at all.

On the other hand, if you take the second-personal response to the foot-treader, you relate to the foot-treader in a very different way. When you tell him to get off your foot, you *summons* the foot-treader to be one possible way, to perform a possible action: removing his foot from on top of yours. In so doing, you bring the foot-treader to consciousness of his own free agency – and in so doing, you get the foot-treader to see *you* as an agent. This is because you effectively say to the foot-treader: “Your treading foot is a bit of the not-*I* that I strive against in trying to relieve myself of pain” – and if *you* strive against the not-*I*, then you are, like the foot-treader himself, an *I*, an agent.⁹⁸ So, contrary to the Realist approach on which one party is an agent and the other only accidentally so, the hallmark of this Second-Personal approach is that *both* parties must be agents. This is a different way of relating to people: one in which you, the second person, are essential to the relation. It involves two people – not merely incidentally, as in the Realist approach, above – but essentially: where the second person (you) brings the first person (the foot-treader) to consciousness of the second person’s consciousness of the first (your consciousness of the foot-treader’s standing on your foot).

Note that the Realist and Second-Personal responses to the foot-treader have, on the Ryle-inspired reading, above, been distinguished without any reference to *reasons*. The difference between them, on this reading, is not that they posit different kinds of reasons in the attempt to best account for the same moral behaviour; rather they posit different ways of behaving: one according to which the way of behaving is premised upon the reciprocal recognition of and by two persons, and the other not. My point here, however, is not at all to eliminate reasons-talk, but, rather, to make clear what we are

⁹⁷ See Ch.5, s.1.3.

⁹⁸ See s.1, above.

actually talking about when we *do* talk about ‘reasons’.⁹⁹ And, a person’s acting for reasons, on this Ryle-inspired reading, is not a matter of being caused to act by a special kind of thing called a ‘reason’ – not even when this special thing has a bipolar structure; it is just a particular way of relating to people: a way of relating to people to which reciprocal recognition is essential.

We are misled into thinking otherwise, however, when the Second-Personal approach is pitted against the category-mistaken Realist approach, as Darwall sets it up in the opening pages of *The Second-Person Standpoint*.¹⁰⁰ Of course, I have only exacerbated this problem by contrasting the Second-Personal approach at the beginning of this project not only with Realism, but also with Hobbesian Egoism, Neo-Kantianism and Divine Command theory.¹⁰¹ My intention in that chapter was of course not to *misrepresent* Darwall’s account, but rather to present it in a positive light, as improving upon all those shortcomings of the rest of the metaethics discussed. However, since subsequent chapters have revealed that Darwall’s account in fact is *not* entirely successful in this regard,¹⁰² the next best charitable move is to read Darwall, not as offering an alternative to all those traditional metaethics, but instead as making a radically different – category-unmistaken – claim, as is my aim in this present chapter. Below, I provide the same Ryle-inspired analysis of Pufendorf’s Point.

3.2. Pufendorf’s Point

On Darwall’s explicit presentation thereof, Pufendorf’s Point is that an agent is held genuinely morally accountable when she holds him-/herself accountable.¹⁰³ I went on to question, however, what it means to be “held accountable”, such that a second person may hold a first accountable or a first may hold him-/herself accountable. And, although he does not present this account explicitly, I was able to piece together on the basis of the rest of Darwall’s text the following account of accountability: genuine moral accountability is the agent’s conviction that what he/she did was wrong, such that he/she may justly be (proportionally) punished for it.¹⁰⁴ Pufendorf’s Point, therefore, is that a genuinely morally accountable agent is one in whom occurs the internal state of conviction, and an agent who is not held genuinely morally accountable but whose behaviour we merely attempt to modify by negative reinforcement is one in whom this state of conviction does not

⁹⁹ Likewise, Ryle (1949) does not endeavour to eliminate mental-state-talk. His project, rather, is to make clear exactly what we are talking about when we *do* talk about mental-states – and we’re not talking about hidden ‘things’ inside a person’s head; we’re just talking about ways of behaving.

¹⁰⁰ See s.3.1.1, above.

¹⁰¹ See Ch.1.

¹⁰² See Ch.2-Ch.5.

¹⁰³ See Ch.1, 6.1.2.

¹⁰⁴ See Ch.5, s.1.2.1.

occur.¹⁰⁵ In Darwall's own words: "To see what Pufendorf is getting at, we need to distinguish between merely *external* censure or blame [...] and *internal* blame, that is, the self-reactive attitude of guilt."¹⁰⁶

However, I hope by now that it is clear how Ryle would respond to the above suggestion: we commit a category-mistake when we posit *internal states* to explain an observed distinction. Just as the Cartesian Dualist commits a category-mistake in positing such things as private, inner mental states in order to explain the distinction between intelligent and unintelligent behaviour,¹⁰⁷ so Darwall commits¹⁰⁸ a category-mistake in positing such a state as *conviction* in explaining Pufendorf's distinction between genuine moral accountability and behavioural modification by negative reinforcement. Rather, on a Ryle-inspired reading, the distinction that Pufendorf draws between genuine moral accountability and behavioural modification by negative reinforcement is not one of a special internal state – conviction – that is present in cases of the former but not of the latter; it is between different *ways of relating to people*.

The Ryle-inspired reading of Pufendorf's Point is fairly straightforward, in light of the preceding section.¹⁰⁹ There, I argued that when you tell the foot-treader to get off your foot, you *summons* the foot-treader to be one possible way – that is, not having his foot planted on top of yours – and, in so doing, you bring the foot-treader to consciousness of his own free agency, while, at the same time, getting the foot-treader to recognize your own agency, as one who strives against the not-*I* of a trod-upon foot. The Ryle-inspired account of what it is to hold an agent accountable is, similarly, that in holding the foot-treader accountable by telling him that what he did was wrong, you present him with other possible ways of acting: he should not have trod upon your foot. And, in so doing, you bring into focus the foot-treader's free agency. Furthermore, in holding the foot-treader accountable in this way, you assert your own agency in telling him that he thwarted your strivings against the world by treading on the foot you were striving to keep pain-free – that you, like him, are an *I* striving against the not-*I*. This is the Ryle-inspired account of Pufendorfian genuine moral accountability. It does not require us to posit the internal state of conviction in the agent who is held genuinely morally accountable.

On the other hand, there is Pufendorf's behavioural modification by negative reinforcement. And, this, it is clear, is a different way of relating to people – if it is a way of relating *to people* at all. Unlike

¹⁰⁵ See Ch.5, s.1.2.1.

¹⁰⁶ Darwall 2006, p.112, as cited in my Ch.5, s.1.2.1 (emphases added).

¹⁰⁷ See Ch.6, s.4.1.

¹⁰⁸ Or: *I commit* a category-mistake, if one holds that my reconstruction of an account of accountability on the basis of Darwall's text (see my Ch.5, s.1.2.1) has gone awry.

¹⁰⁹ See s.3.1.2, above.

the case of genuine moral accountability (above), it is incidental to a case of behavioural modification by negative reinforcement that an *agent* threatens or imposes the feared punishment, and so there is no reciprocal recognition present. The negative reinforcement may be implemented by various machines even, as in “aversion therapy” by electric shock. Above, I suggested that the foot-treader’s attention might be drawn to the fact of his causing you pain by any of a numbers of means, such as a pressure-sensitive alarm wired into your shoe.¹¹⁰ Imagine, further, that this system is so sophisticated that if, after sounding the alarm, it monitors no decrease in pressure (after maybe two seconds or so), it subsequently emits an electric shock (from which your foot is somehow protected, but the foot-treader’s is not). This would act as an aversive stimulus, and would negatively reinforce non-foot-treading behaviour in the foot-treader. So, your involvement as an *agent* who negatively reinforces foot-treading is irrelevant to this set-up; if you’re involved at all, you fulfil only the same role that a sophisticated alarm system might have fulfilled.

Moreover, not only is it incidental to cases of behavioural modification by negative reinforcement that an agent threatens or imposes the feared punishment – it is also incidental that the *subject* of the threatened or imposed punishment is an agent. For instance, behavioural modification by negative reinforcement is often the approach taken to animals, such as dogs, which are not agents. Thus, *no* person is actually essential to this way of relating to way of behaving that is behavioural modification, and so it is clearly distinct from the way of relating to people by holding them genuinely morally accountable, to which reciprocal recognition is essential.

In summary, Darwall apparently understands Pufendorf’s Point as the presence or absence of a certain internal state – what I have called *conviction*¹¹¹ – writing, “To see what Pufendorf is getting at, we need to distinguish between merely *external* censure or blame [...] and *internal* blame, that is, the self-reactive attitude of guilt.”¹¹² However, we could account for Pufendorf’s distinction between genuine moral accountability and behavioural modification by negative reinforcement, on a Ryle-inspired approach, just as different ways of relating to people, rather than in terms of the presence or absence of an internal state such as conviction. Such a Ryle-inspired reading should avoid the category-mistake of positing internal states in order to explain observed differences in behaviour.

¹¹⁰ See s.3.1.2, above.

¹¹¹ See Ch.5, s.1.2.1.

¹¹² Darwall 2006, p.112 (emphases added), as cited above.

Since Pufendorf's Point plays such a large role in the main argument of *The Second-Person Standpoint*,¹¹³ this Ryle-inspired reading thereof will have several repercussions. I consider these, below.

3.2.1. Realist accountability

This Ryle-inspired reading of Pufendorf's Point¹¹⁴ will have implications for the Realist account I advocated in Chapter 5. There, I defended Realism against the objection that it misconstrues the nature of moral accountability, where the particular misconstrual under examination was Realism's treating moral accountability as behavioural modification by negative reinforcement.¹¹⁵

In response to the charge that Realism treats all cases of punishment as cases of behavioural modification, rather than genuine moral accountability, I argued that this is simply false and that no (non-Utilitarian) Realism is committed to construing moral accountability as behavioural modification by negative reinforcement. However, in light of this Ryle-inspired reading of Pufendorf's Point, I will again consider the objection that Realism misconstrues the nature of moral accountability – not by construing it as behavioural modification by negative reinforcement, but by failing to construe it as a way of relating to people to which reciprocal recognition is essentially.

In Chapter 5, I argued that Realism, unlike Darwall's Second-Personal account, is able to accommodate those broader aspects of Mill's Point – namely, that non-victims, such as the agent him-/herself or members of the general moral community, also have the authority to hold agents accountable. I argued that this was so, because, on the Realist account, any individual – victim, agent or third-person observer – may in principle possess the epistemic authority to hold an agent accountable when that individual has knowledge of the facts of wrongdoing (and the means of communicating it to the agent).¹¹⁶ And, as we noted, there are various ways in which such information might be communicated: the individual's sitting down and explaining to the agent why what he/she did was wrong, renting out a billboard along the agent's daily commute proclaiming the message "What you did was wrong!", putting the agent in prison, or even – as in the Milgram

¹¹³ In particular, it is by Pufendorf's Point that Darwall deduces Morality as Equal Accountability: since, Darwall argues, what it is for an agent to be held genuinely morally accountable is for that agent to hold him-/herself accountable, it follows that the first and second persons will share equal authority (see my, Ch.1, s.6).

I consider the repercussions of a Ryle-inspired reading of Pufendorf's Point for the argument for Morality as Equal Accountability specifically, in s.3.2.2, below.

¹¹⁴ See s.3.2.1, above.

¹¹⁵ See Ch.5, ss.2.1 and 2.1.1.

¹¹⁶ See Ch.5, s.1.2.2.

experiments¹¹⁷ – sitting in a room next door and pressing a button to pass an electric shock through the agent’s body.¹¹⁸

It follows, therefore, that when, on the Realist proposal, an individual holds an agent accountable, that particular individual is merely incidental to the occurrence: the agent him-/herself, or a member of the moral community, could have performed the task just as well (if that individual possessed the relevant knowledge of the facts of wrongdoing and the means of communicating it¹¹⁹). So, an individual who holds an agent accountable, on this Realist proposal, is in the same position as you are when you draw the foot-treader’s attention to your trod-upon foot, on Darwall’s Realist account:¹²⁰ you, as the second person, are incidental and irrelevant to the exchange. This is not a way of relating to people premised on reciprocal recognition.

On the Realist account of accountability, the following scenario would count as a case of holding accountable: one day, on the evening news (in a very, very boring city), Barry announces that Sally has suffered a broken promise – and I realize with a shock that *I* was the breaker of the promise to Sally. Barry’s news piece produces in me the state of conviction, and thus – according to our previous, category-mistaken analysis¹²¹ – I am held genuinely morally accountable by Barry. Such a case involves two agents: the individual – Barry – and the guilty agent – myself – but these two agents have no reciprocal recognition of each other in this set-up. When asked in an interview what the Second-Personal standpoint is, Darwall replies: “it is the standpoint you and I are in right now, though the people who are “overhearing” our conversation may not be.”¹²² That is, Darwall and his interviewer have reciprocal recognition of each other, but we readers of that interview’s transcript do not. We merely “overhear” the discussion of two second persons; although we are conscious of their words, they are not reciprocally conscious of our consciousness of them. Likewise, Barry the newsreader reports a fact that happens to be a fact about me, but Barry does not in so doing *address* me; he is not conscious of me, so obviously I am not reciprocally conscious of his consciousness of me. So, although Barry’s news piece produces in me the internal state of conviction, it is not a way of relating to people premised upon reciprocal recognition.¹²³

In so far as Realism allows for a case like this as a case of moral accountability, Realism goes wrong in construing moral accountability as a way of relating to people that does not require reciprocal

¹¹⁷ See Ch.4, s.3.4.

¹¹⁸ See Ch.5, s.1.2.1.

¹¹⁹ See Ch.5, s.1.2.2.

¹²⁰ See s.3.1.2, above.

¹²¹ See Ch.5, s.1.2.1.

¹²² Marshall 2013, citing Darwall.

¹²³ For a similar scenario, see my discussion of the Milgram experiments in Ch.4, s.3.4.

recognition, *contra* the category-unmistaken reading of Pufendorf’s Point.¹²⁴ (Of course, if I am correct that Realism is guilty of a category-mistake,¹²⁵ then we should hardly be surprised to find that it goes wrong in this way also.)

3.2.2. Morality as Equal Accountability

In Chapter 5, I considered the objection that Realism misconstrues the nature of moral accountability – but went on to counter this objection, arguing that Realism can indeed accommodate reasons “of the right kind” for holding agents accountable.¹²⁶ The Ryle-inspired reading of Pufendorf’s Point, however, apparently has the advantage of affirming (a version of) this objection to Realism: the Realist proposal *does* misconstrue the nature of genuine moral accountability, because it fails to identify reciprocal recognition as essential to this way of behaving.¹²⁷ However, this Ryle-inspired reading of Pufendorf’s Point may also bring with it a *disadvantage*. In particular, it will no longer support Darwall’s argument for Morality as Equal Accountability.

Darwall’s argument for Morality as Equal Accountability, recall, begins with the Pufendorfian premise that an agent is held genuinely morally accountable when he/she holds him-/herself accountable. (Of course, this is precisely the interpretation of Pufendorf’s Point denied by the preceding Ryle-inspired reading thereof.¹²⁸) On this basis, Darwall concludes that the first-person agent and the second person holding the agent accountable must therefore share the same, equal authority. That is, Morality as Equal Accountability follows on the basis of Pufendorf’s Point, according to Darwall’s major argument of *The Second-Person Standpoint*.¹²⁹

But, on the Ryle-inspired reading of Pufendorf’s Point, above,¹³⁰ genuine moral accountability is distinguished from mere behavioural modification in terms of their being different ways of relating to people (if behavioural modification by negative reinforcement is a way of relating to people at all), and not in terms of the agent’s holding him-/herself accountable – that is, not in terms of the

¹²⁴ Note, however, that this is still not the strong claim that Realism is *committed* to the necessary misconstrual of moral accountability (see Ch.5, s.2.1, as referenced at the beginning of this present s.3.2.1). Rather, it is the claim that Realism *allows for* cases of moral accountability as ways of relating to people that do not require reciprocal recognition, and it therefore misconstrues the nature of moral accountability insofar as it does so. Even on the Realist account, cases of an individual holding an agent accountable will not all be cases in which a newsreader announces to everyone and no-one that an agent – whomever that agent may be – is guilty of a moral wrongdoing, but will be the more ordinary cases in which an individual *tells* an agent that what he/she did was wrong – and such tellings will, I think, constitute a way of relating to people that is premised upon reciprocal recognition, as will all cases of genuine address (see Ch.5, ss.2.2.2.2.3 and 2.2.2.3).

¹²⁵ See Ch.6, s.4.2.

¹²⁶ See Ch.5, ss.2.1 and 2.1.1.

¹²⁷ See s.3.2.1, above.

¹²⁸ See s.3.2, above.

¹²⁹ See Ch.1, s.6.1.2.

¹³⁰ See s.3.2, above.

presence or absence in the agent of the internal state of conviction. On the Ryle-inspired reading of Pufendorf's Point, genuine moral accountability requires us to posit only the first-person agent's consciousness of the second person's consciousness of the first, in order to distinguish it from mere behavioural modification, which does not necessarily involve reciprocal recognition. That is, the Ryle-inspired reading of Pufendorf's Point does not require us to posit the first-person agent's authority to hold him-/herself accountable. And, without the first person's authority to hold him-/herself accountable, it does not follow that the first person shares the same authority as the second – and, in the absence of this conclusion, Morality as Equal Accountability does not follow.

This is, however, not a *new* criticism. I have already contended that Darwall's argument for Morality as Equal Accountability does not go through: the fact that the first-person agent has the authority to hold him-/herself accountable as the second person holds the first person accountable does not entail that the first person shares equal, reciprocal authority *over the second*; all that is entailed is that the two share the same authority over the *first* person only. The argument for Morality as Equal Accountability is, I claimed, a *non sequitur*.¹³¹

And, as we saw, before,¹³² the failure of the argument for Morality as Equal Accountability reopens the possibility of the amoralist. The point of Darwall's argument for Morality as Equal Accountability, recall, is to establish that all persons are equally accountable – that is, that all persons are subject to the same moral requirements, and morality is *universal*.¹³³ So, unless we can demonstrate by way of Morality as Equal Accountability that morality is universal, we will not have adequately responded to the amoralist: even if we are able to establish that *some* persons are subject to moral reasons, the amoralist may just retort that he is one of those lucky few who are not subject to moral reasons.¹³⁴

So, if the argument for Morality as Equal Accountability does not go through – even on the Ryle-inspired reading of Pufendorf's Point – then we are apparently still vulnerable to the amoralist's challenge.

3.2.3. Addressing the amoralist's challenge

However, one may be able to argue that that we actually don't *need* Morality as Equal Accountability in order to address the amoralist's worries. The Fichtean way of relating to people, to which reciprocal recognition is essential, can do the work for us.

¹³¹ See Ch.2, s.1.

¹³² See Ch.2, s.1.

¹³³ See my Ch.1, s.1.3.

¹³⁴ See Ch.1, s.2.2ff.

Above,¹³⁵ I argued that the Fichtean summons is significantly different from a Darwallian second-personal reason in that the Fichtean summons leaves the first person free to accept or reject that summons. Unlike a Darwallian second-personal reason, the Fichtean summons has no obligating power over the first person and may be freely accepted or rejected. However, I should not overstate this claim, for the Fichtean summons certainly does have *some* power over the summonsed first person. It is not, unlike a Darwallian second-personal reason, the power to *obligate*, but is, rather, the power to get the summonsed first person to see the summoning second person *as an agent*. That is, the Fichtean summons has the power to lock the first person into the way of relating to people to which reciprocal recognition is essential. My summoning you does not give you a Darwallian second-personal reason to do whatever I have told you to, but it does commit you to *responding* to me. In Darwall's words, "What makes all the difference is that a summons [...] comes with an RSVP."¹³⁶ Although a summonsed first person may reject (or accept) a summons, he/she can't go on as if the summons had not occurred. However the first person freely chooses to respond, it will constitute *a response* or "RSVP" to the second person. Korsgaard expresses a similar idea by her example of calling your name as you walk past. She writes: "You may carry on walking, but it wouldn't be as before, because now you would be ignoring me."¹³⁷ Even if you choose to reject Korsgaard's summons, your response constitutes one of ignoring her – it is still a response *to her*. The power of the Fichtean summons, therefore, is that it locks the summonsed first person into this way of relating to people; it forces you to see the Other seeing you.

And, the power of such a summons is arguably sufficient to respond to the amoralist's challenge. If I am an amoralist foot-treader and you tell me to get off your foot, I may wonder why I should be moral. But, on the Ryle-inspired approach I have advocated in this chapter, the amoralist's question apparently answers itself.¹³⁸ If your demand that I get off your foot gives me any cause to even wonder whether I should, then I *already* relate to you in a moral way. Just your telling me to get off your foot, leads me to see you as an agent striving against the not-*I*, and leads me to see myself as an agent choosing to act one way rather than another as a function of the possible ways of acting that

¹³⁵ See s.2.1, above.

¹³⁶ Darwall 2006, p.256.

¹³⁷ Korsgaard 1996, p.140. Note that Korsgaard, like Darwall, frames this in terms of *giving reasons*: in calling out your name as you walk past, I give you a reason to stop. However, as Raymond Geuss (1996, p.195) points out in response, the sense in which someone is "obligated" just by calling out their name is "weak and thin", because, as G.A. Cohen (1996, p.176) points out, in calling out my name as I walk past, you might actually have "given me a reason to speed up" rather than to stop. Geuss and Cohen are right to point out these problems – but, I argue, they are problem with explaining Korsgaard's observation in terms of *reasons*: what my calling your name does is not to give you a reason (neither Korsgaard's suggested reason nor Cohen's), but what it *does* do, if you hear it as a summons, is lock you into a way of relating to me in which reciprocal recognition is essential.

¹³⁸ See Ch.1, s.4.

you present to me. That is to say, your summoning me to get off your foot locks me into a way of relating to you that is premised upon our reciprocal recognition of each other.¹³⁹ And, that way of relating to people, to which reciprocal recognition is essential, is moral behaviour on this Ryle-inspired reading.¹⁴⁰ So the amoralist's question answers itself: Why should I be moral? You are *already* being moral. There is no other way for you to be; if there's an *I* at all it has got to be an *I* reciprocally aware of another *I*,¹⁴¹ so it has got to be a moral *I*.

However, I think this argument is too quick. Although it is correct that any consideration of any summons will lock the summonsed first person into a way of relating to the second person premised upon the first and second persons' reciprocal recognition of each other, it does not necessarily follow that such a first person relates to the second in a *moral* way. This is because, I have argued, there are other kinds of Fichtean summonses than those involving reasons, such as asking for the salt¹⁴² or calling out a passing person's name.¹⁴³ And, if this is the case, then it is clear that my relating to you in a reciprocally aware way need not necessarily entail moral action. It may just entail my recognition of your having tried to catch my attention as I walk past, which falls short of anything like a *moral* action.¹⁴⁴ Or – more pointedly – my relating to you in a reciprocally aware way may even consist in my *rejection* of the moral reason you try to 'give' me to, for instance, get off your foot¹⁴⁵ – in which can't I surely *don't* have such a moral reason. This will obviously not satisfy the amoralist.

So, we are left without a satisfactory answer to the amoralist's challenge, because not any way of relating to a person premised upon reciprocal recognition will constitute a *moral* way of relating to a person. But, let me note that, even though this response to the amoralist is ultimately unsuccessful, I do think it is a much more promising approach than others. Consider the Realist response to the amoralist, which simply asserts that the world is such that there are moral reasons – that is, tries to get the amoralist to see that there are moral reasons, when this is, after all, what the amoralist has stipulated he/she *can't* see.¹⁴⁶ Consider the Hobbes-like approach, that tries to show how the amoralist has moral reasons in virtue of the non-moral reasons he/she already has.¹⁴⁷ Or consider the Neo-Kantian approach, that tries to show that the amoralist has moral reasons in virtue of the

¹³⁹ See s.1, above.

¹⁴⁰ See Ch.6, s.5.

¹⁴¹ See s.2.2, above.

¹⁴² Yaffe 2010, p.251, as referenced in s.2.1, above.

¹⁴³ Korsgaard 1996, p.140, as referenced above.

¹⁴⁴ But see fn.137, above.

¹⁴⁵ See s.2.1, above.

¹⁴⁶ Korsgaard 1996, p.38, as referenced in my Ch.1, s.2.2.

¹⁴⁷ See Ch.1, s.3.1.

kind of being that he/she is.¹⁴⁸ What all of these non-Second-Personal approaches overlook is that morality is about “*how we should relate to one another.*”¹⁴⁹ So, looking outwards to the world or inwards to oneself will not address the amoralist’s challenge. Rather, a more promising approach is to look at one’s relationships with others, as this Ryle-inspired reading does. Even if it is ultimately unsuccessful because not any way of relating to a person premised upon reciprocal recognition will constitute a *moral* way of relating to a person, it is at least not fundamentally misdirected. Brewer echoes this sentiment:

Still, I think that by turning his attention to second-person[al] encounters, Darwall has opened up the right philosophical terrain on which to place moral concepts in such a way as to illuminate their normativity without changing the subject, and he has taken some large strides towards this crucial and elusive philosophical goal.¹⁵⁰

Brewer is, as I am, of the opinion that, despite the “large strides” Darwall takes in the right direction, his Second-Personal account ultimately fails. Brewer’s grounds for this claim are essentially the same as a critique I have raised here. As he writes:

The ubiquity of second-person address casts doubt on the idea that what makes it wrong to violate a moral obligation is that it violates a second-person[al] demand. The range of cases in which such demands are violated stretches from bad manners (e.g., refusing to answer a question or to shake a proffered hand) to genocide, and one cannot tell whether one is on the morally blameworthy portion of this continuum simply by determining whether a legitimate second-person[al] demand has been flouted. To appreciate what it is that qualifies a demand as moral, and lends it the special urgency associated with paradigm moral considerations, it seems necessary to engage in non-second-person[al] assessment of the stringency of the interests or values that are at stake.¹⁵¹

That is, Brewer notes, as I have (above), that relatively innocuous actions, such as asking for the salt¹⁵² or “refusing to answer a question”, will presuppose the first person’s awareness of the second’s awareness of the first; however, such actions surely do not fall on the “morally blameworthy portion of [the] continuum”. In response to this, Brewer’s suggestion is that, in order to determine which ways of relating to people with reciprocal awareness are of moral import and which are not, “it seems necessary to engage in non-second-person[al] assessment”.

I assessed a similar suggestion as the first possible account of validity, examined in Chapter 2.¹⁵³ This was the natural suggestion that what it is for a command to be valid is that it is legitimized by independent reasons, that is, that second-personal commands are authorized by non-second-personal reasons. However, as we noted there, Darwall’s Second-Personal account cannot

¹⁴⁸ See Ch.1, s.5.

¹⁴⁹ Korsgaard 1993, p.275 (emphasis added), as cited in Ch.6, s.5.

¹⁵⁰ Brewer 2008, p.167.

¹⁵¹ Brewer 2008, p.165.

¹⁵² Yaffe 2010, p.251, as referenced above.

¹⁵³ See Ch.2, s.2.3.1. See also Ch.5, s.1.1.

accommodate non-second-personal criteria of validity, because this would be to “break into [the Second-Personal] circle from outside it.”¹⁵⁴ Similarly, if we can only tell which ways of behaving premised upon reciprocal recognition are moral and which are not by “engag[ing] in non-second-person[al] assessment of the stringency of the interests or values that are at stake”¹⁵⁵ then we will be attempting to “break into [the Second-Personal] circle from outside it.” Brewer’s suggestion, then, is not acceptable.

But, there may be a way out for the Second-Personal approach, one that does not reduce it to the non-second-personal. If the moral way of relating to people is one premised upon reciprocal recognition of the other, what distinguishes it from non-moral reciprocal awareness? Darwall has, in fact, already suggested an answer to this question. The answer comes in the form of Mill’s Point – but, again, it will require a little help from Ryle.

3.3. Mill’s Point

Darwall’s Mill’s-Point argument is an inference to the best explanation.¹⁵⁶ It argues that our second-personal practice of holding each other accountable is best explained by the fact that the nature of moral reasons is, likewise, second-personal.¹⁵⁷ If this Mill’s-Point argument goes through, then we have good reason for accepting that there are indeed such things as second-personal reasons: we must posit them in order to explain our practice of holding each other accountable. However, I went on to argue in Chapter 5 that Realism can provide a better explanation of Mill’s Point than does Darwall’s Second-Personal account. As such, we are again left without any good reason to accept that there are indeed such things as second-personal reasons. Darwall’s Mill’s-Point inference to the bipolar nature of moral reasons fails.

However, if we heed Ryle’s advice, then the above line of argument is the paradigm of a category-mistake.¹⁵⁸ The “origins” of both Cartesian Dualism and Realism are inferences to the special cause of an observed phenomenon: in the case of Cartesian Dualism, private mental states are inferred to explain intelligent behaviour, and in the case of Realism, Real, moral reasons are inferred in order to explain moral action.¹⁵⁹ Likewise, according to my (mis)construal of Mill’s Point, the second-personal

¹⁵⁴ Darwall 2006, p.12, as cited in Ch.2, s.2.3.2.

¹⁵⁵ Brewer 2008, p.165, as cited above.

¹⁵⁶ See, for instance, Darwall 2006, p.76, fn.26, or 2010c, p.32, as referenced in my Ch.3.

¹⁵⁷ See the introduction to my Ch.3.

¹⁵⁸ This goes both for Darwall’s inference to Second-Personality (as I have presented his Mill’s-Point argument) (see my Ch.3) and my Realist argument (see Ch.5), as both share the form of arguing from observed phenomena to the nature of the things – reasons – that must explain them. The discussion to follow will focus only on Darwall’s Second-Personal inference, but the same goes for the Realist version of the argument I presented.

¹⁵⁹ See Ch.6, s.3.5.

nature of reasons is inferred in order to explain the observation that we are held accountable by others.

Part of Ryle's argument for the claim that Cartesian Dualism commits a category-mistake in positing internal, mental states in order to explain intelligent behaviour, recall, is that we already knew how to distinguish intelligent from unintelligent behaviour prior to arriving at the theory of Cartesian Dualism.¹⁶⁰ If, therefore, my prior treatment of Mill's Point is likewise category-mistaken (above), then we should respond as Ryle responds to the Cartesian Dualist: we *already know* how to distinguish genuine moral accountability from behavioural modification. To see how this is so, consider the argument we have followed in this project (prior to any alternative, Ryle-inspired readings). Darwall's argument for Morality as Equal Accountability claims that the address of any second-personal reason demonstrates that the first and second person share an equal, mutual authority;¹⁶¹ this follows on the basis of Pufendorf's Point, which is that an agent is held genuinely morally accountable when he/she holds him-/herself accountable, in the absence of which an agent's behaviour is merely modified by negative reinforcement.¹⁶² I, however, have questioned the antecedent of Darwall's argument for Morality as Equal Accountability: why should we accept that there are such things as second-personal reasons in the first place?¹⁶³ And, the suggestion presently under consideration – Mill's Point – is that there are such things as second-personal reasons because they are the best explanation of our practice of holding each other accountable; that is, they are the best explanation of the distinction between behaviour for we are liable to being held accountable and behaviour for which we are not so liable.¹⁶⁴ But, notice that this – the distinction between behaviour for we are liable to being held accountable and behaviour for which we are not so liable – is something we already knew all along; this distinct is the very *starting point* of Pufendorf's Point. Mill's Point, therefore, category-mistakenly infers the special nature of moral reasons in order to explain a distinction that we knew how to make all along. My presentation of Mill's Point,¹⁶⁵ therefore, is apparently category-mistaken in that it posit special states – the second-personal nature of the moral reason – in order to explain a distinction, where this is a distinction we already knew how to draw all along.

On a Ryle-inspired reading of Mill's Point, on the other hand, we do not posit the special, second-personal nature of moral reasons in order to explain behaviour for which we are morally responsible,

¹⁶⁰ See my Ch.6, s.5.

¹⁶¹ See my Ch.1, s.6.2.

¹⁶² See Ch.1, s.6.1.2.

¹⁶³ See Ch.2, s.2.

¹⁶⁴ See Ch.3.

¹⁶⁵ See Ch.3.

that is, moral action. Instead, on the Ryle-inspired approach, moral action is a way of relating to people premised upon reciprocal recognition – *and*, we can now add in light of Mill’s Point: for which we are liable to being held accountable. That is to say, Mill’s Point arguably enables us to refine the Ryle-inspired reading of moral action: it is not just any way of relating to people premised upon reciprocal recognition – because, as we have seen, just asking your companion to pass the salt will meet this description¹⁶⁶ – it is, instead, that way of relating to people that is premised upon reciprocal recognition *and* for which we are liable to being held accountable. This will hopefully supply the solution to Brewer’s worry, above¹⁶⁷ – that is, it will hopefully supply the criterion by which we can “tell whether one is on the morally blameworthy portion of [the] continuum” when “a legitimate second-person[al] demand has been flouted”, and it will do so without having to “engage in *non-second-person[al]* assessment of the stringency of the interests or values that are at stake”,¹⁶⁸ as I have argued would unacceptably constitute “break[ing] into [the Second-Personal] circle from outside it.”¹⁶⁹ Moral action, on this Ryle-inspired reading of Mill’s Point, is a way of relating to people premised upon reciprocal recognition – where that is the particular reciprocal recognition of the second person holding the first accountable.

This Ryle-inspired reading of Mill’s Point is subtly – but importantly – different from the category-mistaken one. On the Ryle-inspired reading, Mill’s Point is that what it is to be (second-personally¹⁷⁰) obligated is to relate to a person in a way that is premised upon reciprocal recognition of the particular variety in which the second holds the first accountable. That is, on the Ryle-inspired reading of Mill’s Point, being obligated is just a way of behaving, as team-spirit is just a way of playing.¹⁷¹ One is *not*, on this Ryle-inspired reading, obligated to perform an action *in virtue of there being a reason to perform the action*, as if a reason (whatever its nature) were an independent state that could explain obligation and our practice of holding accountable. Rather, on this Ryle-inspired reading, being obligated is a way of relating to a person – a way that is premised upon reciprocal recognition, where this is reciprocal recognition of the particular variety in which the second holds the first accountable, as Mill’s Point has revealed.

¹⁶⁶ Yaffe 2010, p.251, as referenced in s.2.1, above.

¹⁶⁷ See s.3.2.3 above.

¹⁶⁸ Brewer 2008, p.165, as cited in s.3.2.3 above (emphasis added).

¹⁶⁹ Darwall 2006, p.12, as cited in Ch.2, s.2.3.2, and as cited in s.3.2.3 above.

¹⁷⁰ I add this qualifier parenthetically because, on the Ryle-inspired reading, there will be no distinction between “second-personal” obligation and “non-second-personal” obligation, because any kind of obligation construed as the function of an independently existing reason (be this a bipolar reason, or a reason of a different nature) will be category-mistaken.

¹⁷¹ See Ch.6, s.5.

At several points in *The Second-Person Standpoint*, Darwall comes very close to presenting this Ryle-inspired reading of Mill's Point. For instance, he writes: "[W]hat is morally obligatory is what we are responsible to one another for doing",¹⁷² rather than: We are responsible for doing what we are morally obligated to do. Likewise, endorsing Strawson, Darwall writes:

What we are morally obligated to do [...] is what members of the moral community can appropriately demand that we do, including by responding with blame or other reactive attitudes if we fail to comply without adequate excuse.¹⁷³

That is to say, in passages like this, Darwall appears to be making the Ryle-inspired claim that it is not our (second-personal) reasons that commit us to behaving in ways that reciprocally recognize the second in the first's liability to being held accountable by the second – but, rather, that it is our reciprocal recognition of each other in the first person's liability to being held accountable by the second that constitutes our obligation – because being obligated, on this Ryle-inspired reading, *just is* a way of relating to another.¹⁷⁴

However, these appearances of a Ryle-inspired reading of Mill's Point in *The Second-Person Standpoint* don't last long. In the same discussion of Mill's Point as that cited immediately above, Darwall goes on to write, this time with reference to Williams:

Williams evidently assumes that it is a conceptual truth that violations of moral obligations are appropriately blamed and that blaming implies the existence of good and sufficient reasons to do what someone is blamed for not doing. The idea is not, of course, that normative reasons follow from the fact of someone's being blamed. Rather, in blaming one implies or presupposes that there are such reasons.¹⁷⁵

Here Darwall clearly falls back into the category-mistaken construal of Mill's Point: taking the practice of accountability to "[imply] or [presuppose]" "the existence of [...] reasons to do what someone is blamed for not doing". That is, instead of taking obligation to be a way of relating to people in which a first person is liable to being held accountable by a second, as on the Ryle-inspired reading, the quote above reveals that Darwall holds reasons to be independently existing states, the existence of which we can infer on the basis of our practice of holding each other accountable. Yet

¹⁷² Darwall 2006, p.35.

¹⁷³ Darwall 2006, p.92, referencing Strawson 1968.

¹⁷⁴ I suggest that this is why much of Darwall's discussion is framed in terms of 'obligation' rather than in terms of 'reasons' (see the note on this in the introduction to my Ch.3): in order to distance himself from the category-mistaken view that we, firstly, occupy a certain state – namely, the having of a reason – and then, subsequently, go on to act because of it. 'Obligation', of course, does not immunize one from the category-mistake (I don't think it has immunized Darwall), but it does appear to have some protective effect, in that 'obligation' is more clearly an *abstract* noun than 'reason' is, and thus presumably less susceptible to being category-mistakenly understood in the terms of a causal thing. (Think, for instance, of how we speak more readily of '*having* reasons' – as we do of *having* things, like cars – but of '*being under* obligations'.)

¹⁷⁵ Darwall 2006, p.94, referencing Williams 1985.

this is the category-mistaken construal of Mill's Point. (Moreover, this argument doesn't actually entail the existence of second-personal reasons, as I have argued previously.¹⁷⁶)

What drives Darwall back into the category-mistake of inferring the existence of reasons, as above, is apparently the worry that, on the Ryle-inspired reading, "normative reasons [would] follow from the fact of someone's being blamed".¹⁷⁷ That is, if all there is, on the Ryle-inspired reading, to being obligated is relating to a person in a way in which the first is liable to being held accountable by the second, then any case of a second person holding a first accountable will apparently entail the first's obligation to the second. But, surely, there have to be some independent facts – *reasons* (whatever their nature) – that make it the case that a first person is obligated to a second. Unless there are such existing reasons that make it the case that a person is *justly* liable to being held accountable, then we could go around holding *anyone* accountable, thereby, on the Ryle-inspired reading, 'making it the case' that those randomly-picked agents were morally obligated – and even 'making it the case' that those randomly-picked agents were *guilty* of violating those obligations, obligations that had just appeared, as if out of thin air, in virtue of their being held accountable. But, naturally, this is to put the cart before the horse. Rather, Darwall's thinking here seems to be, we hold agents accountable *because* they are antecedently guilty of transgressing an independently obtaining reason – not the other way around.

But, this is *precisely* the category-mistaken thinking Ryle has warned us against. Thinking that we hold agents accountable *because* they are independently guilty is to claim that there is a thing called 'guilt' that causes us to hold agents accountable. But, what could this thing called 'guilt' be? Where is it? These are category-mistaken questions.¹⁷⁸ Guilt isn't that kind of thing in the first place. Guilt is not a state, but a *status*: a standing one has amongst fellow persons (and: just a standing one has amongst fellow persons, not a standing one has amongst fellow persons *in virtue of* some state of one's being). It is a way of relating to people. It is not *implied by* the reciprocally recognitional practice of holding accountable – it *is* a way of reciprocally recognizing another. *That's* the relevant category here: ways of relating to people – and when we leave that category, we make a category-mistake.

Darwall, I have argued in this section, has trouble keeping within the appropriate category: sections of *The Second-Person Standpoint* approximate it, but others transgress the boundaries into category-mistakenness. The impulse behind such excursions, it seems, is the worry that we need some

¹⁷⁶ See Ch.3 and 4.

¹⁷⁷ Darwall 2006, p.94, as cited above.

¹⁷⁸ Cf. The Cartesian Dualist's presumption that the mind must be of *immaterial* substance when he fails to find the mind anywhere in the world (Ch.6, s.2).

independent states in order to justify our holding agents accountable. I consider a possible Rylean solution to this worry, in closing, below.

4. Conclusion

The Ryle-inspired suggestions of this chapter are intriguing, but difficult, ideas. I am unsure if they are Darwall's ideas. As I have tried to show in this chapter, there is room for reading *The Second-Person Standpoint* as a statement of these quite radically different, category-unmistaken ideas, although not all of Darwall's claims in that book will be consistent with this reading. On the other hand, *The Second-Person Standpoint* might receive a much more natural reading – as I have done for the bulk of this project. That more natural reading, however, does not fare so well, as I have shown over the course of Chapters 2 to 5. And, presumably, that more natural reading does not fare so well, because, as I hope to have shown over the course of this present chapter and the preceding one, that more natural reading commits a category-mistake in the way it conceives of moral reasons. Moral reasons are not things – especially not things given to first persons by second persons – that cause moral action. Rather, they are a matter of a special way of relating to people: a way of relating to people premised upon reciprocal recognition, in particular that variety of reciprocal recognition by which a second person holds a first accountable.

If Darwall does wish to advocate this radically different approach, then I conclude with a suggestion. Treating moral reasons as a way of relating to people according to which a first person is liable to being held accountable by a second leaves us with the question of when, under what conditions, an agent is *justly* held accountable.¹⁷⁹ And, in order to avoid backsliding into the category-mistake, the Second-Personal approach needs to be able to address this question without positing a *thing* – like a category-mistaken conception of a reason – that makes it the case that an agent is guilty and justly held accountable. Now, such a category-mistaken metaethic strikes me as analogous to a Referential theory of meaning,¹⁸⁰ according to which the meaning of a word is given by the *thing* it picks out. Likewise, Ryle¹⁸¹ and – more famously – his contemporary, Wittgenstein¹⁸² were struck by the Referential theory of meaning's analogy to the category-mistaken Cartesian Dualist theory of mind. If a Rylean or Wittgensteinian philosophy of language is able to provide conditions under which a word is used *correctly*, without making reference to the things that those words pick out, then I suggest that an analogous metaethic will be able to provide conditions under which an agent

¹⁷⁹ See s.3.3, above.

¹⁸⁰ Or, perhaps suggestively, a *Millian* theory of meaning.

¹⁸¹ Ryle 1932, as referenced by Tanney 2009, p.liii.

¹⁸² Wittgenstein 1953.

is held accountable *justly*, without inferring things like reasons by which those agents failed to be led to action.¹⁸³ I suggest that this is the way forward for a successful Second-Personal approach to morality.

Asking, as I did at the beginning of this project, what the nature of a moral reason is – as Ryle would put it – an “improper question”.¹⁸⁴ Reasons simply aren’t of the category of things that have ‘natures’ in this sense. They are moral reasons of our own, intrinsic to persons – but not in the sense that they are things *possessed* by us nor *given* to each other. Rather, on the Ryle-inspired reading of this chapter, being obligated is a matter of relating to another person in a particular kind of way – a Second-Personal way.

¹⁸³ We might read Korsgaard’s (1996) employment of Wittgenstein’s (1953) Private Language argument in this manner.

¹⁸⁴ Ryle 1949/2009, p.12ff.

References

Please note that, where more than one date of publication has been provided, the more recent version has been used. Page numbers referenced in the body of the text refer to the pages of the more recent version.

- Anscombe, G.E.M. (1958) 'Modern Moral Philosophy'. Retrieved March 23, 2014, from: <http://www.pitt.edu/~mthomпсо/readings/mmp.pdf>
- Appiah, K.A. (2003). 'An example: Free will and determinism', s.9.10 in *Thinking It Through: An Introduction to Contemporary Philosophy*. New York: Oxford University Press. pp.365-377.
- Avramides, A. (2001) *Other Minds*. Abingdon: Routledge.
- Ashford, E., & Mulgan, T. (2012) 'Contractualism', in E.N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved May 29, 2015, from: <http://plato.stanford.edu/entries/contractualism/>
- Blackburn, S. (1990) 'Just causes', *Philosophical Studies*, vol.61(1). pp.3-17.
- Boyd, R.N. (1988) 'How to be a moral realist', Ch.9 in G. Sayre-McCord (Ed.) *Essays on Moral Realism*. Ithaca, New York: Cornell University Press. pp.181-228.
- Boyd, R.N. (2003) 'Finite beings, finite goods: the semantics, metaphysics and ethics of naturalist consequentialism, Part I', *Philosophy and Phenomenological Research*, vol. 66(3). pp.505-553.
- Brewer, T.M. (2008) 'Morality and the second person', *The Philosophical Quarterly*, vol.59(234). pp.163-167.
- Brink, D.O. (1989) 'Objective utilitarianism', Ch.8 in *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press. pp.211-290.
- Cohen, G.A. (1996) 'Reason, humanity, and the moral law', in O. O'Neill (Ed.) *The Sources of Normativity*. Cambridge: Cambridge University Press. pp.167-188.
- Darwall, S. (1989) Review of J. Hampton's (1986) *Hobbes and the Social Contract Tradition*, *The Philosophical Review*, vol.98(3). pp.401-404.
- Darwall, S. (2006a) 'Contractualism, root and branch: a review essay' (Review of T.M. Scanlon's (2003) *The Difficulty of Tolerance*), *Philosophy & Public Affairs*, vol.34(2). pp.193-214.
- Darwall, S. (2006) *The Second-Person Standpoint: Morality, Respect and Accountability*. Cambridge, Massachusetts: Harvard University Press.
- Darwall, S. (2007) 'Reply to Korsgaard, Wallace, and Watson', *Ethics*, vol.118(1). pp.52-69.
- Darwall, S. (2009) 'The Second-Person Standpoint: an interview with Stephen Darwall', *Harvard Review of Philosophy*, vol.XVI. pp.118-138. Retrieved May 3, 2014, from: <http://www.harvardphilosophy.com/issues/XVI/Interview%20with%20Stephen%20Darwall.%20The%20Second-Person%20Standpoint.pdf>

- Darwall, S. (2010a) 'Authority and reasons: exclusionary and second-personal', *Ethics*, vol.120(2). pp. 257-278.
- Darwall, S. (2010b) 'Bipolar obligations', presented at the *Seventh Annual Metaethics Workshop: University of Wisconsin, Madison*. Retrieved August 12, 2011, from: https://mywebspace.wisc.edu/shaferlandau/web/metaethics/workshop_2010/papers/darwall.pdf
- Darwall, S. (2010c) 'Moral obligation: form and substance', *Proceedings of the Aristotelian Society*, vol.110(1). pp.31-46.
- Davidson, D. (1963) 'Actions, reasons, and causes', *The Journal of Philosophy*, vol.60(23). pp.685-700.
- Davies, M., & Stone, T. (1998) 'Folk psychology and mental simulation', in A. O'Hear (Ed.) *Contemporary Issues in the Philosophy of Mind*. Cambridge: Cambridge University Press. pp.53-82.
- DeLapp, K.M. (2009) 'The merits of dispositional moral realism', *The Journal of Value Inquiry*, vol.43(1). pp.1-18.
- Dennett, D.C. (1981) 'True believers: the intentional strategy and why it works', Ch.3 in J. Haugeland (Ed.) (1997) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Massachusetts: Massachusetts Institute of Technology Press. pp.57-79.
- Dullstein, M. (2012) 'The second person in the theory of mind debate', *Review of Philosophy and Psychology*, vol.3(2). pp.231-248.
- Estlund, D.M. (2009) *Democratic Authority: a Philosophical Framework*. Princeton: Princeton University Press.
- Fleischacker, S. (2009) 'Review of Stephen Darwall, *The Second-Person Standpoint: Morality, Respect and Accountability*', *Utilitas*, vol.21(1). pp.117-123.
- Frankfurt, H.G. (1971) 'Freedom of the will and the concept of a person', *The Journal of Philosophy*, vol.68(1). pp.5-20.
- Frankfurt, H.G. (2004) *The Reasons of Love*. Princeton: Princeton University Press.
- Gallagher, S. (2005) *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Gertler, B. (2008) 'In defense of mind-body dualism', in J. Feinberg & R. Shafer-Landau (Eds.) *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*. Belmont, California: Thomson Wadsworth. pp.285-297.
- Geuss, R. (1996) 'Morality and identity', in O. O'Neill (Ed.) *The Sources of Normativity*. Cambridge: Cambridge University Press. pp.189-199.
- Harman, G. (1977) 'The problem with ethics', Pt.1 in *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press. pp.3-26.
- Hobbes, T. (1651) *Leviathan*, in R. Tuck (Ed.) (1996) *Cambridge Texts in the History of Political Thought: Hobbes: Leviathan: Revised Student Edition*. Cambridge: Cambridge University Press. pp.86-482.
- Hornsby, J. (2000) 'Personal and sub-personal: a defence of Dennett's early distinction', *Philosophical Explorations*, vol.3(1). pp.6-24.

- Hughes, P.M. (2013) 'Forgiveness', in E.N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved August 1, 2015, from: <http://plato.stanford.edu/entries/forgiveness/>
- Hume, D. (1739-1740) *A Treatise of Human Nature*. Edited by J. Bennett (2010). Retrieved January 2, 2013, from: <http://www.earlymoderntexts.com/pdfs/hume1739book1.pdf>, <http://www.earlymoderntexts.com/pdfs/hume1739book2.pdf> and <http://www.earlymoderntexts.com/pdfs/hume1740book3.pdf>
- Hurka, T. (2014) 'Moral realism', s.4.1 of *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press. pp.86-92.
- John Paul II (1993) *Veritatis Splendor*. Retrieved May 29, 2015, from: http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_06081993_veritatis-splendor.html
- Joyce, R. (2015) 'Moral Anti-Realism vs. Realism: explanatory power', in E.N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. Retrieved September 7, 2015, from: <http://plato.stanford.edu/entries/moral-anti-realism/moral-realism-explain.html>
- Kant, I. (1785) *Groundwork for the Metaphysic of Morals*. Translated by M.J. Gregor (1997) and edited by J. Bennett (2010). Retrieved January 2, 2013, from: <http://www.earlymoderntexts.com/pdfs/kant1785.pdf>
- Kar, R.B. (2007) 'Symposium: the second-person standpoint and the law: introduction', *Loyola of Los Angeles Law Review*, vol.40(3). pp.881-890.
- Korsgaard, C.M. (1993) 'The reasons we can share: an attack on the distinction between agent-relative and agent-neutral values', Ch.10 in (1996) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press. pp.275-311.
- Korsgaard, C.M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, C.M. (2007) 'Autonomy and the second person within: a commentary on Stephen Darwall's *The Second-Person Standpoint*', *Ethics*, vol.118(1). pp.8-23.
- Lavin, D. (2008) 'Review of *The Second Person Standpoint: Morality, Respect and Accountability*', *Notre Dame Philosophical Reviews*. Retrieved May 2, 2014, from: <http://ndpr.nd.edu/news/23292-the-second-person-standpoint-morality-respect-and-accountability/>
- Lerm, J. (2012) 'Second-personal reasons: why we need something like them, but why there are actually no such things', *South African Journal of Philosophy*, vol.31(2). pp.328-339.
- Lerm, J. (2013) 'The agent-relative/agent-neutral distinction: my two sense(s)', *South African Journal of Philosophy*, vol.32(2). pp.137-148.
- Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*. London: Penguin Books.
- Marshall, R. (2013) 'From the second person', interview of S. Darwall, *3:AM Magazine*. Retrieved May 3, 2014, from: <http://www.3ammagazine.com/3am/second-person-ethics/>
- McMyler, B. (2011) *Testimony, Trust and Authority*. Oxford: Oxford University Press.
- Metz, T. (2013) *Meaning in Life*. Oxford: Oxford University Press.
- Miller, C. (2006) 'Shafer-Landau and moral realism', *Social Theory and Practice*, vol.32(2). pp.311-331.

- Moore, G.E. (1903) *Principia Ethica*. Retrieved January 2, 2013, from: <http://fair-use.org/g-e-moore/principia-ethica>
- Moore, M.S. (2010) 'Causation and the permissibility of consequentialist justification within agent-relative morality and the law', Ch.3 in *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics*. Oxford: Oxford University Press. pp.34-80.
- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Morissette, A. (1995) 'You oughta know' in *Jagged Little Pill*. Milwaukee: MCA Music Publishers, H. Leonard Corp.
- Nagel, T. (1970) *The Possibility of Altruism*. Oxford: Clarendon Press.
- Nagel, T. (1986) *The View from Nowhere*. Oxford: Oxford University Press.
- Phillips, D. (2011) 'Sidgwick's metaethics', Ch.2 of *Sidgwickian Ethics*. New York: Oxford University Press. pp.10-51.
- Prichard, H.A. (1912) 'Does moral philosophy rest on a mistake?', in (1949) *Moral Obligation: Essays and Lectures by H.A. Prichard*. Oxford: Clarendon Press. pp.1-17.
- Pufendorf, S. (1672/1729) 'Of the duty of keeping faith, and the diversity of obligation to it', in *On the Law of Nature and Nations*. Translated by B. Kennett (1729). London: J. Walthoe, R.Wilkin, J. & J. Bonwicke, S.Birt, T.Ward, and T. Osborne. pp.252-259. Retrieved December 20, 2012, from: <https://archive.org/details/oflawofnatureat00pufe>
- Pufendorf, S. (1672/2003) Excerpts of *On the Law of Nature and Nations*, translated by C.H. Oldfather & W.A. Oldfather (1934), in J.B. Schneewind (2003) 'Samuel Pufendorf' in *Moral Philosophy from Montaigne to Kant*. Cambridge: Cambridge University Press. pp.170-182.
- Railton, P. (1986) 'Moral Realism', *The Philosophical Review*, vol.95(2). pp.163-207.
- Russell, B. (1951) 'My mental development', Ch.2 in (1961/2009) *The Basic Writings of Bertrand Russell*. New York: Routledge. pp.9-22.
- Ryle, G. (1949/2009) *The Concept of Mind* (60th anniversary ed.). Abingdon: Routledge.
- Scanlon, T.M. (1998) *What We Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press.
- Shafer-Landau, R. (2003) *Moral Realism: a Defence*. Oxford: Clarendon Press.
- Shoemaker, S. (1994a) 'Self-knowledge and "inner sense": lecture I: the object perception model', *Philosophy and Phenomenological Research*, vol.54(2). pp.249-269.
- Shoemaker, S. (1994b) 'Self-knowledge and "inner sense": lecture II: the broad perceptual model', *Philosophy and Phenomenological Research*, vol.54(2). pp.271-290.
- Stace, W.T. (1952) 'The problem of free will', in J. Feinberg & R. Shafer-Landau (Eds.) (2008) *Reason & Responsibility* (13th ed.). Belmont, USA: Thomson Wadsworth. pp.248-258.

- Sturgeon, N.L. (1985) 'Moral explanations', in J. Rachels (Ed.) (1998) *Ethical Theory 1: The Question of Objectivity*. Oxford: Oxford University Press. pp.180-209.
- Sturgeon, N.L. (2003) 'Moore on ethical naturalism', *Ethics*, vol.113 (3). pp.528-556.
- Strawson, P.F. (1962) 'Freedom and resentment', Ch.7 in S. Guttenplan, J. Hornsby & C. Janaway (Eds.) (2003) *Reading Philosophy: Selected Texts With a Method For Beginners*. Malden, MA: Blackwell Publishing. pp.193-209.
- Tanney, J. (2009) 'Rethinking Ryle: a critical discussion of *The Concept of Mind*', in *The Concept of Mind* (60th anniversary ed.). Abingdon: Routledge. pp.ix-lvii.
- Van Helden, A., featuring Harden, D. (1999) 'U don't know me', in (1998) *2 Future 4 U*. Somerville, Massachusetts: Armed Records.
- Wallace, R.J. (2007) 'Reasons, relations, and commands: reflections on Darwall', *Ethics*, vol.118(1). pp.24-36.
- Ware, O. (2009) 'Fichte's voluntarism', *European Journal of Philosophy*, vol.18(2). pp.262-282.
- Watson, G. (2007) 'Morality as equal accountability: comments on Stephen Darwall's *The Second-Person Standpoint*', *Ethics*, vol.118(1). pp.37-51.
- Werner, R. (1983) 'Ethical realism', *Ethics*, vol.93(4). pp.653-679.
- Wikipedia (2013) 'Gilbert Ryle', in *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. Retrieved October 17, 2013, from: http://en.wikipedia.org/wiki/Gilbert_Ryle
- Wikipedia (2014) 'Armin Meiwes', in *Wikipedia: The Free Encyclopedia*. Wikimedia Foundation, Inc. Retrieved March 10, 2014, from: http://en.wikipedia.org/wiki/Armin_Meiwes
- Williams, B. (1979) 'Internal and external reasons', in (1981) *Moral Luck*. Cambridge: Cambridge University Press. pp.101-113.
- Williams, B. (1985) *Ethics and the Limits of Philosophy*. Cambridge, Massachusetts: Harvard University Press.
- Wittgenstein, L. (1953) *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.
- Wood, A.W. (2006) 'Fichte's intersubjective I', *Inquiry*, vol.49(1). pp.62-79.
- Yaffe, G. (2010) 'Comment on Stephen Darwall's *The Second Person Standpoint: Morality, Respect and Accountability*', *Philosophy and Phenomenological Research*, vol.81(1). pp.246-252.