The Pursuit of Repair

Kirstine la Cour

University College London (UCL)

Ph.D. Philosophy

Declaration

I, Kirstine la Cour, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis develops an account of the possibility and desirability of interpersonal moral repair. A dominant strain of philosophical thought has cast interpersonal moral repair as a problem of redistributive or corrective justice. According to this picture, to wrong another is to unfairly deprive her of goods that are rightfully hers, and to repair is to compensate her for these losses, leaving her no worse off than she would otherwise have been. Despite its wide acceptance, I argue that this conception of repair is both explanatorily and ethically untenable. In its place, I develop an account that centres the need for mutual understanding. Philosophers have underestimated how agentially demanding the pursuit of mutual understanding can be, but also failed to see its profound creative potential. The pursuit of mutual understanding, I argue, provides opportunities for challenging and revising our interpersonal norms and values, and not just for applying or expressing the commitments we already hold. Appreciating this creative potential of repair illuminates the value, and not just the cost, of our mutual vulnerability. In the course of developing and defending this proposal, I also offer an account of testimony as a joint project and of protest as a means for generating and sustaining a sense of self-respect.

Table of Contents

Declaration 2

Abstract 3

Impact Statement 6

Acknowledgements 7

UCL Research Paper Declaration Form 9

Chapter summaries 11

Part I 16

Chapter One 17

The problem of repair 17

- The need for repair 19
- The neglect of repair 22
- Explanatory alternatives and Scepticism about repair 26
- The shape of the problem 28

Chapter Two 32

The Aims of Apology: 32

The Bookkeeping Model of Repair and its Discontents 32

- A top-down approach 35
- The bookkeeping model illustrated 37
- Apology as Penance 43
- Apology as self-denigration 49
- Apology as Reassurance 57
- Conclusion to Part I 65

Part II 72

Chapter Three: 73

Telling takes two 73

- From Indicative signals to social acts 76
- Social acts and the Active-Passive distinction 84
- Telling and the impossibility of resistance 88
- The threat of unilateralism 91

Chapter Four 94

Complex testimony and the social character of speech in interaction 94

From simple to complex testimony 98

Communicative repair 103

Co-narration..110

Speech in interaction as a collaborative undertaking 114

Conclusion to Part II: Exchange Fetishism and how to avoid it 118

Part III 124

Chapter Five..125

Repair as the Pursuit of Mutual Understanding 125

The what of repair: Repair as a defence of intersubjectivity 129

Wrongdoing as an epistemic challenge 136

The *how* of moral repair and the role of narrative..139

Exculpation, Condonation, and the threat of collapse 145

The alchemy of apology revisited..151

Chapter Six..155

Standing up for yourself: Protest as self-creation and self-discovery 155

Introduction 156

Boxill's Puzzle..158

Self-respect and worth: The Dignity View 161

Self-respect and Worth: The Honour View 164

An impasse?..167

Self-creation: Taking Responsibility for Oneself 168

Self-discovery: Knowing oneself as self-respecting 172

Conclusion 175

Conclusion 176

Bibliography 182

Impact Statement

The work presented in this thesis stands to impact future scholarship on repair, communication, and epistemic and normative agency, not just within philosophy, but also in neighbouring fields such as law, public policy, political theory, sociology, or the experimental communication sciences, as well as projects involving collaborations between these fields. Its expected outputs will in the first place include academic journal articles and conferences presentations, but will also in the longer term include both trade and specialist books, as well as publications for a general audience on blogs, news sites, and in online magazines.

The problem of repair I take up in this dissertation is familiar and significant in most people's lives. By reaching not just academic researchers interested in the topics of this project, but also members of civil society, policy makers, educators, and ordinary citizens, the work presented here hopes to spur deeper reflection and richer conversations, and thereby itself to further support the pursuit of repair.

Acknowledgements

This thesis is in part about the value of communication, and the manner in which what is said by one speaker can be influenced by the agency of her interlocutors. In developing this project, I have had many valuable interlocutors.

John Hyman, Stephen Darwall, Lucy O'Brien and Han van Wietmarschen have all supervised my work for parts of this thesis. I thank them for their support and contributions.

Ancestors of the chapter materials presented here have also benefitted from the input of audiences at several academic fora, workshops, and conferences. I would like to extend my thanks to audiences in the University of Kent, the Hebrew University, Vienna University, Yale University, Tel Aviv University, Birmingham University, Manchester University, Lund University, the UCL Department of Linguistics and Psychological Sciences, and the UCL Department of Philosophy.

I am particularly grateful for the extensive and insightful feedback from two work-in-progress groups I have had the pleasure of participating in. My thanks to Lea Salje, Daniel Morgan, Alex Geddes, Edgar Phillips, Lucy Campbell, Alexander Greenberg, Ashley Shaw, Alec Hinshelwood, and Vanessa Carr; and to Daniel Ranweiler, Daniel Telech, Larisa Svirsky, James Lewis, Romy Eskens, Shawn Wang, Samuel Dishaw, and Sandford Diehl.

For conversations and suggestions at different stages of this project that pointed me in the right direction, I offer my sincere thanks to Marcia Baron, David Enoch, Elizabeth Fricker, Miranda Fricker, Bart Geurts, Maximilian Kiener, Isabelle Laurenzi, Colleen Macnamara, Yuuki Ohta, and David Zapero. Thank you also to Elizabeth Camp for providing access to an in-press manuscript.

A special thanks to treasured present and former colleagues at UCL, including Rebecca Rowson, Nikhil Venkatesh, Luke Devereaux, Catherine Dale, Tasnim Ismael George Britten-Neish, Alexander Murphy, Christabel Cane, Jack Hume, Filippa Ronquist, Alice Harbert, Lizzy Holt, Fiona Wittingham, Kenta Sekine, Michael Thorne, Victor Braga, Mark Kalderon, Simona Aimar, and Ulrike Heuer.

My thanks above all to Julian Bacharach without whose unfailing support, encouragement, and extensive conversational participation this project would not exist.

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. For a research manuscript that has already been published (if not yet published, please skip to section 2)

a) **What is the title of the manuscript?**

Click or tap here to enter text.

b) **Please include a link to or doi for the work**

Click or tap here to enter text.

c) **Where was the work published?**

Click or tap here to enter text.

d) **Who published the work?** (e.g. OUP)

Click or tap here to enter text.

e) **When was the work published?**

Click or tap here to enter text.

f) **List the manuscript's authors in the order they appear on the publication**

Click or tap here to enter text.

g) **Was the work peer reviewed?**

Click or tap here to enter text.

h) **Have you retained the copyright?**

Click or tap here to enter text.

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

Click or tap here to enter text.

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

□

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

a) **What is the current title of the manuscript?**

REDACTED

b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

No

c) **Where is the work intended to be published?** (e.g. journal names)

REDACTED

d) **List the manuscript's authors in the intended authorship order**

Kirstine la Cour (single authored)

e) **Stage of publication** (e.g. in submission)

REDACTED

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

N/A

4. **In which chapter(s) of your thesis can this material be found?**

Chapter Six

5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Kirstine la Cour

Date:

29.11.2024

Chapter summaries

This thesis comprises six chapters over three parts. Chapter One, *The Need for Repair*, introduces the central problem of the thesis: the problem of repair. Sometimes wrongful actions have an afterlife: whether done by us or to us, they can lodge themselves recalcitrantly in the texture of our relationships, imposing moral psychological burdens, and making it hard or even impossible to sustain our bonds with others. In at least some of these cases, we generally think it is both *possible* and *desirable* for the wrongdoer to do something to set things right. Wrongdoing gives her a kind of task – a task at which she might succeed.

The problem of repair as I shall pursue it in this thesis asks for an account of this task. How is it undertaken, and what are the norms and goals that structure its pursuit? I also motivate both the practical and the theoretical importance of this problem, briefly address its uneven philosophical history, and clarify the terms and bounds of my inquiry.

Chapter Two *The Aims of Apology: The Bookkeeping model of repair and its discontents* provides an extended critique of a dominant approach to the problem of repair – I call it the Bookkeeping model of repair. According to the Bookkeeping model, to wrong another is to unfairly deprive her of goods that are rightfully hers, and to repair is to repay or compensate her for these losses, leaving her no worse off than she would otherwise have been.

I first note a number of apparent advantages of this model. For instance, it promises precision and determinacy in the conditions for success and secures explanatory continuity between interpersonal morality and social-institutional normative domains. In addition, the bookkeeping model is resistant to critiques often made of it, such as that it makes it possible to pay-off victims, or makes all moral values fungible.

However, the bookkeeping model faces other challenges. I survey three versions of the view that differ above all in their proposed *currencies* of wrongdoing and reparative repayment. The Penance view focusses on welfare and suffering; the Self-denigration view on status and social power; and the Reassurance view on doxastic states like beliefs or expectations.

Each version, I argue, faces serious explanatory and ethical challenges that threaten to make the pursuit of repair irrational, impossible, or morally misguided. A proponent of the bookkeeping model could try to revise her position to evade these problems, or come up with a fourth or fifth possible currency for reparative restitution. However, I end by arguing that there are reasons to be suspicious of the bookkeeping approach to moral repair as such.

Part II, comprising Chapters Three and Four, departs from the problem of repair to gather resources for the positive proposal developed in Chapter Five. It does so by considering a debate over the agential character of testimony, taking as its starting point a contribution to this debate made by Richard Moran. Moran argues that it takes two to tell: a speaker who openly and intentionally offers her assurances for the truth of some claim, and a hearer who gives it recognition and uptake. By means of this exchange, a normative relationship is generated between the parties: the speaker confers a set of normative privileges and entitlements on a hearer, including the right to hold her to account if what he has been told turns out to be false.

Though I endorse the notion that telling takes two, I argue that Moran's conception both over- and under-estimates the social character of testimony. The first part of this argument is made in Chapter Three, *Telling Takes Two,* and the last in Chapter Four, *Complex Testimony and the Social Character of Speech in Interaction*. In Chapter Three, I argue that Moran *over-estimates* the social character of telling by drawing too close a comparison between testimony, promising, and contract. The central problem for this account, I argue, is that it is possible to tell someone something despite their being unwilling to hear or understand it (what I term *the impossibility of testimonial resistance*); *pace* Moran, in transmissions of information of this kind, the hearer is not a collaborative partner but rather a passive recipient for the speaker's unilateral exercise of a normative power to bind.

However, as I argue in Chapter Four, this description applies only to individual, unambiguous, self-standing utterances (what Rachel Fraser has called *simple testimony*). Much of our testimony is not like this, however; instead, it is offered in ordered sequences of interconnected claims, sometimes over multiple conversational 'turns' at talking. *Complex* testimony like this, I argue, involves and requires an altogether different degree of cooperation between speaker and hearer than philosophers have generally acknowledged. To establish this, I draw from a variety of studies in the empirical communication sciences, and particularly from conversation analysis and socio- and psycho-pragmatics. As these resources show, speakers rely on the active interventions of their interlocutors in several ways: to signal when communication is going awry (in cases of misspeaking, mishearing, or misunderstanding), or what is also termed *communicative repair*; to assist and facilitate the narration of first-personal experiences, termed *co-narration*; and to elicit and give shape to a speaker's self-understanding.

These findings seriously challenge traditional mainstream epistemological assumptions and ideals about the autonomy and self-sufficiency of epistemic agents, and about the relative passivity and intersubstitutability of hearers. If I am right, hearers themselves bring distinctive abilities and capacities to bear in communication, and have the power to curtail or obstruct a speaker's exercise of epistemic agency if they are unwilling or unable to collaborate.

Part III, comprising Chapters Five and Six, returns to the problem of moral repair, and to the other motivations an agent may have for responding to wrongdoing committed against him.

In Chapter Four, *Repair as the Pursuit of Mutual Understanding*, I draw from the findings of Chapters Three and Four to answer the challenge raised across Chapters One and Two. I first argue that communicative repair as introduced in the previous chapter provides an apt and illuminating model for moral repair: in both cases, the pursuit of repair arises in response to a manifest lack of, gap in, or threat to the parties' *mutual* *understanding* – their underlying operative stock of assumptions and procedures for reciprocal intelligibility and attitudinal alignment. And in both cases, the necessary remedy is an extended collaborative undertaking that aims to reestablish a common footing.

I propose that wrongful action, like a misdirected conversational move, raises a kind of epistemic challenge: it reveals or suggests that one's partner is not operating with the same rules and norms of conduct, not applying them in the same way, or not giving the same weight to the risks and costs. It therefore calls for a kind of inquiry to determine where the parties have diverged and why. Accordingly, the pursuit of repair is the pursuit of mutual understanding. This pursuit is undertaken by communicative means - the parties need to name and coordinate on the source of their problem, acknowledge its significance, and place it within a narrative understanding of their present situation.

I also consider a set of central challenges to my proposed account of repair, including the notion that it collapses repair into either exculpation or condonation of wrongdoing. The challenger concedes that understanding one another can be a morally valuable undertaking, but denies that doing so in itself amounts to repair. If understanding reconciles, the objection holds, it is because it has struck upon an exculpatory consideration, or else because it has motivated or made possible simply *setting the matter aside*. I argue that this challenge can be set aside as illegitimately begging the question against my proposal by implicitly importing the assumptions of the bookkeeping model itself.

I end by offering a positive vindication for the moral value of mutual understanding both as an essential resource for interpersonal intimacy with our loved ones, and as an occasion for normative contestation and development.

Chapter Six, *Standing Up for Yourself: Protest as Self-creation and Self-discovery* carries over many of the themes of the foregoing, but extends the inquiry by considering a significantly different case: responding to unjust treatment for one's own sake when one has no prospect for or interest in repair with the wrongdoer. More specifically, I take up a proposal due to Bernard Boxill that a victim of injustice should stand up for herself to show and know her self-respect.

Considering two prominent accounts of the basis for self-respect, the Dignity View and the Honour View, I argue that neither is capable of resolving Boxill's Puzzle. They make the connection between an agent's action I argue that neither view can explain the epistemic role of protest, making the connection between a person's worth and her knowledge of it either too tight or too tenuous. Instead, I develop and defend an alternative. Starting from the idea that a person's worth can be 'up to her', I propose that one can *take responsibility* for one's self-respect – for the person one is, or is becoming. Doing so, however, requires an act of public commitment; a person can stand up for herself by standing up before others. In this manner, protest can be a form of self-creation. I then consider whether an act of commitment can also be genuinely self-revelatory. Its apparently self-fulfilling character might seem to preclude that it can, but I suggest this conclusion can be avoided. If so, protest can be a form of self-discovery as well.

I end by summarizing the central findings of the project and suggesting avenues for further research.

Part I

Chapter One

The problem of repair

Oliver: "Jenny, I… I'm sorry!"
Jenny: "Don't! Love means never having to say you're sorry!"
Love Story, 1970

Dr Bannister: "That's the dumbest thing I ever heard!"
What's Up Doc? 1973

It became the catchphrase of the 1970 hit movie *Love Story:* "Love means never having to say you're sorry". Drawn from the best-selling Erich Segal (Segal, 1970) novel of the same name, the line is spoken by the rain-drenched and quivering cold Jennifer Cavalieri to her lover Oliver Barrett as he attempts to apologise to her for an angry outburst earlier that day. The line so resonated with audiences that in 2005, it was voted into 13th place of the American Film Institute's 100 most memorable movie quotes of all time (coming in only a couple of places after Taxi Driver's "You talkin' to me?", and several places ahead of both "E.T. phone home" and "Rosebud"[1]).

The line expresses a conception of the relationship between loving relationships[2] and apologising, but which one? I offer you three interpretations. On the first

[1] ("AFI's 100 YEARS…100 MOVIE QUOTES," n.d.)

[2] In talking about loving relationships and lovers, I mean to include not just romantic relations, but also familial or platonic ones.

interpretation, love means never having to say you are sorry because loving someone precludes doing to them such things as would require an apology. Lovers unfailingly treat each other with kindness, care, and respect, and as a result never have anything to for which to apologise[3].

On the second interpretation, love means never having to say you are sorry, not because lovers never transgress against one another, but rather because within a loving relationship, forgiveness is always immediate and unconditional; whatever wrongs are done between lovers are always already forgiven, and never dependent upon an apology being given and received.

Finally, on the third interpretation, the emphasis should go on the word 'say': love means never having to *say* you're sorry - not because lovers never transgress, and not because such transgressions are automatically and unconditionally forgiven, but rather because what matters is not *saying* sorry, but *being* sorry; and when you are sorry, your beloved will already know that you are. Between people who understand one another as lovers do, this interpretation has is, what one feels does not require explicit articulation.

Most of us, I submit, do not subscribe to any of these conceptions: we occasionally (even frequently) both give and receive apologies within relationships our loved ones. If this is so -and I here invite the reader to check against their own biography- none of these interpretations provides a faithful description of love, or of apology, as we practice it.

This is perhaps not so surprising; after all, the kinds of relationships depicted in schmaltzy romance movies rarely match real life. But is it regrettable? We might think that the notion that *love means never having to say you're sorry* expresses an ideal (or perhaps several ideals at once) to which our loving relationships aspire, or ought to aspire. If we were only less fallible, or more forgiving, or had a sufficient depth of

[3] The famous line makes a second appearance in *Love Story* in a confrontation between Oliver and his father at the movie's end, and on this occasion the intended interpretation is arguably the first one: Oliver is rebuking his father for hostility and rejections that have estranged the pair from one another in Oliver's great time of need.

understanding of one another, apologising really *would be* superfluous within our loving relationships, as it supposedly is between Segal's star-crossed lovers.

I believe each of the three 'ideals' articulated above should be resisted; neither we nor our relationships would be better if we never had recourse to apology[4]. This may sound strange; surely, a world without apology, i.e. a world in which we were all perfectly considerate and self-possessed, perfectly generous with others' flaws, or perfectly understanding of each other would be a better world than this one? In defence of the contrary position, I shall be making a case for the positive value of apologising, and for its integral role not just in loving relationships, but in the social reproduction of morality itself.

The need for repair

It is a fact of social life that we sometimes transgress against others, or others against us. Only the naïve, the self-deceived, or the comprehensively isolated could expect never to hurt or harm another, or to be hurt or harmed in turn. These transgressions range from the trivial (shoving and pushing on public transport; inconsiderate or injudicious remarks; impositions and inconveniences) to the serious (betrayals of loyalty, or honesty, or trust; breaches of promises; neglect of others' interests and concerns) or the outright egregious (assaults on our fundamental rights and freedoms; violations of human dignity and bodily autonomy; destruction of lives and limbs); many of them are commonplace, and others thankfully rare.

The reality of such transgressions gives rise to the need for repair, at least some of the time. That it does so is itself noteworthy and even somewhat surprising: why do we not simply let bygones be bygones, put our troubles behind us, and move on? But wrongful actions, it seems, have an afterlife; mistakes and missteps once committed cast long shadows and weigh us down with regret or resentment, grief or guilt from afar, allowing a troubled past to persist in the present.

[4] The line itself was parodied in the Peter Bogdanovich comedy "What's Up Doc?" a few years later, this time delivered by Barbra Streisand's lovable but impetuous and trouble-prone Judy to Dr Howard Bannister, portrayed by *Love Story*'s Ryan O'Neal. This time, it meets a more sceptical response.

The peculiar staying power of wrongdoing is not the only instance in which we observe the moral psychological hold of the past. We see similar dynamics in cases of regret over bad moral luck or roads not taken, in grief for the loss of a loved one to illness, accident, or old age, anger at past injustices, or in the continuing toll of traumatic formative circumstances. We also see in the positively valanced analogues of all these cases – continuing delight or pride in past triumphs, gratitude for the good fortune and generous services one has received, or a sense of security and contentment built upon a long history of nurture and care.

While some would argue this beholdenness to the past is irrational (or at least that it is irrational whenever it pains or burdens us[5]), others have argued that it is itself a sign of a healthy moral psychology, a realistic, humane outlook, or a natural and unavoidable corollary of meaningful attachments to others or to oneself and one's projects.

The latter perspective has, I believe, been the philosophically dominant one. In the case of responses to wrongdoing, this likely owes at least in part to the ongoing influence of Strawson's seminal *Freedom and Resentment* (Strawson, 1962) with its insistence that

> "in the absence of any forms of these attitudes it is doubtful whether we should have anything that we could find intelligible as a system of human relationships, as human society."(Strawson, 1962, p. 210)

By 'these attitudes', Strawson means the reactive attitudes, e.g. our responses to the conduct or quality of will evinced by others or by oneself, a list that includes "resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other"(Strawson, 1962, p. 194). The alternative to vulnerability to these attitudes, Strawson held, is a comprehensive isolation or (what arguably amounts to the same) a thoroughgoing indifference to others that make interpersonal relationships as we know them impossible.

[5] See e.g. Bittner on regret (Bittner, 1992); or Sommers on the objective attitude (Sommers, 2007).

However, it is one thing for these attitudes to arise at all, and another for them to persist. Most of those who maintain that being emotionally affected by the past can be a sign of good moral psychological health also generally acknowledge that the presence of the past *can* become pathological. Righteous resentment can turn into senseless grudge-holding or navel-gazing nursing of old sores[6]. What prevents this from happening in the 'good' case? If we are initially right to be affected by what others do, but wrong to remain so indefinitely, what gives?

An old adage has it that *time heals all wounds*, and some philosophers have recently taken a similar line. For instance, Oded Na'aman (Na'aman, 2021, 2020) and Berislav Marusić (Marusić, 2022) have each proposed that our past-directed attitudes and emotions are processes that fittingly evolve and conclude over time. If so, perhaps the best response to past wrongdoing, whether done to us or by us, is simply to sit tight and wait it out.

In some contexts, this approach seems perfectly adequate. A slight or disappointment is sometimes rather like a grazed knee or an elbow to the ribs: the sting consumes your attention in the short term, but after a while it subsides and passes of its own accord[7]. At other times, however, past transgressions seem to recalcitrantly lodge themselves in our hearts and minds, or, in Aurel Kolnai's phrase, in the 'texture' of our relationships (Kolnai, 1974, p. 101). Left unaddressed, they continue to fester and may even grow more painful and significant than they initially seemed. More pressingly still, we feel we cannot, and should not, move forward without doing something about them. It is this phenomenon – the sense of past wrongdoing as not just a moral, but a *practical* problem, something that sets us a certain kind of task – that I shall have in mind by the label *the problem of repair*.

Most of us, I submit, recognise the need for repair, so construed. We at least sometimes feel called upon to answer for ourselves when we have done wrong, and at least sometimes expect and desire the same from those who wrong us. I say 'answer for'

[6] Not everyone accepts that lasting emotional responses must be pathological. See e.g. (Callard, 2020). However, as the rest of that collection suggests, Callard's is a minority position.
[7] It may also prove an attractive characterisation of other past-directed attitudes and affective states, such as grief over the loss of a lost one.

purposefully, because the practices we typically use for this purpose are communicative – paradigmatically, we apologise, express our sorrow, or ask for forgiveness.

Even more interestingly, we sometimes experience these practices as successful – as sufficient for allowing both parties to move on, put an end to their conflict, and even resume a harmonious relationship. What makes the problem of repair so puzzling, I suggest, is quite how unequal the apparent remedy can seem to the affliction at hand; how is it that simple bits of speech –apologising– can overcome breaches of trust, hurt feelings, or inflictions of suffering and distress that are themselves irreversible?

Confrontation with this challenge has led some authors to suggest that there is something magical, mystical, or otherwise supernatural in apology's reparative powers, something which defies attempts at rational explanation, or at least appears to do so, lest we fail to look closely enough[8].

The neglect of repair

I have argued that confrontation with the problem of repair can hardly be avoided; in sharing a life with other people, we are constantly liable to being the victims or the agents of wrongdoing. The impact of this reality is no mere momentary frustration or discomfort; the tolls of wrongdoing can be lasting, and may both burden the individual and obstruct her valuable relationships with others. The problem so characterised should be familiar to anyone with the slightest bit of moral experience; the question of how to respond to wrongdoing may be one of the few ethical quandaries every person is all but certain to confront in their own life, and to confront again and again and again.

This combination of ubiquity and significance makes the relative neglect of the problem in the history of moral philosophy all the more notable. Few of the discipline's celebrated figures or traditions devote attention to repair at all; what is

[8] See e.g. (Rushdy, 2015); (Tavuchis, 1991); (Jankélévitch and Hobart, 1996); (Arendt, 1958); (Nussbaum, 2016). This tension between irreversibility and change is often a starting point for philosophical explorations of forgiveness, e.g. as in (Arendt, 1958); (Kolnai, 1974); (Calhoun, 1992); or (Hieronymi, 2001).

more, several ethical positions are set up so as to deny the possibility of a distinct problem about responding to wrongdoing at all. For instance, as e.g. Charles Griswold has argued, the resolute perfectionism of Aristotelian ethics precludes serious investigation of repair; the *megalopsuchos* has no need either to forgive or be forgiven, since his own conduct is by definition morally faultless, and since he is immune to moral injury from those inferior to himself (Griswold, 2006, pp. 8–9).

Consequentialism, too, provides little scope for raising or even recognising a distinctive question about *what one ought to do now*, in light prior wrongdoing. For a consequentialist, what matters in deciding what to do is always the same: how much good one can (expect to) do on balance by selecting each available alternative. A given agent's particular situation and history can shape the answer to this question in that it places them in *particular proximity to* or perhaps *in exclusive command of* or certain levers of influence – for example, the fact that I am the one who hurt your feelings may indicate a present opportunity to do good (say, by cheering you up). However, if someone else is better placed to take up this task, or if I myself could do more good by investing my energies elsewhere, then that is what I ought to do instead, and there is nothing regrettable about it - no loss in letting past failures to act well go unaddressed.

Some historical figures did acknowledge and contend with the problem, at least briefly – perhaps most notably, Kant and Hegel[9], both of whom connect it to defence of a retributivist position on punishment. A kind of principle for reparation also appears in the work of both Henry Sidgwick and W. D. Ross[10]. But other moral philosophers may have found it a matter best left for theologians or lawyers.

This approach has broadly been mirrored in the academic discipline. First order moral theorising as we learn it and teach it in introductory ethics courses generally concerns itself with a different set of questions altogether: What is it to act well? What is the relation between the right and the good? What makes for a virtuous life or agent?

[9] In the *Religion within the Limits of Bare Reason* and the *Elements of the Philosophy of Rights*, respectively. My thanks to Dan Ranweiler for valuable discussion on this point.

[10] I owe the observation of this point to Jeremy Watkins (manuscript).

What ought I to do, and to refrain from doing? Starting in this way gives the impression that ethical questions are questions of individual action and reflection, arising for those who are free to follow an argument wherever it leads, and starting from a blank page, more or less *ex nihilo*.

The problem of repair exposes the inadequacy of this characterisation of the situation of an ethical subject. As moral agents, we are constrained in myriad ways by limitations to our powers of thought and action, and encumbered by existing bonds and histories that are themselves not of our making. Questions about what to think, feel, or do always arise for us *in situ,* and typically from positions we have not freely chosen and cannot easily shed. We not only want but need to share our lives with other people, whose ideas and desires often diverge significantly from our own. Operating within these constraints, we often get it wrong, and are perhaps bound to do so, and not least with those people we care about most of all. The problem of repair, then, compels us to confront our liability to moral failure, but also our deep concern for and vulnerability to other people.

Recent moral philosophy marks a significant change. The enduring influence of a small number of seminal contributions from the latter half of the 20th century have spurred and held the interest of a new generation of thinkers, and a flurry of activity in the last few decades has put especially *blame* and *forgiveness* on the map of central ethical concerns[11]. But even here, repair itself (and apology as its principal practical manifestation) have received significantly less attention than its sequential neighbours.

That is, philosophers have devoted a lot of attention to uncovering and investigating the normatively appropriate antecedents or pre-conditions of repair – to what it takes to be responsible; who or what is to be excluded or included in the practice and on

[11] I have in mind the altogether unavoidable "Freedom and Resentment" (Strawson, 1962), but also Aurel Kolnai's "Forgiveness" (Kolnai, 1974), and Jean Hampton and Jeffrie Murphy's jointly authored *Forgiveness and Mercy* (Murphy and Hampton, 1988), in the absence of which the contemporary debate would look and sound significantly different.

what grounds; and what it takes in practice to hold someone to account in light of their wrongful action.

Much of this work *presupposes* that repair is possible and sometimes normatively difference-making, at least in theory if never in practice – for instance, those who emphasise the communicative character of responsibility practices like blame, or more broadly its role in agential formation and development typically suppose that blame can motivate and spur pursuits and attainments of repair. Even those who are sceptical that judgements of moral responsibility are or can be placed on a metaphysically robust footing can find pragmatic justifications for retaining these practices by pointing to their role in promoting restitution for victims or better future behaviour on the part of the wrongdoers[12].

Philosophers have also devoted a lot of attention to the normatively appropriate consequences of repair – the reliance of forgiveness upon repair; the value or virtue of forgiveness; or the limitations on forgiveness on what and how we can forgive.

Again, much of the work on these questions *presupposes* that repair is possible and sometimes normatively difference-making, at least in theory if never in practice – for example, the distinction between gifted and un-gifted or earnt forgiveness relies on this possibility, as does much work distinguishing forgiveness from condonation, or explaining how forgiveness can be consistent with retaining self-respect.

Despite its intimate connections with many of these well-developed research areas, then, the possibility of repair is comparatively under-explored. If we think of accountability practices as a sequence of normative changes, each depending upon, and in an intuitive sense *responding to* the one that precedes it[13], we find a major gap at the centre of our pictured sequence of blame, apology, and forgiveness.

For the purposes of this dissertation, I mean my investigation of this gap – of its conceptual shape, and the possibilities of filling it in - to be neutral between different

[12] See e.g. (Pereboom, 2012)

[13] I take this description to be one many participants in the debate would accept. One particularly explicit endorsement of it is in McKenna (McKenna, 2011); but even here, it is notable how little is said about how repair and apology effects the transition from blame to forgiveness.

conceptions of the phenomena that precede or follow it, viz. paradigmatically blame and forgiveness. I take on this objective so as not to pre-judge or beg the kinds of questions I cannot explore sufficiently thoroughly in this thesis without wandering too far afield of my primary focus - the pursuit of repair itself. However, no element in the sequence exist in a vacuum. Once the elements of my negative and positive projects in this thesis start to fall into place, it should become clear that the account of repair I defend has both upstream and downstream consequences – it affects what positions we can and ought to hold on the conditions of responsibility, on inclusion and exclusion within our accountability practices, and, most directly, on whether, when, and why we should forgive.

Explanatory alternatives and Scepticism about repair

There are a few different ways of taking on board the characterisation of the problem of repair as I've described it, yet remaining sceptical about either the desirability or the possibility of repair itself.

For instance, some philosophers recommend that the task of responding to wrongdoing be placed with victims themselves. Despite the many differences in their outlooks, both Martha Nussbaum (Nussbaum, 2016) and Cheshire Calhoun (Calhoun, 1992) adopt positions along these lines. Both are principally interested in forgiveness, and both argue that philosophers have overemphasised the need to base forgiveness on a wrongdoer's subsequent acts of repair – her apologies, repentance, amends-making, restitution, and whatever else we may see fit to include in this list. The kind of forgiveness that a wrongdoer earns or deserves through these responses (what Calhoun calls *minimalist forgiveness*, and Nussbaum *transactional forgiveness)* is both morally worse and rationally unstable or misguided[14], they argue – it is not the form of forgiveness to which we aspire or ought to aspire.

[14] In Nussbaum's description, the mistake is a form of retributivist magical thinking about which more shall be said in the next chapter; for Calhoun, it is a 'double vision' that impossibly tries to see the wrongdoer as both culpable and no-longer-culpable at once. (Calhoun 81-83).

There are many reasons for commending these forms of forgiveness and jettisoning or deemphasising a concern with the wrongdoer's repair. For instance, it might be argued that victims of wrongdoing benefit from seizing the initiative and reasserting their agency of their own accord, rather than waiting around for responses from wrongdoers that may never materialise. In other cases, wrongdoers are no longer available, or, while available, cannot be trusted to respond appropriately and ought instead to be steadfastly avoided. Alternatively, one may hold that there is simply nothing a wrongdoer can do to make up for her prior transgressions, and that a victim of wrong therefore needs to rely only on himself, or perhaps on his community of peers; perhaps the wrongdoer truly is incorrigible, or her deed to heinous for repair. Forgiveness may still be possible in such cases, and where it is, it is attributable to the victim's generosity, grace, or personal emotional development, and not to anything the wrongdoer has done to merit or otherwise promote the change.

I happily concede that the visions of gifted forgiveness recommended by Calhoun, Nussbaum, and others who to some extent turn away from repair[15] may be both appropriate and admirable in such cases. We should indeed make space for the possibility that a victimised person's perspective on the wrongdoing can change unilaterally in the absence of any effort by the wrongdoer, and that this can be an achievement that is to his moral credit.

Nonetheless, this perspective need not encourage a wholesale indifference to or suspicion of repair. Even if forgiveness without repair is both possible and morally valuable, we clearly sometimes do pursue repair, whether as the wronged or wrongdoing party, and find these pursuits morally meaningful and difference-making. To hold that we are always wrong to do so – embracing a morally defective practice in its place – would be to embrace a thoroughly revisionist position on our moral practices. In what follows, I will work on the assumption that we ought to at least try to make sense of our manifest commitment to repair before embracing the

[15] See e.g. (Fricker, 2018) or (Norlock, 2008).

kind of position Calhoun, Nussbaum, and other sceptics about repair would recommend.

I shall also argue that, while unilateral forgiveness has its place, it is not without its costs. When we individually decide to forego repair, sweeping wrongdoing under the rug, or privately deciding to set it aside or rise above it, we also cultivate a distance and alienation within our relationships. Repair, I will argue, can itself cultivate the kind of intimacy our close bonds with others require. Though repair itself requires a great effort, and though our pursuits of it may not succeed, the attempt is often worth the risk.

There is one further repair-sceptical attitude to briefly canvas. It might be agreed that apologising can provide repair, but also held there is simply no unified story to be told about when, how, or why. Perhaps interpersonal morality is irreducibly particularistic and a search for general explanatory mechanisms and principles is in vain. While I have considerable sympathy for this kind of particularistic scepticism, in what follows I will set it too aside, and instead pursue lines of argument that aim to provide a unified account of the underlying mechanism and practice of repair.

While saying this, I am also committed to making space for normative criticism. While we want our account to encompass the real variability in the phenomenon as we know it, we must not simply take all aspects of it as given and outside the scope of critique. It should be possible both to make sense of repair, and to raise and pursue ethical questions about it.

The shape of the problem

In starting investigations from the reality of the phenomenon as we know it, it will be useful to make note of some of the additional features we are looking to capture.

First, instances of repair do not always match incidents of wrongdoing. Some instances of wrongdoing go unaddressed, and some instances of morally unproblematic behaviour give rise to pursuits of repair.

In some cases, repair does not occur where it could or should because the parties fail to recognise that something wrongful has taken place. Wrongdoings that are not acknowledged or understood as such do not become repairable. But in other cases, one party makes a conscious choice to *forego* repair although they suspect, or even confidently know that they have wronged or done wrong. We should be able to make sense of this choice and the basis for making it, or for commending or criticising it in one another[16].

We should also note and bear in mind that repair is not always attained where it is pursued. Even earnest attempts at repair can fail. Resolution sometimes comes far easier and more quickly than expected, and at other times comes not at all.

As may already have been clear, cases where a conflict is unilaterally set aside will not count as cases of repair within my picture. Repair as I shall understand it must involve some kind of intentional pursuit (though not necessarily intentional under that guise), and thus excludes reconciliation brought on by external conditions – for instance, if a sudden disaster instinctively leads us to set our petty lovers squabble aside, we will have reconciled but not have repaired. It is hard to draw a precise line here – a sudden disaster can spur and facilitate repair as well - the intended contrast is hopefully sufficiently clear. In characterising repair as a *task*, I am committed to the view that it the distinctive kind of reconciliation it secures cannot be brought about willy nilly by a sudden bout of double amnesia or a magical pill.

The key point for my purposes is that we should be able to accommodate a relatively wide-range of variation in how repair is actually pursued or achieved – cases diverge on many different points; if there were a one-size-fits-all method for repair, it would not be as philosophically and practically vexing a problem as it evidently is – whilst being able to maintain that some central explanatory mechanism recurs across cases. Our search, then, is for this mechanism, and with it hopefully some tools to help us

[16] Though this will not be my focus, it also seems possible to pursue repair when one in fact knows it is not appropriate to do so. One could pursue repair disingenuously and strategically to gain some sort of advantage, just as mock politeness is possible and sometimes advantageous even when one knows it is inauthentic.

understand why more specific demands arise under certain circumstances and not others.

A final note is in order on my choice to focus on repair that occurs within interpersonal relationships. Why start an inquiry about apology and moral repair here in particular? Some would argue that doing so is sure to distort the search from the beginning, seeing as such relationships are, by their nature, *special*; they involve a richer set of demands and expectations, levels of familiarity and intimacy, or intertwined personal histories that set them apart from the kinds of bonds we have to mere acquaintances, to strangers who pass through our lives, or to the extended moral community as such, most of the members of which we will never know, meet, or even think about.

I by no means deny that loving relationships are different in morally relevant respects; however, like a generation of writers in the feminist ethics literature, I believe the majority of our ethical concepts and practices are most at home in this setting, and so that ethical inquiry ought to start from within it and work outwards, rather than the other way around – indeed, I am convinced by those who argue that our ethical inquiries have a tendency wind themselves into irresolvable knots when we try to work in the other direction and justify our special attachments or obligations from a vantage point where they are designedly absent.[17].

It is far from obvious why personal attachment to a person or relationship should *preclude* application of properly ethical concepts to it, or, on the contrary, why a purely abstract schematic relationship, stripped of all specificity, provides a superior setting for ethical investigation[18]. Clearly, one of these is significantly 'cleaner', and therefore in one sense easier to 'work with', philosophically speaking. If I am describing to you my relationship with a purely hypothetical agent A, it is easy to determine that I have captured all and only the features of that agent and that relationship that are 'relevant' to the moral question at hand; no such features exist but the ones I have stipulated for

[17] See for instance work by Margaret Urban Walker on partiality and special obligations (Walker, 1997, p. 85pp), Annette Baier's argument about the primacy of interpersonal trust in (Baier, 1986); or Bernard Williams' famous discussion on having a thought too many(Williams, 1981).

[18] Strawson seems to have taken this position, as have many figures in mainstream analytic moral philosophy.

the sake of the argument (i.e. that A is a child, or a psychopath, or a moral expert or saint; that A has made me a promise to Q, or has a stepped upon my gouty toe, or betrayed my trust; that A is a stranger to me, or a member of my moral community, or 'a friend', as though that label alone were sufficiently informative). When I ask of such circumscribed conditions what I ought to do with respect to A, it looks perfectly possible to extract, within some perimeters, some concrete and plausible implications.

By contrast, suppose I were to raise a similar question about my mother –not a thought experiment cipher of my mother, but the real person that she is, complete with a history and perspective and mind of her own. How would you even begin to describe all the features and conditions and all of the history that bear on what I have done and ought to do next?

There is a careful balance to tread here between a stance that looks to make moral theorising resolutely impossible (since everything is too complexly situated and contextual to ever be captured by theory) or on the other hand a stance that stipulatively tidies the moral setting to such an extent that the clear implications and prescriptions our theorising offers us lack application to any real world situation. It is worth noting that both of these stances stifle the possibility of action guidance.

Accordingly, I shall start from the assumption that close interpersonal relationships provide an appropriate setting for raising and pursuing these questions – at a minimum, a setting that does not compromise or distort the inquiry itself.

Chapter Two

The Aims of Apology:

The Bookkeeping Model of Repair and its Discontents

In the previous chapter, I set out the central concern of this thesis: the problem of moral repair. I also noted some of the parameters that will structure the search for a conception of repair capable of addressing this problem. In particular, we are looking for an account that can explain why repair is sometimes both possible and desirable, and which allows us to see a unified mechanism operating across diverse instances. The account should be such as can sensibly be held to a degree of fit with recognisable experiences of repair -a standard with its own attendant methodological challenges– but also one which allows for the possibility of normative critique; we should be able to explain why someone might choose to forego repair, and explain why pursuits of repair sometimes succeed and sometimes do not. Finally, it should make some sense of the appearance that communicative practices like apologising are typically a crucial part of repair and intuitively sufficient for repair in many cases.

The objective of this chapter will be to examine a specific way of talking and thinking about repair which has, I propose, become prominent in the philosophical literature. There are different ways into and through this story, differences in the detail and the presentational emphases, and variation in how openly the organising assumptions are articulated and embraced in the work of different authors. Nevertheless, the core of the view is, I believe, sufficiently stable and sufficiently widespread to be worth investigating. I call it the Bookkeeping Model of Repair.

A few passages from Joel Feinberg (1970) will help bring the model into view:

> "We say that persons deserve compensation for harm wrongly inflicted by others, in which case it is called "redress of injury", "amends", or "reparation" and functions not only to repair the damage but also to "restore the moral equilibrium", as would an apology or expression of remorse. Reparation "sets things straight" or "gives satisfaction"". (p. 74)
>
> "If reparation is to be received by a victim, it would seem that it must be given by a wrongdoer; and it seems to follow that, if one person deserves to take, another deserves to give. … [T]he wrongdoer deserves to be held liable for the harm he has caused; he deserves to be forced to compensate his innocent (or relatively innocent) victim." (p. 75)
>
> "Reparation can express sympathy, benevolence, and concern, but, in addition, it is always the acknowledgement of a past wrong, a "repayment of a debt", and hence, like apology, the redressing of the moral balance or the restoring of the *status quo ante culpum*."(p. 76)

From this initial presentation, we can assemble our own list of the central elements; first, wrongdoing affects us and our relationships by disturbing a pre-existing moral balance or equilibrium, and the aim of repair is to restore that order, and thereby reestablishing the *status quo ante*. Second, wrongful actions themselves are instances of unjust deprivation or loss; accordingly, the type of rebalancing they call for is compensation or repayment. Third, the wrongdoer deserves to be held liable for these restitutive efforts, while by contrast the victim is innocent, or at least relatively so; it is her, i.e. the wrongdoer that must *give*, in order that her victim can *receive*[19]. Finally, while repair *may* do various other things besides (for instance, express sympathy, benevolence, and concern), the balance-restoring repayment is their fundamental function and the one which repair *always* performs or ought to perform[20].

I call this the Bookkeeping model because it so neatly fits the elements of basic accounting; each party to a conflict is represented by a column on a balance sheet; when a wrongdoing occurs, goods or values are removed from one column and/or added to another without adequate reason; the remedy against this is for the 'debitor' to repay or compensate the 'creditor', thereby rebalancing the books and restoring the equilibrium that existed before the wrong. While the model (or something sufficiently like it) is sometimes recognised by that label in the literature, it is also variously described as an economic model (Warmke, 2016), an indebtedness model (Griswold, 2006; Radzik, 2009b), or – particularly by its detractors - a transactional model (Nussbaum, 2016).

The bookkeeping model has a number of apparent attractions. Most obviously, it makes the otherwise puzzling remedial potential of moral repair look not just practically tractable but comprehensible and recognisable. When the normative significance of wrongdoing is cached out in terms of an unjust allocation of goods or distribution of benefits and burdens, the supernatural-seaming ability to 'reverse' or 'undo' the past is domesticated into something much more familiar: the prosaic process of balancing the books. While the wrongful action itself remains irreversible, its normative impact no longer is: an illegitimate transaction can be cancelled out and

[19] Throughout this thesis, I shall use the female pronoun for the wrongdoer (and eventually for the speaker) and the male pronoun for the victim of wrongdoing (and eventually the hearer), unless explicitly indicating otherwise.

[20] Note that there is something ambiguous about the relation Feinberg draws between 'reparation', 'amends making', and 'redress' on the one hand and 'apology' or 'expressions of remorse' on the other. It is somewhat unclear whether he believes reparation to be a *functionally equivalent alternative to* apology, or rather that apology is one of the forms – perhaps the primary form – reparations take. Since my interest here is to explore our attachment to apology in particular, I will be working with the latter interpretation.

nullified by a corresponding or reversing transaction that restores the previous allocation of goods.

In addition to familiarising the mechanism of repair, the model seemingly affords us clear and repeatable action-prescriptions for its pursuit, and determinate conditions of success: Repair requires no more and no less than restoring to each party what they are rightfully owed by the other, viz. that which they each had before the wrongdoing. To say this much is not to say that repair is easy; clearly, some complex moral tasks remain for one who is applying this framework to a concrete situation – identifying and typing each incurred cost or gain; accurately 'pricing' them in restitutive actions; putting the reversing repayment into practice, etc. However, it is reasonably clear how the procedure would go and where it should conclude. The bookkeeping model, then, supplies both a unifying underlying mechanism for repair, and a set of practical guidelines for its implementation.

A top-down approach

The bookkeeping model has another noteworthy advantage. By emphasising notions like restitution, desert, and allocative fairness, the bookkeeping model stresses the continuity between interpersonal moral repair and notions of corrective or (re)distributive justice that are recognisable from legal and political philosophy. The problem of moral repair, the model proposes, is simply one more species or variant of these latter, and can consequently be elucidated and resolved by applying the conceptual tools, explanatory frameworks, and justificatory strategies already familiar from legal and political normative domains.

I call this argumentative strategy a *top-down approach* to repair: approaching a phenomenon in interpersonal morality by starting from the more abstract and general principles found within legal, political, and institutional normativity and reducing down.

Moral repair is not the only interpersonal practice with a legal, political, or institutional analogue, and so not the only candidate for top-down theorising. For instance, though she does not use the same terminology, Annette Baier argues - albeit in critique – that philosophers have approached interpersonal trust by taking contractual agreement as a theoretical model and explanatory starting point(Baier, 1986). If Baier is right, the philosophers she critiques can accordingly be said to take a top-down approach to trust[21].

A top-down approach to repair itself promises us further advantages; if, as I argued in the introductory chapter, interpersonal moral repair, and apology in particular, has been afforded relatively scant attention in the philosophically literature, conceptions of legal and political corrective justice have not. Though there is by no means universal agreement in these literatures, there is at least broad consensus around certain ideas and approaches, and a wealth of contributions devoted to stress testing and refining them and working out their implications across contexts[22]. Applying these well-developed resources in a 'new' or underexplored area of inquiry therefore allows us a significant head start. It also promises to make philosophical analyses satisfyingly parsimonious; there is no need to reinvent the wheel in the interpersonal context and needlessly multiply accounts if we already have models at our disposal that can cover and unify both interpersonal and institutional forms of repair.

There may be a further reason why the bookkeeping model's top-down approach has recommended itself to many philosophers. Many philosophical investigations of repair and forgiveness foreground instances of wrongdoing that are particularly egregious and shocking in their scale and scope; genocide, political persecution and oppression, systematic and structural violence, profound losses of human dignity, or

[21] I will later return to Baier in making a similar proposal about testimonial commitment in Part II.

[22] I am particularly thinking of the distributive paradigm as an approach to justice in political philosophy, and the consideration of retributive or distributive motivations for tort and criminal law. Among the proponents of the former, I count Rawls and those who have taken up his legacy in focussing attention around questions about the currencies of equality and the conditions for fair liability to have and keep various goods. Iris Marion Young (Young, 2011, 1990)and Elizabeth Anderson (Anderson, 1999) are among the notable critics of this approach to justice in political philosophy, as is – albeit for very different reasons – Robert Nozick(Nozick, 1974), whose own conception of justice nonetheless remains staunchly allocative. In tort law, I particularly have in mind Gardner(Gardner, 2013, 2011), who in turn aligns himself with the work of Weinrib and Coleman, and in the debate on retributivism in criminal punishment particularly the work following Morris(Morris, 1971, 1968), including Hampton(Hampton, 1991; Murphy and Hampton, 1988).

deprivations of basic rights and protections[23]. Clearly, these are cases that directly raise legal and political concerns. If much philosophical work on repair is animated by a concern with cases of this kind, it is unsurprising that notions of legal or political justice are salient, and that the resulting understanding of interpersonal repair strive (implicitly or explicitly) to be answerable to and continuous with standards of justice applicable in these contexts.

The bookkeeping model illustrated

To get a better handle on the bookkeeping model, it may help by starting with the kind of case where it is easiest to apply, viz. wherein the wrongdoing to be repaired involves a concrete and quantifiable material loss. Suppose A wrongs B by taking B's bicycle without his knowledge or agreement. To repair this wrongdoing, says the bookkeeping model, A must make up for the shortfall thereby imposed on B: A should return the bike, and cover any additional costs B incurred as a result of the unexpected lack of the bicycle during the period in question – for instance, if B ended up having to get a cab, A should pay the cost of it, or cover the late fee for the appointment B missed because of her inference. The standard for A's having successfully repaired her transgression is that she makes B no worse off than he would have been, were it not for her wrongfully taking the bike.

In the background of this picture is the assumed satisfaction of a variety of other conditions contributing to the appropriateness of repair. For instance, we are assuming that A acted intentionally, was neither excused nor justified in taking B's bike, as she would have been if, say, she has mistaken B's bike for her own similar looking one in their communal bike shed[24], or if she had taken it only because she was rushing to aid someone in a medical emergency. We are assuming also that it really is

[23] E.g. (Walker, 2006a) on the political violence and disappearances in Argentina and Chile, and Black Reparations for enslavement; Nussbaum on South African Apartheid(Nussbaum, 2016); (Radzik, 2009b)on the Magdalen Laundries; Griswold (Griswold, 2006)on the Holocaust.

[24] Most philosophers think that while justification absolves a person of the duty to repair, excuses simply *lesson* or *modify* the duty to repair, without removing it altogether. Excuses, then, are scalar (A is more excused if B's bike is identical to hers than it if is merely similar and would have been distinguishable with more care and closer scrutiny, say) whereas justifications are binary. I am setting these complications aside for the moment, focussing just on cases where the wrongdoing is *neither* justified nor excused to any extent.

rightfully B's bike, and that the use B had intended to put it to when he was prevented by its absence was not a morally nefarious one[25].

Even given the assumed satisfaction of these background conditions, however, further complications ensue almost immediately: what should count as 'additional costs' incurred by B as a result of A's wrongdoing? Suppose B, unexpectedly deprived of his intended means of transportation, booked a luxurious limousine ride, when he might instead have made his journey by bus. Or suppose he rushes to purchase a new bicycle during the period he lacks the first one, not knowing that A means to return it later. Should A's repair additionally cover these costs? Or suppose that the meeting B missed was about a promising investment opportunity, which, it transpires, would have landed B a substantial windfall. Is A liable for the loss of the windfall? If the Kingdom is lost for want of a nail, does the nail-thief owe the value of a Kingdom in return?

But this problem can be addressed. A proponent of bookkeeping can agree that there is no *pre-theoretically obvious* or *uncontroversial* cut-off point separating the ramifications of wrongdoing the repairing agent is liable for from those she is not liable for. However, the model does not require an obvious or universally agreed-upon boundary; it requires only that it is possible to draw a line that *could* be defended and agreed upon by rational and reasonable people, whether on moral, metaphysical, or pragmatic grounds[26]. That is to say, perhaps what A owes to B is not obvious to us *now* from the brief schematic description I gave of their case; however, with sufficient time and information, we could come to a stable and satisfactory answer.

But what about losses not easily priced in monetary terms? Say the lack of the bike caused B to be late for an important job interview or workplace meeting, doing damage to B's career prospects and professional reputation. Or say it caused B to miss his daughter's piano recital, leading to a serious rift in *their* relationship. Or say was an appointment for cancer treatment, which cannot be rescheduled for several weeks

[25] Again, it is not uncontroversial that B would lack a right to repair if these conditions were not met, but I am setting these complications aside as well.

[26] Tort law has made a significant headway on the challenge with the aid of additional subsidiary principles, such as contributory negligence or remoteness of damage. See (Gardner, 2013, p. 24)

or months, during which time B's health will deteriorate. How could A repay B for these kinds of injuries and losses?

This line of questioning suggests a more fundamental challenge for a bookkeeping model: that it treats all gains and losses resulting from wrongdoing as fungible, and therefore always in principle offsetable for a price. Doing so, the objector holds, is to commit a pernicious moral mistake. It is false that all moral values can measured and compared on a single scale, and to claim otherwise is both offensive to victims and dangerous for the moral community at large: it in effect treats any act of wrongdoing as acceptable so long as the offender subsequently pays the 'fine', which is disrespectful to the wronged, and generates a de facto moral immunity for those wealthy enough to pay up again and again[27].

But again, a proponent of the bookkeeping model can argue that it is perfectly capable of addressing or avoiding these problems. It is true that not all effects of wrongdoing involve *monetary* losses or allow obvious conversions for compensatory damages in monetary terms; it is indeed hard to put a price on damage to reputation, or losses of affection, say. This clearly poses a theoretical challenge for the bookkeeping model, but not obviously an insurmountable one. There are two different tacks to explore. First, a proponent of the bookkeeping model could dig in her heels, insisting that while monetary equivalences are hard to come by, they are nonetheless possible. Second, a proponent of the bookkeeping model could concede that not all losses are fungible but maintain that we can still use the bookkeeping model's reasoning to guide our efforts in repair, since repayments and compensations can themselves be granted in a variety of moral currencies. I shall briefly consider each of these in more detail.

In pursuing the first response, a proponent of the bookkeeping model would argue that putting a monetary price on a non-monetary moral value or good is frequently hard, but nonetheless possible. Indeed, she might argue, we do it all the time – in prioritising limited funds in healthcare, in setting up insurance markets and

[27] See e.g. (Griswold, 2006; Helmreich, 2015; Radzik, 2009b) .

purchasing insurance products, or in making personal financial decisions, we regularly do assign a price to losses and hardships of various kinds. Some may find it uncomfortable to confront this fact, but such squeamishness must not be confused for theoretical impossibility. This first strategy therefore mirrors the one noted above about drawing a line between consequences of wrongdoing for which a wrongdoer is liable and those for which she is not.

In response, the objector might maintain that the basis of his objection is not mere squeamishness about fungibility, but an insistence that important moral values are being neglected, overlooked, or relevantly misrepresented. Healthcare policy, insurance markets, and personal financial decision-making all involve attempts to navigate complex and multi-faceted moral realities under pragmatic constraints. The apparent financial equivalences we arrive at in these contexts are only approximations, and only fit for narrow, localised purposes – not genuinely reflective of our real first-order moral commitments.

A disagreement on this issue will be difficult to resolve. Proponent and opponents of fungibility are working from fundamentally different moral starting points, and neither is likely to be impressed by the other's arguments[28]. However, if we are impressed by the further prong of the critique, according to which the fungibility of losses and gains would imply that the wealthy can simply buy their way out of repairs demands, this first line of response will not do, and can therefore serve as a reason for favouring the second approach: that repayments or compensations can be provided in many different currencies, not all (or even any) of which are inter-translatable.

A proponent of the second response would concede that the conversion of moral losses to monetary damages is sometimes (or often, or always) impossible: some forms of moral injury cannot be priced and repaid in cash alone. However, she will argue, the bookkeeping model itself does not require that it can; instead, it can accommodate that wrongdoing may imposes losses or debts in a variety of diverse 'currencies'; some

[28] The self-same disagreement occurs in many other moral contexts, and most notably in confrontations between consequentialists, who maintain that different costs and benefits can be offset against each other, and non-consequentialists, who deny this very fact.

monetary, some material, some physical, some psychological, and so on, and accordingly that repair itself may require both monetary, material, physical, and psychological restitution. As such, simply buying one’s way out of repair’s demands is not always (or perhaps ever) possible.

In sum, then, the bookkeeper needs not be guilty of assuming the crude commensurability of all goods and ills on a single value metric, let alone a monetary one. Nothing in the model precludes the possibility of multiple incommensurate ‘currencies’ for repair; indeed, it might be thought a strength of the view that it allows us to make sense of a principle or mechanism that unifies the pursuit of different such values.

Moreover, a proponent of a bookkeeping model need not hold that perfect or complete restitution is always possible. For instance, Swinburne writes:

> “Sometimes [a wrongdoer] can literally restore the status quo. If I steal your watch and have not sold it, I can return it to you. Sometimes I can only make things rather similar to the way they were, so that the victim is almost equally happy with the new state. I can compensate him adequately, that is. If I steal and sell your watch I can buy you another one. If I smash up your car, I can pay for the repairs. The harm done by stealing, injuring, and similar acts is not only the physical damage, but the inconvenience of temporary loss and the trauma and anxiety resulting from it, and for these too compensation is needed. Sometimes, alas, full compensation is not possible. If I run you over with my car, and paralyse you for life, nothing I can do can compensate you fully for that. But some things which I can do can compensate you in part. I can pay for wheelchairs, and machines to life you out of bed in the morning. But clearly reparation, as far as lies within the wrongdoer's power, is essential for removal of the taint of guilt”(Swinburne, 1989).

This sets us two standards for repair: in the ideal case, the wrongdoer makes the victim as well of as he would (and therefore, ought) to have been; in the realistic case, wrongdoer approximates this level *as far as lies within her power*. The occasional inaccessibility of complete repair does not mean partial repair cannot be a worthy and relevant goal, nor that the bookkeeping model does not have the right conception of the task.

Finally, in considering the question of the relevant baselines for repair, it should also be noted that in requiring for repair that the parties be restored to the *status quo ante culpum* (or as close to it as lies within one's power), the bookkeeping model is implicitly assuming also that the antecedent state was at least morally adequate. However, this is not always the case. If by contrast the relationship between the parties that pre-existed wrongdoing was itself a morally defective one –say, if the future victim was himself oppressing or attacking the future wrongdoer when she struck back– the question of what the demands of repair require become more complex. In such cases, a proponent of the view can hold that repair instead requires establishing a well-balanced relationship *for the first time*. If a history of unsettled moral debts already exists between the parties, both of them will have a role to play in bringing this about.

This introduction has hopefully provided an initial impression of the bookkeeping model as an approach to the problem of repair. However, while I have emphasised its theoretical strengths and attractions, and argued it can satisfactorily defend itself against common lines of attack, my defence is qualified and temporary. Ultimately, the argument of this chapter will be that the model does not succeed, or rather, succeeds only at far too high a cost.

Over the next few sections, I will be considering a number of different iterations of a bookkeeping model found in the philosophical literature. They vary principally in their favoured currencies of repair, and in the good they see restitution in these currencies to deliver for the parties. At first glance, it may sound strange that there should be a task of spelling out what good restitution provides; restitution, it might be thought, is morally desirable as a matter of definition. But this is too quick. John Gardner argues that to say some norm is a norm of *justice* is only to describe its subject matter; it remains still to be established that it is a sound norm, a norm we have good reason to adopt and uphold (Gardner, 2011, pp. 14–17). So too the bookkeeping model's conception of repair as restitution: restitution is its subject matter, but to determine if repair so conceived is a norm, or a normative practice, we should adopt and uphold, we need to examine if it is morally sound. We need to ascertain, in Gardner's terms, *what it has going for it*.

In carrying out this assessment, I will have my eye on a few different matters: First, normative adequacy: Does the account provide a moral vindication of repair, rather than show repair to be normatively problematic or defective?[29] Second, explanatory adequacy: Does the account provide a satisfactory elucidation of the phenomenon of repair? A satisfactory elucidation, as I shall understand it, must make some sense of the underlying mechanism(s) that deliver repair, minimising explanatory gaps and substantive open questions. It must also find a proper place for the practice of apologising in particular. If, as I argued in the introductory chapter, we are particularly attached to this practice, why might that be? What is the aim of apology, and how does apologising promote or achieve it?

Apology as Penance

The first version of the view I will consider holds that the aim of repair is to restore proper balance in suffering or hardship between the parties. In wronging the victim, the thought goes, the offender has not only imposed material costs on him, but also physical, emotional, or psychological distress; she has deprived him of a degree of wellbeing which he would otherwise have enjoyed, and to which he is entitled. This has introduced an unfair disturbance of the well-being equilibrium that previously existed between them. While she may have gained some advantage or satisfaction as the result of her act, he has unjustly lost out. To set things right between them, the wrongdoer must now bear a correspondingly painful cost herself; and apology is the means for doing so. I call this iteration of the bookkeeping model a *penance* view.

The proposal shares a great deal with conceptions of retributivism in criminal justice, and with theological satisfaction theory. Both domains affirm the claim that a

[29] It may of course provide neither of these, showing repair to be a 'morally neutral' practice, which is neither morally virtuous or vicious in itself. For my purposes, what matters is avoiding the latter implication: the conclusion that there is something inherently morally troubling about the pursuit of repair.

wrongdoer must repay her transgressions (whether against the state or against God) through her own suffering. [30]

In considering this view, notice first that a mere disparity in wellbeing between the parties does not suffice for establishing a need for repair. If lighting strikes and ignites your house, but not mine, or if you lose your savings playing roulette, I do not owe you repair for being worse off. Your relative deprivation must be somehow or other unjust, and it must brough about by my misconduct.

Critical assessment of this view might start by asking whether the phenomena under consideration really fit the penance view's description of them. First, is the commission of wrongdoing really certain to generate an imbalance in wellbeing in the wrongdoer's favour? On the face of it, it would seem there are instances of A's wronging B which make A worse off, or which improve the situation of B, or which simply make no discernible difference to their wellbeing one way or the other. Take for instance, a case where a false belief causes me to think I am harming you when in fact I am not; in an example due to Davidson, I may trample on my hat, believing it to be that of my enemy(Davidson, 2001, p. 229). In a case from Bovens (2008, p. 222), a unscrupulous doctor means to kill her patient by administering a drug overdose, but instead cures the patient of a debilitating disease. In both of these cases, a malicious intent to injure backfires on the immoral actor.

In other cases, it seems there is simply no discernible difference made to the wellbeing of the parties one way or the other; for instance, in cases where moral luck foils a planned assassination plot. Provided one thinks, as many people do, that simply planning, intending, or attempting to injure another suffices for wronging him, it seems the penance view goes wrong in aligning wrongdoing with the generation of an imbalance in suffering in the wrongdoer's favour.

The penance view could treat all these cases as aberrant, fringe, or non-paradigmatic, but a better strategy might be to maintain that there is *always* and as a matter of definition a loss of wellbeing involved in being the victim of wrong – whether you

[30] See e.g. (Radzik, 2009b) on antecedents in Anselm.

know it and care about it or not, the wrongdoing has set back your interests, say, or failed to accord you the care and respect you are due. Even if you never consciously experience your victimisation as a loss – indeed, even if you somehow gain from the experience instead - you will all the same have lost in an objective sense.

This response is fine as far as it goes, but why, we would want to know, must this objective loss definitionally involved in victimisation always weigh more heavily than any subjective gain that may accompany it? And vice versa, for the wrongdoer. However, my critique of the penance view does not rest on this point, so let us grant for the sake of argument that wrongdoing does indeed correspond to a relative loss of welfare on the part of the victim.

Our next question is whether it is true that the practices we use to pursue repair, and in particular the practice of apology, really is painful for the one who undertakes it? If repair is to count as penance – as a way of imposing on oneself an onerous burden the weight of which corresponds to the victim's suffering – it seems it needs to be painful to repair, and by extension painful to apologise.

To make this case, proponents of a penance view point to the fact that apology characteristically involves assuming a self-humbling or even self-degrading posture, and that the paradigmatic language involves reports on pained or painful affective states:

> "[A]pology has two fundamental requirements: the offender has to be sorry and has to say so. These are the essential element of an authentic apology. Other features, for example, offers of reparation, self-castigation, shame, embarrassment, or promises to reform, may accompany an apology, but they are inessential because, I submit, they are implicit in the state of "being sorry"." (Tavuchis, 1991, p. 36)

All the same, we may question whether apology is *always* painful to undertake, or always painful *enough* to amount to proportional penance for the wrongful injury caused. Swinburne, questioning this, proposes that apology must then be backed up by additional 'costs':

> “Apology can often be very difficult, it costs many a person a lot to say ‘I'm sorry’. But sometimes for some people, apology can be very easy. We all know the smooth amiable people who say ‘I'm frightfully sorry’ with such a charming smile that our reaction is ‘Yes, but do you really mean it?’”(Swinburne, 1989)

Considering an example of an unreliable borrower, he writes:

> “You lend your friend £1,000. He forgets to return it, until you remind him five times; in consequence of which you have to borrow money yourself and disappoint your own creditors. He then acknowledges his wrongdoing and resolves not to do it again (publicly, and, let us suppose, also privately). He pays you the money back and compensates you for any interest payment and loss of time, and says that he's sorry. And yet that's still not quite good enough, is it? We feel something else is required.”(Swinburne, 1989)

The ‘something else’ turns out to be a ‘token of his sorrow’: a ‘costly gift’ that manifestly sets the repairer back in time, effort, or financial resources. Without taking on this burden, the repairer’s apology does not have the same significance: “The penitent constitutes his apology as serious by making it costly.” (Swinburne, Ibid.) For Swinburne, then, an apology that is given too easily is not as valuable and not as reparative as one that is burdensome to make, or one made more burdensome by being accompanied by a considerate and expensive gift or favour.

In defending what he at one point terms the ‘Penance Argument’(Bennett, 2012) agrees that “merely *saying* sorry” (Bennett, 2008, p. 118) is sometimes insufficient, and that other remedies may called for as supplements or alternatives to apology to make it adequately burdensome. In considering a teacher who has failed to live up to her responsibilities to her students, he writes that she

> “might do penance by undertaking unpaid remedial work for students who are in difficulty (particularly, though perhaps not exclusively, those who are in difficulty as a result of her negligence). She might also do some voluntary teaching outside of the university, say in schools or as an evening class.”(Bennett, 2008, p. 118)

The “important general conclusion” reached is as follows:

> “that the amount of penance that we expect someone to do in order to redeem herself (its duration and onerousness) is the way in which we express our sense of the seriousness of the action. The penance therefore has to be proportional to the offence.”(Bennett, 2008, p. 119)

Suppose we are satisfied that the penance view can account for both the idea that wrongdoing brings about an imbalance in wellbeing, and that it is possible for repair to be correspondingly burdensome for the wrongdoer – either because apology itself is onerous, or because, and insofar as, it is backed up with further acts of penance that add to the onerousness of repair. What would a practice requiring repair through penance have going for it? Is it normatively or explanatorily attractive?

Several different lines of critique, I argue, suggest that it is not. The first worry to consider is that even if we stipulate that the model can indeed achieve a kind of equality of welfare between the parties, it achieves this by *levelling down* – by depriving the wrongdoer of wellbeing, so she ends up being just as badly off as her victim. But finding something good in equality achieved through levelling down has struck many philosophers as highly implausible, or even morally indefensible.

What makes the view particularly liable to this line of objection is the insistence that repair must involve the proportional loss of welfare for the wrongdoer, rather than the proportional gain in enjoyment for the victim – a feature that is present in both Swinburne and Bennett’s presentations. It would not suffice, on these views, if the wrongdoer were simply to cheer the victim up, and particularly not if she were able to do so relatively cheaply[31]. To answer the objection to levelling down, the view must point to some good that comes from the wrongdoer doing penance by suffering under an onerous burden.

The most obvious candidates are not attractive, however; first, we might argue that the wrongdoer’s suffering is good for the victim. Indeed, as some retributivists have

[31] “We all know the smooth amiable people who say ‘I'm frightfully sorry’ with such a charming smile that our reaction is ‘Yes, but do you really mean it?’” (Swinburne, 1989)

pointed out, victims of wrongdoing sometimes report a desire for the suffering of their assailants[32]. However, at least as frequently, other theorists warn that this impulse is the sign of a bloodthirsty vengefulness that must not be indulged, let alone celebrated as morally righteous or morally necessary. (Radzik, 2009, p. 51).

Alternatively, a proponent of a penance view could argue that suffering is good for the wrongdoer. It is, as both Swinburne, Moore, and Bennett argue, what a guilty person *ought* to feel and *ought to want* to feel, provided at least that she were right-thinking. Suffering penance could allow one to expiate the guilty deed and provide the opportunity for moral renewal.

But there are two problems with this response. First, if apologising is only good for the wrongdoer, pursuit of it is liable to a kind of egoistical self-concern. Offering an apology, Radzik writes:

> "might be painful or humiliating to some degree but need not be. Furthermore, when an element of suffering is present in such actions, it is far from obvious that this is the source of their value as responses to punishment. The victim may have been hurt financially or physically as a result of one's misdeed and may continue to suffer from hurt pride, alienation, or self-doubt. In light of such problems, wrongdoers who express their sense of self-guilt through the single-minded pursuit of their own suffering appear shortsighted and self-absorbed." (Radzik, 2009b, p. 38)

Second, the notion that penitential suffering can itself be the source of a positive transformation is suspect. It looks like a slide into precisely the kind of resort to the magical or supernatural that we were hoping to evade. There is, in H.L.A. Hart's terms "a mysterious piece of moral alchemy in which the combination of the two evils of moral wickedness and suffering are transmuted into good."(Hart, 1968)

The same point applies if it is held that the wrongdoer's suffering is simply good in itself, or good for their relationship she has with the victim – a way of putting them on an even keel and restoring them to a balanced state. Martha Nussbaum describes

[32] See e.g (Moore, 1988)

the idea that the wrongdoer's penitential suffering could out-weigh or cancel out the victim's as "magical thinking" and "a restoration fantasy"(Nussbaum, 2016).

The problem that afflicts the penance view, then, is the same as afflicts retributivist conceptions of punishment in the legal and political philosophy: The claim that justice will be done through the wrongdoer's suffering requires further support. We need some additional story about why this kind of justice is valuable, and in particular, why it is valuable in the pursuit of repair.

But more than that, when we consider pursuits of repair that happen within the context of an interpersonal relationship, and particularly a relationship between intimates, the insistence on suffering only becomes harder to defend. Even defenders of retributivism about criminal punishment have argued that these demands are appropriately slackened between people who love and care for one another(Levy, 2014, p. 654). The view looks liable to this style of objection insofar as it exemplifies something akin to the eye-for-an-eye principle of the *lex talionis*. This kind of objection is particularly pertinent for those who allow that penance may be undertaken in ways that do not involved the injured party at all, such as Bennett's.

Added to these concerns about providing a respectable mechanism for repair, apology itself appears somewhat inessential or even potentially treacherous; if apologising comes all too easy for some, it is better substituted for hard physical labour, acts of service, or a costly present.

Apology as self-denigration

So far, I have introduced the basic contours of the bookkeeping model as an approach to the problem of repair. I have then considered in some detail a particular iteration of such a view – the penance view – according to which the imbalance to be addressed by repair is to the relative levels of suffering or hardship between the parties arising from wrongdoing. I've argued that this view confronts both explanatory and ethical challenges.

Most notably, like the retributivist positions it resembles, it struggles to explain the positive value of the offender's suffering. However, as an account of repair, I argued, the insistence on suffering is even harder to defend, for it does not suffice to make the case that the wrongdoer *deserves* to suffer, or that her suffering is inherently just. *Even if* we agreed that a wrongdoer's subsequent suffering is deserved and just – for instance, by taking this as a kind of moral bedrock – we are left wanting an explanation as to its contribution to mending the conflict between the two parties. Here it looks at best like a necessary first step, and arguably not even that.

However, we could learn from retributivism in developing the penance position to help us overcome this problem. Confronted with the challenge of justifying hard treatment, many retributivists about punishment do not rest their case simply on the (purported) intuitive obviousness of its justice, or its fit with the pattern of emotional reactions commonly experienced when one has been wronged. Instead, retributivists often seek to justify impositions of hard treatment on the basis that punishment has a distinctive expressive or performative significance that is needed to counteract and oppose the original wrongdoing.

The underlying idea here is that wrongdoing itself is in some sense communicative. A classic expression of this view is found in Murphy (Murphy and Hampton, 1988)[33] who writes that what we resent in wrongdoing is

> "not simply that [the wrongdoer] hurt us in some tangible or sensible way; it is because such injuries are also *messages* - symbolic communications. They are ways a wrongdoer has of saying to us, "I count but you do not", "I can use you for my purposes," or "I am here up high and you are there down below." Intentional wrongdoing *insults* us and attempts (sometimes successfully) to *degrade* us – and thus it involves a kind of injury that is not merely tangible and sensible. It is moral injury, and we care about such injuries."(Murphy and Hampton, 1988, p. 21)

The idea of such moral messaging has been very influential in the subsequent literatures on blame and forgiveness , and it helps to fill some of the explanatory gaps we encountered above: in addition to harming us in ways that are "tangible and

[33] See also (Ekins, 2012) for a survey of historical antecedents in the law.

sensible”(Murphy and Hampton, 1988, p. 25), wrongdoing also imposes a different kind of injury viz. by degrading or insulting its victim, treating him as inferior in status or importance. These kinds on injuries are present even in cases where the victim happens to gain materially from being wronged, and this helps us to make better sense of the idea central to the bookkeeping model that victims of wrong suffer a loss. The relevant loss (and corresponding gain for the wrongdoer), we now see, is not (just) to his material, physical, or psychological welfare, but instead to his symbolic standing as the wrongdoer’s equal. Accordingly, and in keeping with the organising principles of the bookkeeping model, it is this standing or status that repair must restore.

Furthermore, apology turns out to be an apt vehicle for carrying this countervailing message:

> “[O]ur moral relations provide for a ritual whereby the wrongdoer can symbolically bring himself low (or raise us up – I am not sure which metaphor best captures the point) – in other words, the humbling ritual of *apology*, the language of which is often that of *begging* for forgiveness. The posture of begging is not very exalted, of course, and thus some symbolic equality – necessary if forgiveness is to proceed consistently with self-respect – is now present.”(Murphy and Hampton, 1988, p. 28)

In a similar vein, Jean Hampton writes:

> “An apology is a way of humbling ourselves in front of the one whose value (and entitlements) we have failed to respect. Such humbling is not easy for us prideful wrongdoers, which is why apologies come hard. But by apologizing, we deny the diminishment of the victim, and our relative elevation, expressed by our wrongful action. And by trying to "make it up" to our victim, we attempt to repair the damage we have done by failing to respect their entitlements. If we are successful, our response annuls the appearance of degradation accomplished by our act, and establishes the right moral relationship between us.

> What these reflections show is that retribution is actually a form of compensation to the victim. Whereas tort damages are supposed to be awarded to place the victim in the situation she would have been in had the tortfeasor not acted, retribution is supposed to be inflicted to nullify the wrongdoer's message of superiority over the victim, thus placing the victim in the position she would have been in had the wrongdoer not acted" (Hampton, 1991, p. 1698)

Finally, Luc Bovens writes:

> "apologies are admissions that I did not treat you with the respect that is due to you. I bow my head to make up for the deficit of respect in my earlier treatment of you. Kant (1793, Part 1, ak 6:332) describes a case in which a rich offender must not only apologize, but also kiss the hand of the victim who is of lower social status. This display of humility expresses an excess of respect, and this excess is meant to put the scales of respect back into balance."(Bovens, 2008, p. 231)

Rather than a form of penitential suffering aimed at restoring a balance in wellbeing, on this version of the bookkeeping view apology is an act of status-restoring self-effacement or self-denigration; it is a means for the wrongdoer to humble herself, and thereby lower her standing relative to the victim she injured.

Notice also how closely both Hampton and Bovens follow the language and the logic of bookkeeping – debts and deficits in respect in must be redressed by compensatory offerings in kind, which, if adequate, will annul the damage done and thereby and 'put the scales back into balance', restoring the victim to the position he 'would have been in' were it not for the wrongful act.

Compared to the penance view discussed above, the self-effacement view gives us a more plausible account of why repair would matter to the wronged party. It is hard to see, I argued, how he could gain from the wrongdoer's subsequent suffering, but much easier to understand his having an interest in not being inferior to her. Indeed, it seems we should affirm this interest as entirely morally appropriate.

The view is not without its challenges, however. Proponents of the idea that wrongdoing involves a form of offensive moral messaging need to contend with the fact that some wrongdoers and some instances of wrongdoing intuitively make for a bad fit. For instance, Helmreich (2015) considers a case where a student breaks a promise by continuing to treat their teacher as an expert with superior knowledge and skills, rather than treating her as an intellectual peer as she requested; Martin (Martin, 2010) argues that many wrongdoers neither *tacitly believe* their victims have lower value than themselves nor *reflexively intend to communicate* any such attitude to their victims[34].

If the proposed account is to cover cases like the ones Helmreich and Martin have in mind, therefore, moral messaging must be understood in a fairly unrestrictive sense. It must involve a form of meaning a given act-type can carry and convey independently of, and even contrary to, the conscious attitudes of its agent. Some will find this idea implausible, but in what follows, I will accept it.

A more serious difficulty is whether it is right to see wrongdoing as adversely affecting the parties' standing as moral equals, and so to see apology as the means for redressing this balance in status. Does wrongdoing which carries an injurious message really *degrade* its victim, and/or relatively elevate the person who enacts it?

If it does, this would be enormously troubling, and oddly paradoxical; firstly, why would those who violate morality's rules thereby gain the power to mould moral reality to their will? Even those who believe equal moral status is not guaranteed and unconditional – surely a minority of moral philosophers – do not accept that a person can become morally inferior simply by being treated as such by another[35]. Human value, *even if hierarchical and conditional*, is not as easily malleable as this.

[34] Martin makes this point to identify the shortcomings of a view defended e.g. by Anderson and Pildes (Anderson and Pildes, 2000), who hold that an action can be expressive by 'manifesting' a state of mind, even if the agent does not intend to communicate it. The problem for this view, according to Martin, is that not all wrongdoers have such offensive states of mind to manifest in the first place – neither directly or tacitly. See also (Ekins, 2012)

[35] For instance, perhaps some would argue that one could lose equal moral status by losing rational capacity or executive function, or by entering a permanent vegetative state, or perhaps by violating basic moral requirements. I return to the latter suggestion below.

The point is therefore not that status cannot be gained and lost when one person is defeated or overpowered by another; it is rather that *this kind of status* cannot. For comparison, think of a pair of prize fighters, one an established champion, and the other a fresh-faced challenger. If the challenger defeats the champion in the ring, this reverses the relative hierarchy of positions between them; the challenger is now ranked more highly, and the former champion as relatively inferior. Hierarchical and flexible status ranking systems like these clearly exist, but even where they do, there are very precisely circumscribed mechanisms for moving 'up' and 'down'; if the challenger treats the champion as inferior by chiding and ridiculing him prior to their fight, or if he simply asserts his own superiority, this does not suffice to make it so. If the challenger is ahead on points but the match is interrupted before the completion of the required number of rounds, then the challenger does not gain superior status. If the challenger overpowers his opponent, but only by breaking the rules of the sport, he does not gain superiority.

Furthermore, there is something strangely self-fulfilling in the idea that being treated as lowly can make it so; if the transgressor literally decreases the value of the person she treats with disrespect, then it seems she thereby *makes it false that she wrongs him*. By failing to act in accordance with his (prior) value, she lowers that value, retrofitting it to her treatment of him. But if she successfully makes him lowly, she also undermines any claim he could have to being treated better, and thus any scope for repair as we understand it.

In sum, the malleability of status implied by this conception of repair faces three problems: First, it requires an inegalitarian conception of human value, according to which people can be differentially morally valuable, and change in value over time; but this is a view most would reject. Second, even if we adopt a hierarchical and conditional conception of human value, we have good reason to reject a view on which a person's moral value can decrease simply by being wronged by another. Such a view is both normatively objectionable on its own terms (affording moral transgressors the power to decide on moral value assignments), and self-fulfilling in a way that undermines the coherence of repair; if the wrongdoer *does* remake moral values, she thereby makes her action morally acceptable, and makes apology and repair inapt.

Hampton takes note of the first of these malleability problems and attempts to address it by amending the view such that wrongdoing does not in fact lower the victim's value and increase the wrongdoer's, but merely *appears* to do so. Instead of remaking moral values to suit her wrongful purposes, then, the wrongdoer's behaviour *betrays* and *misrepresents* them. Hampton calls this 'diminishment'(Hampton, 1991, p. 1673)[36].

Diminishment, Hampton argues, can injure a victim by *damaging the realization* of his value, or by *damaging the acknowledgement* of his value (Hampton, 1991, p. 1678); it does the former when it treats the victim in a way that it violative of his entitlements. For instance, in my earlier example A violates B's ownership rights over his bicycle by removing it and using it without B's consent. A wrong damages the acknowledgement of the victim's value when it "threatens to reinforce belief in the wrong theory of value by the community" (ibid): if A treats B's property rights as casually violable, she sends the message, not just to A, but also to their wider community that others can similarly treat A's belongings as they please.

This version of the view is an improvement in that it is consistent with more stable and egalitarian conceptions of human value, and allows us to retain the idea that the transgression is wrongful rather than self-fulfillingly appropriate. However, the wrongdoer's behaviour is now wrongful in a rather different way: it is simply incorrect.

Now, false claims about moral value can clearly be damaging. For instance, mistaken views about the moral value of people of different genders, races, or physical abilities have been enormously harmful for the individuals in those groups. But is a wrongdoer's false claim of superiority relevantly like this? The false claim that *women have a lower moral value than men*, say, is pernicious because and insofar as it is widely

[36] Gert and colleagues (Hampton, 1991, p. 1678) raise some doubts about whether this professed position is really a faithful representation of Hampton's view. Hampton certainly makes a number of claims that are difficult to square with the notion that diminishment is a mere appearance of lowered status. However, whether Hampton holds the mere appearance-position or not, it is surely an option worth considering.

accepted and engrained in our structures, institutions, and social practices. But a singular false one-off claim about relative value ('I am more important than him' ; 'my desire for a bike ride outweighs A's claims to his property) does not become injurious in these same ways simply by being expressed by one person to another. At a minimum, the wrongdoer's claim would have to persuade, or stand some chance of persuading, a preponderance of others to adopt her false view. The idea that it could do so is made more tenuous still when we recall that the false view of relative values the wrongdoer's action expresses is not 'hers' in the usual sense; the wrongdoer may consciously reject it and earnestly disavow it, yet have expressed it in her action all the same.

Hampton insists that diminishment is an 'objective' matter ((Hampton, 1991, p. 1683), not something that depends on anyone's actually being convinced by the implicit false message or beginning to lose confidence in the victim's equal value. The problem in moving to this construal of wrongdoing and its relevance to status is therefore to explain why a false claim expressed by the wrongdoer's action needs to be ceremoniously gain-said through a self-humbling ritual, rather than simply being laughed off or ignored. If it is agreed that the wrongdoing does not in fact lower the victim, but merely misrepresents him as lowly (perhaps unconsciously and unintentionally), why is opposing it so important? And why would it need to be opposed in this specific way?

What gives bite to this concern is that ritual self-denigration itself looks somewhat morally troubling and in need of positive defence. The suspicion is liable to arise that what we have here is just another mode of penance payment, somehow supposed to cancel one evil out with another of the same kind and proportion[37]. But in addition, if the problematic impact of wrongdoing is a false representation of one party as having

[37] (Gert et al., 2004) make a proposal along these lines about Hampton, proposing that her position is affected by a stable underlying commitment to retributivism. Bennett (Bennett, 2022) could arguably also be read in this light, albeit with the central difference that it is the wrongdoer's equal status, and not the victim's, that is to be earned back through the payment of penance. This arguably makes Bennett's position a kind of hybrid between what I label Penance and Self-denigration views. Martin (Martin, 2010) also considers the possibility that it is wrongdoer's status, and not the victim's, that is at stake.

a lowly status, why would a further performance of defence and inferiority by the other be the right way to counter it?[38] If the goal is to affirm the parties' equality, how would it help to performatively *deny* equality for a second time?(Radzik, 2009b, pp. 43–44)

Apology as Reassurance

So far, I have considered the bookkeeping model of repair as exemplified by two different restitutive currencies, wellbeing and status, and argued that neither variant, viz the penance view and the self-denigration view, gives us a compelling conception of the underlying phenomena. Amongst their various other problems, the penance model struggles to demonstrate the reparative value of the wrongdoer's suffering, and the self-denigration model does better only by adopting a normatively troubling conception of the inequality and malleability of moral worth.

Confronted with these problems, a persistent proponent of the bookkeeping model's central commitments will cast about for a different placeholder value for reparative restitution – a different kind of loss that might be counteracted by the wrongdoer's apology. In doing so, she may consider what can be gleaned from the evaluations conducted so far.

One lesson she might draw is that there is something irredeemably problematic about the implicit retributivism of the earlier views. Both the penance view and the self-denigration view focus their attention on the discomforts that befall a victim of wrongdoing and look for ways of matching these – suffering for suffering, or humiliation for humiliation. As the vehicle for repair, the wrongdoer's apology was made to serve as a proportionately unpleasant penalty for what *was*, rather than an opportunity for constructively looking ahead and rebuilding what could be. As such, though each view tries to give additional purpose and structure to the pursuit, we kept coming up against the suspicion that their characterisations of the drive to repair remain uncomfortably close to vengefulness: a conception of justice that is both

[38] The worry that the self-denigrating character of apology is itself morally troubling is raised amongst others by (Garrard and McNaughton, 2003) and (Nussbaum, 2016) and it is natural to think that it has to some extent motivated proponents of unconditional or gifted forgiveness to set demands for repair aside altogether.

morally unsavoury on its own terms and fundamentally unhelpful for enabling the parties to move on.

But perhaps there is a different way of conceiving of restitution – one that does not rely on magically extracting a *good* from an equality of *ills*, but instead focusses on what the victim needs to put the matter behind him and securely renew bonds with the wrongdoer. If we strip away the retributive drive, what the victim really needs, we might think – what he *should* need and *would* need, if he were not in the grip of vengefulness or irrationality - is not retribution, but reassurance; he needs to know that the wrongdoer's transgression will not be repeated, and that she now understands, appreciates, and intends to respect his interests; that he will be safe if he decides to make himself vulnerable to her once again.

The idea here is that we can retain the central elements of bookkeeping, but purify it of those backwards-looking parts that proved morally troublesome and explanatorily vexing. The equilibrium to be restored is not measured in wellbeing, or in status and standing, but in those beliefs and expectations the parties have of one another, and which sustain and facilitate their social cooperation. The victim was unduly deprived of those beliefs and expectations when he was wronged, and the object of repair is for the wrongdoer to reinstate them by reassuring him that they are now once again justifiably held. As before, however, the standard for repair is set by the antecedent state of the relationship – in this case, its epistemic or doxastic profile.

The reassurance view, as I shall call it, even gives us a way of recovering elements of the earlier accounts that seemed attractive: when victims desire the wrongdoer's humble self-lowering, or when wrongdoers feel compelled to endure something painful, what is valuable in these acts is not that they establish symbolic equality of status or suffering, but that they are ways of expressing the wrongdoer's reformed outlook. As Swinburne proposes, the wrongdoer constitutes her apology as serious by making it painful(Swinburne, 1989); but, we can now say, it is its seriousness, and not its painfulness that is reparative. If she can achieve the former without the latter, so much the better.

Similarly Bennett writes: "It is the fact that the wrongdoer fully understands that what they have done is wrong that achieves reconciliation and the restoration of relationships. It is this that allows them to be reaccepted as a member of the relationship from which their action removed them." (Bennett, 2008, p. 112)

And Morris states:

> "The satisfaction that one obtains in the self-inflicted or accepted pain here comes from the very character of the conduct as painful, for it is this that evidences how much what has been done counts for one and how much it means for one to restore. When doubt has been raised about care and commitment, assumption of pain is a preeminent mark of their presence".(Morris, 1971, p. 431)

Or Radzik:

> "Given certain circumstances, self-imposed suffering is a sign that one cares about the victim's pain and thus it is evidence that one is no longer discounting the victim's point of view as less important that one's own. These are precisely the messages that reconciliation requires. Thus self-punishment might be justified as a means to communication." (Radzik, 2009c, p. 101)

As Radzik notes, however, this justification for self-punishment is tentative; though self-inflicted suffering *might* express commitment to the right values, it could also communicate a variety of other ideas: self-hatred or narcissistic self-concern, for instance. Our real interest is in the wrongdoer's present commitment to the right values, and so in knowing that we can safely rely on her in the future. As such, self-imposed suffering or self-abasement are dispensable means to our ultimate end.

As a staunch critic of the 'retributive drive' and its purported connection to forgiveness, Martha Nussbaum(Nussbaum, 2016) takes a similar view. Though apology must not be imposed to settled scores in suffering or status, it "can certainly be useful evidence (as with children) that the wrongfulness of the behaviour is understood" (Ibid. 124), or serve as a "useful sign of what we can expect of the offender in the future" (Ibid. 141). Apology may lead us to "expect good things from the apologizer, all else equal," and thereby remove the victim's need for further protest and anger (Ibid. 154).

Again, the transition to a different version of the bookkeeping model mirrors an approach to punishment in legal philosophy. Some expressive conceptions of punishment hold that hard treatment is inessential, because what really matters is that the victim (and the wider community) can feel confident that the wrongdoer has now understood and adopted the correct moral norms and will not violate them again. Subjection to something unpleasant (monetary fines, community service, deprivations of liberty) *may* be a step in this process, since it may be what leads the wrongdoer to reform her position, e.g. by focussing her attention on the wrongful deed (Duff, 2006). However, it needs not be – if there is another way to instil the right values, or if the wrongdoer already has adopted them, further penalties are unnecessary, and so inappropriate. The issue then becomes one of deciding under what circumstances the victim can feel assured.

The principal challenge for the reassurance view is that it struggles to supply the right connection between reparative function and reparative practice: if the objective of repair is to reinstate the victim's confidence in the wrongdoer's future behaviour, then this is something a great many things could supply. Would it suffice for repair to learn that the wrongdoer has undergone a moral conversion, for instance? If so, this is something one could learn from overhearing her, or from the testimony of a trusted third party, or even from a psychological assessment.

The trouble with this proposal, then, is that it transforms the task of repair in way that threatens to do away with it altogether. If what the victim is owed is secure reliance on the wrongdoer, or reasons to renew trust, why must the wrongdoer herself play any part whatsoever in supplying it? If apology is 'useful' only as a 'sign of what to expect for the future', the communicative practice and the notion of repair as a task for the wrongdoer is itself entirely superfluous.

It is worth spelling out more clearly why this would be a surprising conclusion. The point is not simply that the particular formulations paradigmatically involved in apologising ('I'm so sorry', 'I apologise', 'Please forgive me', etc) are inessential; that much is not controversial, insofar as it is often possible to deliver repair without using such stock phrases. The point is rather that the wrongdoer's active role in repair *as such* is undermined. She does not need to say or do anything at all for the relevant function (restoring credal states) to be served.

Of course, we could *stipulate* that not just any odd way of restoring credal states counts as repair; it has to be done *by* the wrongdoer and *to* the victim and *by means of* a performative utterance of such and such a kind. But without further argument, these stipulations would be ad hoc. *Why* must the relevant reparative function be served in the specified way?

Richard Moran argues that there is an important difference between *giving* an expression and merely *giving off* an expression. A person's facial expression, behaviour, or tone of voice, say, can express that she feels remorse for her behaviour without her explicitly or intentionally communicating any such thing. Aspects of observable appearance can serve as good evidence of her internal states. However, when she freely declares her remorse, openly and intentionally communicating it, she does something else:

> "When the person *expresses herself*… she doesn't simply provide a window onto her state of mind, but also "owns up" to the attitude in question, acknowledges it, and assumes a certain kind of responsibility for it, and for the hearer's knowledge of it. None of this is part of the story when her remorse or gratitude simply manifests itself, clear as day, in how she looks or what she does."(Moran, 2018, p. 86)

Her freely chosen words, then, achieve something altogether different than providing evidence of the inner state they express: by means of them she *owns up* to the attitude and *takes responsibility for* the victim's knowledge of it.

Both Jeffrey Helmreich and Adrienne Martin each adopt similar positions. Martin writes:

> "there is a performative element of apology that goes beyond demonstrating that resentment's expectation has been satisfied—namely, a pseudo-contract

> with the recipient; a second-personal, remorseful taking of ownership". (Martin, 2010, p. 547)

Both Martin and Moran, then, find that the performative element in apology grants the victim a kind of contractual guarantee or IOU. She does not merely lead him to believe that she will do better, but assures him that she will, thereby affording him a commitment he can hold her to. When she says "it won't happen again", she is not making a prediction, but a promise.

On the face of it, this should help address the problem just noted. A victim who merely has a (highly reliable) third party report or behavioural prediction does not have the wrongdoer's explicit ownership or commitment. However, under closer scrutiny, the epistemic contribution of this pseudo-contractual guarantee is hard to comprehend: is the speaker's open declaration of commitment supposed to give the victim *more* and *better* reasons for relying on the wrongdoer than he would have had without it? The answer, it seems, would need to be a yes, but on what grounds?

Moran's idea seems to be that the taking of responsibility provides a distinctive reason for belief because it grants the hearer a right of complaint: by making a free assertion, the speaker stakes her reputation on the truth of what she says, and entitles the hearer to criticise her if it turns out to be false. This is also Helmreich's position; if the wrongdoer's normative orientation changes, but goes unannounced, she will be in a position to abandon it again "without anyone in particular having a claim to [her] commitment"(Helmreich, 2015, p. 104). By contrast:

> "If I openly and explicitly apologize to you, on the other hand, you *do* have that claim. Then you may say, "But you apologized." Speech acts, in other words, commit the speaker to certain listeners that she will remain in some way consistent with having performed them. It is possibly for this reason that we sometimes talk of apologies as not only offered but "given": the apology is in some way no longer the wrongdoer's to realize with her subsequent behavior; the victim-listener now has a right to hold her to it".(Helmreich, 2015, p. 104)

I will have much more to say about this idea as a conception of testimony in general in the next two chapters. However, for present purposes, I shall focus on its application to the phenomenon of apology. In this context, the notion that the normative structure just described could *strengthen* the reliability of a wrongdoer's present commitment to good behaviour, or otherwise afford reasons for belief that enhance the victim's epistemic position, should already strike us as odd: firstly, why would the victim put much stock in the word of the wrongdoer at all? She has just shown that she was both willing and able to transgress against him, so ought he not to be somewhat sceptical and require a bit more than her bare assertion to renew his reliance?

Secondly, and more importantly, if the wrongdoer does indeed go back on her commitment and wrongs the victim again in the future, it is not the fact that *she promised otherwise* that gives him a right to complain. He already had such a right, whether she promised or not.

This marks a relevant difference between testimony and apology that breaks the analogy these authors are trying to draw. A person can lead you to adopt a given belief (e.g. that she is going on a trip) through observable behaviour (e.g. hauling out suitcases and packing her car in your view), or by testifying to its truth. If your belief turns out to be false, you will have a right of complaint in the latter case, but not in the former. If you merely observed trip-going behaviour and drew the natural inferences, she may have tricked you, and possibly even intentionally so, but she has not assumed responsibility for your belief in a way that makes her acccountable for its being false. By contrast, if she told you she would be going, but did not, and never had any intention of doing so, she has failed to live up to her word in a blameworthy way. She has violated her commitment and made herself liable to critique.

By comparison, a person may lead you to adopt the belief that she will not wrong you again through observable behaviour (she looks mortified at having done it once, you see her diligently attend anger management classes, and display a reformed outlook in interactions with other people, etc) or by promising you that she will not. But if indeed she does do it again, you manifestly have a right to complaint in either case.

The basis of your complaint is that she wronged you, *not* that she had said she would not.

Has she perhaps wronged you in a distinctive way? She is now blameworthy for being *both* a hothead *and* a liar? This does not seem the right construal; it seems perfectly possible that she was sincere when she committed to better behaviour, even if it turns out her self-confidence was misplaced. On most conceptions of lying, her earlier statement would not qualify. We *could* blame her for her overconfidence, or her lack of self-knowledge, but do we really want to classify these, too, as morally noteworthy failings? If so, repentant wrongdoers would be better off aiming to under-promise and overdeliver by never making direct commitments to betterment at all.

Affording a distinctive role to the performative character of apology was supposed to put the wrongdoer's active agency back into picture, but so far, the search seems to hit a dead-end. If the goal of repair is epistemic restitution for the victim, then actively apologising looks like a superfluous ritual: no more than a distracting bit of ceremony we might as well cut out or forego.

I said from the beginning that my project in this dissertation is to provide a vindication of the practices of repair as we know them. As such, arriving at a revisionist position is already a significant cost. However, kinds and degrees of revision can vary; perhaps this one can be welcomed without too much reservation? Should we accept the reassurance view, and therefore accept that providing doxastic restitution by whatever means suffices for repair?

Again, I think our hopes for the reassurance view must be disappointed. If the view offers the right conception of the aim and function of apology, viz restoring the victim's confidence in the wrongdoer's future behaviour, then the work of repair can happen entirely without the wrongdoer's involvement. True, she is *responsible* for its coming to pass, but she does not actually need to do anything at all. A conception of repair along these lines does not just revise it, but threatens to shrink it to an extensionless point.

Conclusion to Part I

In these first two chapters, I have first outlined the problem of repair as setting a practical task for the wrongdoer, and then argued that a particular approach to the problem –what I have labelled the bookkeeping model- has become prevalent in the philosophical literature. Despite significant variation in the ways versions of the view are formulated and motivated, a few key elements recur: Most centrally, a conception of repair as restitution, and so of apology as the wrongdoer's means for repaying or compensating the victim for the unjust deprivation of something valuable to which he was, and remains, entitled. Repair is successful, on this conception, when the wrongdoer has restored the parties' respective holdings to their *status quo ante*, or approximated that level as closely as practically possible.

I have then given closer examination to three specific variations of a bookkeeping model: the penance view, on which restitution is to be paid in the wrongdoer's penitential suffering or hardship; the self-denigration model, on which restitution is sought in relative status or standing through the wrongdoer's humble self-lowering; and the reassurance model, on which the target of repair is restitution in the victim's beliefs and expectations and a provision of reasons for future trust and reliance.

Each of these versions of a bookkeeping view, I argued, confronts significant ethical and explanatory challenges that are typically overlooked or underestimated. Consequently, each in its own way threatens to render the pursuit of repair as we know it irrational, incoherent, or morally ugly. Summarising our findings in very broad terms, I argued that versions of the bookkeeping model encounter one of two difficulties: it either reduces repair to a form of retributive punishment, or shrinks it away altogether. In the former case, we face the difficulty of explaining the restorative function of the (self-)imposition of a penalty or sanction on the wrongdoer, and of reconciling the insistence on punishment with common ideals for loving relationships between intimates. In the latter, where the goal of repair is to redress the damaged expectations and bases for future reliance between the parties, we encounter the difficulty that restoring the victim to his former position does not seem to require anything in particular on the part of the wrongdoer herself.

A proponent of the bookkeeping model, or of any of the specific iterations of it I have considered, could take up the task of addressing these challenges, or adjusting their views to avoid them, or setting out a different version of the view in another moral 'currency'[39]. However, in the final section of this chapter, I want to raise some more general concerns about the bookkeeping framework as such to suggest that this would not be the right way to proceed.

In offering my tentative defence of the bookkeeping model at the beginning of this chapter, I argued that it is not guilty of a crude reduction of all moral values to monetary payouts. It can respect and accommodate the notion that multiple incommensurate, non-fungible moral and interpersonal goods may be at stake in a conflict arising from wrongdoing. However, the model still approaches these goods in a fundamentally transactional manner – they are goods that are owned, owed, and repaid; distribuenda for distributive justice. Even if we think this distributive approach is the correct one in the social, political, or institutional domain (which, of course, is by no means universally agreed)[40], we should still be concerned that it distorts the complex realities of the interpersonal realm.

First, the bookkeeping framework as such faces additional explanatory challenges over the connection between repair and forgiveness. I mentioned in the introductory chapter that we sometimes find that a victim's unilateral forgiveness takes the place of repair. Forgiveness, it seems, can precede and motivate repair (Fricker, 2018), or even arise entirely in its absence and offer an alternative route to the parties' reconciliation (Calhoun, 1992; Garrard and McNaughton, 2003; Nussbaum, 2016). Though some believe forgiveness is always in some respect defective if it occurs

[39] Of course, it is also possible that none of the philosophers I have identified as subscribing to a bookkeeping conception of repair would accept that characterisation; the very notion that repair is the repayment of a debt, I suggested, has sometimes enjoyed a bad reputation, and it is therefore not unlikely that some would baulk at the comparison. Naturally, the interpretations I offer here are not indisputable; but it bears noting that it is hard to replace the function the bookkeeping idea serves in providing a unifying explanatory mechanism for repair. When participants in the philosophical debate point to *this* or *that* role which apology purportedly fulfils or accomplishes (or, as is often the case, provide an extended list of such roles), the following question remains: how does doing those very things offer repair? The bookkeeping model's idea of restitution promises to fill this gap without resort to mystery or magic.

[40] See esp. (Anderson, 1999) and (Young, 1990)

without antecedent repair, most everyone agrees that so-called 'gifted' forgiveness – forgiveness unearned by the recipient – is possible and sometimes morally acceptable.

The bookkeeping model initially looks well-placed to capture and make sense of this. When you conceive of wrongdoing as the generation of debts, it becomes natural to think of forgiveness *as* release from obligations to repay, and thus as an intelligible alternative to such repayment; indeed perhaps the surface similarity between the moral notion of forgiveness and the financial one has itself lent intuitive support to the bookkeeping model and led some authors to adopt its ideas.

Moreover, we see this move in some of the authors I have described as proponents of a bookkeeping model. In reckoning with the fact that complete restitution is not (or not always) practically possible, some authors explicitly propose that the sequence must conclude with the victim forgiving (the remainder of) the debt. Such a position is found most clearly in Helmreich, who holds that apologies are a form of partial repayment, presented and offered *as* insufficient (Helmreich, 2015, p. 94). The debt itself is "unpayable", like the destruction of a priceless family heirloom, and accordingly, "apologies themselves do not relieve an offender's moral debt to the victim", but instead "solicit the victim's own act of relieving that debt."(Ibid. 102)

> "One need only find an action that presents oneself to the victim as having wronged him, and having no way to make up for it, but seeking his acceptance of what one has done and will do anyway, or seeking his forgiveness of the outstanding moral debt. That explains why the plea, "Will you please forgive me for the wrong I've done, though I don't deserve it?" can be a workable substitute for apology."(Helmreich, 2015, p. 95)

However, on closer inspection, the inducement to forgive without complete restitution starts to look morally suspect within the bookkeeping model's framework; the victim is being asked to waive his entitlements to more (entitlements stipulated to be fair and just), and to do so just *on the grounds that* the wrongdoer has provided inadequate compensation.

The comparison Helmreich draws with monetary debt forgiveness serves his own purpose badly. If I have destroyed your valuable heirloom, and say "Here's a fiver. I know it is not enough, but please scratch the rest anyway", it would not be amiss for you to feel that I had added insult rather than redressed injury. That a debtor cannot pay and openly acknowledges this is in itself no reason at all to forgive their debt.

There could be other reasons to do so, of course; that the sum is unimportant, for instance, or to display one's generosity, or for old times' sake. But on the bookkeeping model, forgiving the debt on these grounds is to act in contravention of what justice rightfully requires. In short, the worry is that the bookkeeping model, rather than elucidating the connections between forgiveness and repair ultimately sets up an unattractive conflict that forces one to choose between mercy and justice. To forego repair, in its terms, is always to choose justice delayed or justice denied. If harmonious relationships require allocative balance and each party's respect for the holdings of the other, then unilateral forgiveness itself starts to look necessarily suspect rather than commendable – weak condonation, rather than righteous justice.

Another concern attaches to the bookkeeping mindset as such. The suspicion is that the bookkeeping mindset is fastidious and antagonistic in morally damaging ways. Not only does it betray normatively unsavoury attitudes (the desire to make the other pay or grovel for her sins), it is also liable to encourage protective withdrawal from the other and suspicion of their motivations and construal of the case. Above all, it risks entrenching a possessive individualism that paints other people as threats to one's projects and holdings, and promotes the notion that we must be worthy and deserving to stand in relationships with others. It makes mutual independence and self-sufficiency our goals, and leaves little scope for appreciating the value we attach to sharing a life.

Jeffrie Murphy later in his career reflected on the influence of what he describes as his "resentful and vindictive" nature and its "tendencies toward self-righteousness" on his own philosophical work and inclination:

> "At most and at worst, these factors may have inclined me to favor philosophical accounts for less than fully honorable reasons, especially those

> that may reveal that my enthusiasm for settling scores and restoring balance through retributive justice may in part have been extensions of what Nietzsche called "a soul that squints" — the soul of a shopkeeper or an accountant. If I had been a kinder person, a less angry person, a person of more generous spirit and greatness of soul, would robust retributivism have charmed me to the degree that it at one time did? I suspect not."(Murphy, 2007)

If these charges against the bookkeeping attitude stick, they affect not only the normative standing of the proposal in isolation, but also its likelihood of success. Viewing other people as enemies to one's interests and possessions and fearing that missteps will incur harsh punishments and condemnation is itself likely to be inimical to the attainment of repair[41].

Finally, Charles Griswold argues that the problem of moral repair is antithetical to a perfectionist moral outlook. If we were all perfectly good – perfectly kind, perfectly self-controlled, perfectly wise, etc. – there would be no wrongdoing, and so no need for repair. The practical and philosophical problem of moral repair arises because we are fallible – in one way or another, or, more plausibly, in all of them at once.

For Griswold, this is part of what recommends serious reckoning with the problem of repair in the first place. If we believe in the value of this insight – that repair compels us to acknowledge and accept our flaws and find a place for them in our moral theories – then we should be suspicious of the bookkeeping model, for it allows perfectionism to slip back in through the backdoor. The bookkeeping model promises that full repair can be achieved – a deficit once created that can be met, a debt that can be repaid, and which will then disappear for good.

Martha Nussbaum, who particularly connects the bookkeeping conception of repair with a religious Christian tradition, makes a similar observation in writing that transactional forgiveness is "perfectionistic and intolerant in its own way"(Nussbaum, 2016, p. 89):

[41] On this point, see also (Walker, 2014, 2013, 2006a, 2006b)

> "The list-keeping mentality that it engenders is tyrannical toward human frailty, designedly so. We must constantly scrutinise humanity, and frequently punish it. … in its exacting control over wayward desires and thoughts the transactional strand of the Christian tradition is highly continuous with (and influenced by) the very Stoicism Griswold criticises. Stoic philosopher Epictetus' instruction, "Watch over yourself as if an enemy is lying in wait," could easily have been said by many a Christian thinker."(Nussbaum, 2016, p. 89)

The desire to settle up once and for all in the face of wrongdoing is an undeniably compelling idea, particularly from the perspective of the wrongdoer; having hurt or harmed another can be a devastating burden to carry. Who has not had the experience of desperately wishing that what you did or said could be taken back or somehow made right again? Who has not hoped upon hope that if only you try hard enough, and earnestly apply yourself to the task, it is always within your power to redress the damage done, and clear your mistakes from the moral record? Who has not wanted to wipe the slate clean and start fresh? Unfortunately, however, as is so often the case in human affairs, things are not so simple.

In his seminal critique of views that strive to immunise the moral agent from the vicissitudes of fortune and unchosen circumstance, Bernard Williams remarks that "[s]uch a conception has an ultimate form of justice at its heart, and that is its allure … it offers inducement, solace to a sense of the world's unfairness" (Williams 1976, p. 116).

A similar allure, I propose, a similar fantasy of self-control and self-discipline as a shield from the taint of moral evil, attaches to the bookkeeping model's conception of the possibility of repair –repair, it promises, is always attainable, if only we try hard enough. The fantasy of control is placed at only one further remove – it applies not to the assessments others can make of us in light of our characters, our actions, or their consequences, but instead the assessments we merit given how we attempt to redress failures and missteps we (inevitably!) commit at this earlier stage. The bookkeeping conception of repair promises that although we cannot have the kind of control that is first best (one that prevents us from ever injuring others), perhaps we can have second best instead: our mistakes can always be corrected and put to rest.

But this is a fantasy that we would do better to give up. As I expand on the alternative picture I propose, I hope also to show that the costs of doing so are themselves worth bearing.

Part II

Chapter Three:

Telling takes two

> You can sail on a ship by yourself
> Take a nap or a nip by yourself
> You can get into debt on your own
> There's a lot of things that you can do alone
> But it takes two to tango
> "Takes Two to Tango", Al Hoffman and Dick Manning, 1952

In the preceding chapter, I examined several different variants of what I labelled a bookkeeping approach to moral repair and apology's role within it. I argued that these views struggle to deliver accounts capable of vindicating both the appearance of these practices as we know them, and the functional role they play in addressing wrongdoing. Summarising my findings somewhat crudely, the views that had a meaningful story to tell about what apologising *must be like* (painful; self-humbling) could not satisfactorily explain *what good it does* for the parties; and the views that gave us a compelling story about what good apology does for the parties (reaffirming moral norms or restoring a basis for trust) could not satisfactorily explain why this role must be served by the practice of apologising in particular, or indeed by anything resembling repair as we know it . That is to say, neither of the views examined gave us an adequate account of both the *content* and the *form* of moral repair and apology's contribution to it.

In the background we also saw a somewhat ambivalent position on repair's communicative character. On one hand, it is uncontroversial that repair's characteristic manifestation – apology – is itself a communicative act. But on the other, several authors emphasise that apology cannot be *merely* matter of the wrongdoer communicating something about herself or her attitudes. For instance, Bennett writes that apology's distinctive function must be to "act on the normative situation directly, and not simply [to serve as] as a source of evidence of the wrongdoer's state of mind"(Bennett, 2022, p. 120). Similarly, Helmreich emphasises that it is the wrongdoer's "behavior, not the attitudes putatively betrayed by it, that constitutes the mistreatment which apologies redress"(Helmreich, 2015, p. 92), and consequently

> "nothing like an expressed "change of heart," or repudiation, could remedy it. What is needed, instead, is a different, less purely communicative account of how one can remedy the mistreatment involved" (Helmreich, 2015, p. 92)

From both Bennett and Helmreich, then, one gets the impression that repair must do *more* than merely communicate, more than revealing or expressing what the wrongdoer thinks and feels; instead, it must somehow or other *act on the normative situation directly.*

Ultimately, however, neither of the bookkeeping views examined in chapter two actually gave linguistic communication, and so the paradigmatic forms of speech usually associated with repair, a necessary role at all. Both the penance view and the self-denigration view allow that the wrongdoer's restitutive acts could find expression in a variety of non-linguistic gestures (service work, gift-giving, or perhaps even self-flagellation on the penance view; humbly bowing down in supplication on the self-denigration view). And while some versions of the reassurance view try to give an essential role to the apologiser's commissive speech act, we saw that this argument is not convincing, and that we are ultimately given no adequate grounds for requiring active participation from the wrongdoer at all, let alone requiring her communicative action in particular.

In short, in the rush to avoid a conception of apology that is 'purely communicative', and instead seek a way to 'act on the normative situation directly', bookkeeping approaches to repair seem to have sidelined the role of communicative practice in repair altogether.

In the chapters that follow, I will argue for an altogether different approach. I shall argue that the wrongdoer's active involvement is indeed essential to repair, as is the communicative character of their contribution. It is no accident that repair proceeds by means of apology in ordinary practice, and a satisfactory philosophical account should both explain this and put its communicative character to relevant use. The account of repair I set out in Chapter Five does exactly this.

But to arrive at this point, I first need to gather resources for the argument. Doing so will require a temporary departure from the problem of moral repair and an examination of both philosophical and non-philosophical perspectives and insights on the character of communication.

In this chapter, I take up a research programme in the epistemology of testimony – already introduced in brief in the Chapter Two - which has given a central role to connection between the *form* or *agential character* communication and its *content* or *epistemic significance.* In particular, I will take up an proposal due to Richard Moran, according to which *telling takes two*. Though Moran's contribution is building on a pre-existing body of scholarship, the account he develops is amongst the most extensive, original, and detailed treatments of this idea. But it also proves unexpectedly pertinent to several of my principal interests in this dissertation – not least because the account of testimony Moran develops has a transactional streak that is, I shall argue, tellingly similar to the accounts of repair just considered in Chapter Two.

Accordingly, while I will align myself with Moran's ambition to explain testimony's epistemic contribution *by* elucidating its agential and intersubjective character, I will also argue that Moran's position misrepresents this character, and thereby misrepresents the way telling is a social phenomenon. As I shall eventually show, communicating something to an addressee requires either less or more than Moran argues; less because it is under some circumstances possible to testimonially transmit some bit of information despite the unwillingness of one's addressee; and more because the vastly larger share of our telling is far more radically cooperative than Moran (and with him most philosophers of testimony) tend to acknowledge.

The case for the latter of these claims will be laid out in Chapter Four, where I make a detour through areas of study in linguistics and sociology that have investigated conversation empirically, and particularly a field known as conversation analysis. The conversation analytic research tradition provides a host of insights that philosophical theorising tends to miss, including pertaining to the differences between face-to-face verbal communication and its written and remote analogues. It will also introduce the notion of communicative repair that will prove a crucial impetus for further development in the next chapter when I return more directly to the problem of moral repair.

From Indicative signals to social acts

Start with a common story about the place of testimony in our repertoire of strategies for expression and interpretation: As epistemic agents, we inhabit a world of indicative, evidential relations; as we move through space and time, we can come to know, not just what is directly observable to us – the objects or the scene presently before us, or the visible or sensible parts of our own bodies, say – but also what (parts of) these *indicate*; the footprints in the mud indicate that a moose has passed through the area recently; the small red spots on my skin indicate that I have measles. The prints and the spots, then, refer onto something else that I can come to know, or know about, *by* coming to know the indicative signs. The indicative relations are in these cases *natural*[42]; they depend for their significance on facts about moose, mud, and measles, and not on any agents' intentions, knowledge, or understanding of these phenomena. If no human beings existed, moose prints would still indicate that a moose has recently passed through; if no one yet knew what measles is, red spots would still signify the presence of the disease. Finally, some of these indicative phenomena are other epistemic agents; hearing someone clear their voice behind me,

[42] The idea of a contrast between natural and non-natural indicative relationships taken up by Moran draws from similar ideas in (Grice, 1957)

I can learn that I am not alone; seeing the small red spots on *their* skin indicates that they have measles, too.

In moving through a world of such natural indicative relations, an epistemic agent can glean a great deal simply *qua spectator*: utilising her powers of discrimination, and drawing on accumulated insights and recall of learnt relations between signs and what they signify, she could familiarise herself with an environment and the other agents within it, and learn to navigate these according to her basic needs and interests.

Of course, epistemic agents are not just spectators; they are actors, too. As a source of natural indicative signs herself, an agent can decide how openly to display them, or whether and when to conceal or suppress them. Some natural signs can be convincingly faked if one is sufficiently skilful.

We can also create new indicative signs by social convention (red light means *stop*; "GF" means *safe to eat for celiacs,* say). Here, unlike in the case of natural signs, the meaning the sign holds *is* dependent upon the understanding assigned to it by a community of users. Accordingly, one and the same sign can diverge in meaning between different such communities, or even vary in subtle ways within it, and shift over time. Consider, for instance, how different countries use slightly different traffic signs, road markings, and behavioural signals to organise an orderly flow of vehicles or navigate unforeseen obstacles together. Or consider the words of a natural language – with few potential exceptions, such as onomatopoetic expressions, word meaning is meaning by convention, and one and the same phonetic expression can have different meanings in different natural languages. The ability to create new signs, and particularly the infinite combinatorial possibilities of linguistic signs, vastly expands our expressive repertoire, and so extends the kinds of meaning we can signal to others. For instance, using the correct combination of linguistic units, one can broadcast the presence internal states like complex or abstract beliefs that do not have easily available behavioural manifestations. Just as you can roll up your sleeve and stick out your arm to display your measle spots, you can use a sentence of natural language ("I am considering becoming an osteopath") to 'display' the presence of a complex mental state.

According to Richard Moran, the above story represents the conception of *how we learn from being told* that is typical in contemporary epistemology. We relate to other people's speech as indicative signs or evidence of the internal states that their speech represents - their beliefs, desires, aspirations, and other attitudes - and confront these as we would any other observable phenomena. Learning from being told, on this view, is fundamentally similar to perceptual learning. We extend our own powers of observation by using other people's reports, just as we might extend them by using a microscope or pair of binoculars.

However, Moran argues, this story contains some critical errors. In particular, it is a mistake to assimilate our ability to convey knowledge from one agent to another by linguistic means with our ability to emit and decode meaningful indicative signs, as the extended perceptual model would have us do[43]. There is, he argues, a categorical difference between the two methods of transmission, and it comes down to the agential interdependence we paradigmatically find in the former case, and not in the latter. This idea will require some unpacking.

If I see someone shudder, or stub her toe and cry out, I can normally conclude that she feels cold or is in pain. In these cases, her observable behaviour functions as a natural sign she leaves behind in the world, which I, or anyone else suitably placed, can inspect to ascertain how things stand. What she does provides reasons to believe independently of whether she intends it; she may not realise she is being seen, and may even have been trying to give off a contrary impression.

By contrast, if she *tells me* that she is cold or in pain, the significance of her words entirely depends on the way she intentionally stakes herself. Unlike her shudders or cries of pain, her words count in the way that they do –namely as an expression of what she knows or believes– only *because* and *insofar as* she means them to. Accordingly, she will not have testified anything at all, not have told anything about

[43] (Korta and Perry, 2024) label a model along these lines the Locke-Saussure model. See also (Bavelas, 2022a, 2022b) for a survey of approaches in the communication sciences.

what she believes, if she is utters them in her sleep, in practicing her pronunciation, or as part of the telling of a joke.

Of course, no speaker is entirely unconstrained. In utilising a language, she is bound by existing conventions of meaning and limits on intelligibility that are not hers alone. But these conventions typically leave her the freedom to perform any number of acts of speech by the use of any particular sentence. For example, "I am cold" can be used to declare one's bodily state, or to demand or request that someone to close the window, and which of these a speaker in fact carries out in a given situation is settled by what she intends.

Saying this much may not suffice to convince anyone that there is anything all that distinctive about learning from being told. A response to the above might go as follows: Although learning from speech depends upon the identification of the speaker's intentions (as well as the background existence of the convention itself), these are up-stream facts that do not change the fundamental task of the inquirer. A unique aetiology or set of signal-emission conditions does not suffice for a categorical difference. Shudders, for instance, depend on complex features of the human nervous system, moose prints depend on properties of water, climate, and soil particles, and traffic signs depend on specific legislative and regulative processes, but all three are indicative signs all the same. In either case, that is, one may require background or contextual information about the relation between signs and the subject matter they signify to understand their meaning; and in either case the more skilful and attentive an interpreter one is, the better able one will be to gain knowledge from confrontation with the phenomena. But requiring intentional production for meaningfulness is not sufficient to set testimony categorically apart from the rest.

To make the case for a categorical difference, Moran's account instead emphasises the distinctive role for the recipient of a telling. It is not *just* that a speaker selects and produces her meaningful signal intentionally, and that her addressee must detect or attribute this intention in order to decode it correctly. Instead, what a speaker does in telling her hearer something is to *take responsibility* for the truth of that very claim and thereby establish a distinctive mode of relating between them (Moran, 2018, p. 18)[44]. She *gives assurance* of what she testifies, and her doing so depends not just on her own understanding and intention to do so, but also on the *reciprocation and uptake* of the person to whom she gives it. Her addressee therefore has his own role to play in the undertaking, a role without which the telling cannot be accomplished; and this complementary role for another person has no analogue in the case of mere indicative signs. If the hearer does not play his part, *there simply is no such item* to be interpreted at all and no reason to believe[45].

It is therefore precisely because *telling takes two* that it is categorically unlike indicative signs. Telling institutes relations between agents, not just as producers and consumers of independently meaningful signals, but as participants in a cooperative endeavour, who together generate and sustain the meaningfulness of the message conveyed.

It bears noting quite how radical a proposal this is on the face of it. Several features of our epistemic dependence on others are familiar and widely accepted in philosophical discourse. For instance, it is common to note that we depend on other people in our *acquisition* of knowledge in that we rely on them as sources for facts we ourselves have not ascertained or personally verified, and indeed could not ascertain or verify for ourselves[46]. Furthermore, as the notion of epistemic injustice has gained recognition within mainstream philosophy[47], it is generally accepted that a speaker's epistemic agency as testifier can be curtailed if her credibility is underestimated, or if her linguistic community lacks the hermeneutical resources to make her experiences intelligible, whether to herself or to others. However, one can accept all of this without

[44] Moran's position on testimony explicitly draws from work by Angus Ross (Ross, 1986), Edward Craig(Craig, 1990), and C. A. J. Coady (Coady, 1992), but also bears similarity to conceptions of assertion put forward by Brandom (Brandom, 1980) or Watson (Watson, 2004). See also (van Roojen, 2020).

[45] Or, strictly speaking, *no reason of that kind*; for speaker's attempt at telling her hearer that P may cause the hearer to form a particular belief (including a belief in P) in some other way – e.g. by prompting the hearer's recall of previously acquired reasons for belief (in P), or in setting the hearer's independent consideration of or reasoning about the matter into motion. In such a case, the speaker's utterance is the *occasion* for the hearer's belief, even though it will not be the case that she has told him. She might as well have had this effect by wearing a T-shirt with P written on it.

[46] For instance, we need other people's reports to find out about the circumstances of the past, or of our own early life, or to inform us about domains that require specialist expertise.

[47] Coined and influentially presented in Fricker (2007); other feminist and social epistemologists had taken note of a similar phenomenon – see e.g. Code (Code, 2011, 1995, 1991).

agreeing that the ability to testify as such depends on receiving reciprocation from one's addressee.

To see the difference this makes, consider a somewhat parallel debate about sexual violation and consent. It is one thing to hold that (say) patriarchal norms and expectations mean women's attempted sexual refusals are often misunderstood or not taken as sincere at face value ("when she says 'no', she is only being coquettish; she *really means* 'yes'"). It is altogether another thing to hold that women whose partners do not understand them *cannot* give or withhold consent. In holding the former view, we can say: Though he misunderstood, she really did refuse sex. If we hold the latter, we would instead say: Since he misunderstood, her attempt to refuse was unsuccessful; under conditions like these, she lacks the ability to[48].

In the case of testimony, accepting the idea that testimony requires reciprocation and uptake means that a speaker has not testified that p if her hearer was in fact out of earshot, or could not understand the language she was using, or had fallen asleep, or thought she was merely joking or practicing her lines for a play. Instead of saying: 'She told him that p, but he did not hear/understand/etc…' , we would have to say: 'She *tried* to tell him that p, but he did not hear/understand/etc …' Telling becomes a success notion, and a new category of *attempted but unsuccessful tellings* suddenly presents itself.

This might seem an ontological extravagance. Why make the additional step and make the power to tell conditional on the addressee's uptake as Moran does and make space in one's scheme for both successful telling, merely attempted telling, and communication by means of indicative signs? A few different grounds are relevant to mention. First, we might simply hold that accounts of testimony that pay no heed to the distinction between successfully telling, merely attempting to tell, and communication by means of indicative signs fails to carve reality at its joints, and so

[48] Importantly, irrespective of how we classify the woman's speech act, this need not be all there is do say about the normativity of the encounter. There may be other grounds for categorising any resulting sexual contact as consensual or not, as morally objectionable or not, as blameworthy by the initiator or not, and so on. All the same, for some theorists, what we are able to say about the normative status of the woman's (attempted?) refusal is a significant question in its own right. For discussion of these issues, see e.g. (Caponetto, 2021; Maitra, 2009, 2004)

that one "risks losing touch with a distinctive philosophical topic altogether" (Moran, 2018, p. 18) by failing to mark these out.

Second, doing without these distinctions makes various features of our common speech practices very hard to understand and explain. For instance, when it comes to ascertaining people's attitudes, we generally afford special authority to those people's own first personal reports. But why? As Moran notes, considered as indicative signs, viz. as mere evidence of what is in the speaker's mind, belief reports are riskier and open to more forms of intentional distortion and deception:

> "When I learn of someone's beliefs through what they tell me, I am dependent on such things as their discretion, sincerity, good intentions—in short, on how they deliberately present themselves to me—in a way that I am not dependent when I infer their beliefs in other ways." (Moran, 2005, p. 5)

Moreover, when it comes to evidence, the awareness that someone has deliberately produced and presented it for one's consumption generally draws its reliability into question. It is now 'doctored evidence', and should be viewed with a greater degree of scepticism. As such, a preference for such indicative evidence over what one can independently discover or infer becomes mysterious:

> "If the epistemic import of what people say is at bottom that of an indication of what they believe, it would seem perverse for us to give any privileged status to the vehicle of knowledge (speech and assertion) where we are most vulnerable because most dependent on the free disposal of the other person."(Moran, 2005, p. 6)

By contrast, if the speaker's intentional utterances are not just one more indicative or evidential source, but a distinctive form of guarantee, its privileged status makes sense. What one gets is here "different in kind, though not necessarily in degree of certainty, from beliefs I might have read off from his behaviors".(Ibid.) [49]

[49] Note that this is precisely the problem I considered for the reassurance version of the bookkeeping model in the previous chapter. There, too, we confront a problem about why we should prefer people's own reports or declarations of their internal states to ways of ascertaining these that bypass their active agency.

Finally, and perhaps most importantly, the very possibility of relating to someone's words as mere indicative evidence of her internal states itself becomes incoherent without also accepting the kind of story Moran promotes. To see this, imagine two hearers who take different attitudes to what they are being told by a speaker. The first hearer, who we will call A for *assurance*, understands and accepts the speaker's intentional telling as such. In doing so, A takes up a place in a normative relationship with the speaker that grants distinctive duties and entitlements to each of them. For instance, if A wants to disseminate what he has been told to a third party, and the third party challenges A's assertion, A gets to pass the justificatory buck back to the original speaker. A also gets a right of complaint against the speaker if what he has been told turns out to be false. The speaker stakes her reputation on the truth of her telling, and by accepting this, A gains the right to hold her to account *for it that p*.

By contrast, the second hearer, who we will call E for evidence, refrains from accepting the speaker's word as a telling, and instead treats it as mere evidence. Though the speaker invites E's reliance in the same way as she invites A's, E forbears from accepting it – he simply does not "go in for that sort of thing"(Moran, 2018, p. 70). What makes the speaker's assertion evidence for E? It seems E must think something like the following: 'Well, the speaker clearly means and offers to take responsibility for the truth of what she says. She would not do this, and so offer to incur the duties and entitlements that go along with it, if she didn't have good grounds for what she says. So, I'll believe her.' Since E does not accept his role in the normative transaction, he does not receive the entitlements to buck-passing, complaint, and so on that A gets. Instead, there is a sense in which he free-rides on the social practices associated with telling – he takes the epistemic good the speaker offers in marking her contribution as a telling without himself participating in sustaining that practice.

Crucially, however, while it is possible for E to free-ride in this way a given case, it would not be possible for *everyone* to do so in *every* case. If no one ever accepted speaker's guarantees and held them to their testimonially offered commitments, then such commitments could not count *as* evidence in the way they do for E.

In this way, Moran argues, telling is similar to promising, contracting, and other act types that depend on a relational social convention. The existence of the practice allows you to invest your words with additional weight at will, but only because the other party is generally prepared to enforce the convention in cases of violation. While a recipient can occasionally stand back from such practices, refusing to play their part in them, and simply treat the speakers' offers of assurance as evidence of what a person believes or will do, their ability to take up this detached attitude depends on its being the exception, rather than the rule. If everyone treated tellings, promises, and the like as mere evidence, they would *cease to be* evidence.

This argument therefore makes the case for a reorientation in the epistemologist's method of cases. It is true that there are cases wherein we can and do treat what we are told as mere evidence, and so relate to the speaker's words as we would any other indicative sign. However, these cases are peripheral, rather than typical – they are necessarily aberration, rather than the paradigm. The real paradigm, and the case that displays the epistemic distinctiveness of testimony, is the case in which the speaker's telling is treated as an act of assurance, and treated that way not just by the speaker, but by the addressee as well. We fail to identify these latter because our representations of testimony have zoomed in too closely: we have been looking just at the emission of a meaningful signal and not at the manifest intention with which it is produced and the uptake it receives from its hearer. But both of these should in fact be included within the phenomenon – testimony is more extended (or composed of more parts) than we thought.

Social acts and the Active-Passive distinction

Intriguing though this proposal is, it raises the further questions: What, precisely, is the role the hearer has to play? What does it mean to give recognition and uptake, and how does doing so contribute to the provision of a reason for belief a hearer acquires in being told? I will argue that Moran's discussion fails to acknowledge a distinction between two sub-categories of social acts, which I will label as the Active and Passive categories. While acts in both categories require another person for their completion, the role played by this other is importantly different. Actions belonging to the Active category require another's active and voluntary reciprocation for their performance. By contrast, actions belonging to the Passive category require only their understanding and awareness, something a hearer cannot ordinarily withhold at will.

For illustration, consider the contrast between the speech acts of marrying and of proposing marriage[50]. A typical contemporary marriage ritual requires the active interventions of two parties (in addition to the officiator), each of them actively voicing their understanding and their willingness to enter into a marriage with the other. Although one of the two speaks last, and so in a sense seals the arrangement by their very action, what they have thereby done only counts as marrying if the other person has spoken their assent as well (in addition to the satisfaction of whatever other felicity conditions have to be in place, such as that they are both unmarried, of legal age, etc)[51].

By contrast, as any reader of a Jane Austen novel will know, it is very much possible to *propose* marriage without the active reciprocation of another and indeed against that other's will and wish. Although a marriage proposal refused is in one sense (i.e. the most obvious one) an unsuccessful proposal, it is not a defective or incomplete one; not unsuccessful in the way it would have been if the person speaking had suffered a coughing fit at the crucial moment, or if they had addressed their beloved in a language that person could not comprehend.

Similarly, contracting a sale, agreeing to meet for lunch, or placing a call bet in roulette[52] requires the active, voluntary participation of two people; offering a sale, inviting someone out to lunch, or placing a standard roulette bet only requires the understanding and awareness of the other party. Nor is the distinction between the categories restricted to act-types involving the use of words: Greeting someone by handshaking requires two people to stretch out their hands, grip one another, and shake; greeting someone by waving only requires one person being seen and

[50] Marrying is a recurring illustrative case for an illocution, e.g. in (Austin, 1975)

[51] A style of case famously discussed by Austin.

[52] In a call bet or announced bet, a roulette player *states* their bet, rather than put their chips on the table square(s) corresponding to the bet they want to make. The bet is not placed unless the croupier repeats the bet back to the player. A player may use a call bet because they are placing a more complicated "French bet" across several parts of the wheel. Call bets are prohibited in some countries since they are in effect bets on credit.

understood to do that very thing by another; and *mutandis mutatis* for congratulating someone by high-fiving them versus by slapping their back.

Finally, it requires a monetary system, the existence of private banking and credit cards, and the authorisation of my account manager, as well as the institution of private property available for purchase, for me to be able to make an overdraft; but once these various background conditions and practices are in place, I can, in an obviously meaningful sense, *go into debt on my own*, as the song has it.

In each case, then, the possibility of carrying out the action at all trades on the existence of various practices, institutions, and conventions, and the general acceptance and comprehension of these by the participating parties; but once these conditions are in place, the actions in each pair vary in how an individual agent can mobilise their meaningfulness and normative import for the achievement of an intended effect. Can an individual agent do so on their own, or only with the active and willing participation of another?

The distinction between the Active and Passive sub-categories is obscured by Moran's use of words like *recognition* and *acceptance*, for these words themselves are commonly used in either an active or a passive sense. For instance, in its passive sense, to recognise means something like to *detect* or *(re)identify* it. In this sense you recognise some item, location, or person spontaneously and non-voluntarily simply upon perceptual confrontation in ordinary circumstances. By contrast, one can recognise something or someone in the active sense through an explicit and voluntary act of acknowledgement, e.g. in introducing some item into a meeting agenda or acknowledging a person as the next speaker in a formal proceeding. Similarly, one can accept some proposition as true, or a joke as offensive, or a work of art as beautiful in a passive sense simply by seeing that it is true, offensive, or beautiful; or accept it as such actively by treating it as such in one's behaviour towards it – by acting in accordance with its truth, offensiveness, or beauty when one also has the ability to do otherwise.

One might wonder whether the distinction I am drawing between active and passive is really as sharp as I am making it out to be; detection, identification, and interpretation, it may be held, themselves require an intentional agential effort – e.g. for memorial recall of the relevant identification criteria, for sustained attentive observation and inferential processing, or to put the right mental label on something or someone one finds opaquely familiar but can't yet place ("I've seen that woman before – but could she be someone from the gym, or a party, or was she on TV?"). Additionally, the act-types I am describing as voluntarily reciprocated by the second party might seem not to be so in either of two ways; firstly, the second person can be compelled or coerced in the fulfilment of their role; secondly, their contribution may be automatic, unreflective, or spontaneous rather than actively and consciously planned, willed, or carried out.

But these objections can be met without undermining the contrast. Firstly, while it sometimes requires effort to remember or reflect on something, and while such efforts at mental agency can therefore to that extent be voluntarily chosen or avoided, one is not free to choose their outcome – not free to choose *what to see* as one is free to choose *what to do*. If you *do* remember or reason that those markings mean the bird is a starling, you cannot decide to recognise it as a blackbird instead. Recognition and acceptance in the passive sense are firmly held to a mind-to-world direction of fit, and this imposes strict limits on agential voluntariness. Secondly, in the cases that concern us here, we are dealing with recognition and acceptance of act-types that are extremely common and designed to be easily and unambiguously interpretable within their communities of use. The possibility of occasional misconstruals notwithstanding, it does not ordinarily take a great deal of memorial or interpretive effort to see a wave for a wave or a proposal for a proposal, particularly when embedded within a context that licenses certain expectations between the parties.

Furthermore, that the alternatives to active reciprocation are sometimes so constrained that an agent will find that he *can do no other* does not prevent his actually doing of it from being the product of his will. Being compelled or coerced in certain ways might absolve an agent of certain legal or moral ramifications that would follow if his act were not so compelled, and at the limit may even invalidate what he is doing as an act of the relevant kind at all (perhaps he could not genuinely *promise* or *confess* for example, if he was held at gunpoint); but we can accept this and nonetheless hold that operating within these limits[53], we can make good sense of the idea that an agent did something voluntarily, and that he could have refrained from doing it, even if only at great (or even exorbitant) cost to himself or others. Nor must we deny that reciprocal actions can be automatically, carelessly, or spontaneously performed, as highly conventionalised behaviours often are. Again, what matters is only that it is in principle open to the agent to reciprocate or not.

Admittedly, these questions are not always so clear cut, and more can be said on behalf of this complaint; once can arguably cultivate perceptiveness as well as ignorance and inattentiveness, for example[54]. All the same, I hope I will have said enough to make the case that there is a genuine distinction here, one that is at least sufficiently clear and robust for present purposes.

Telling and the impossibility of resistance

So far, I have introduced Moran's distinction between tellings and indicative signs on the basis that tellings, but not indicative signs, requires for the transmission of reasons for belief that a certain intersubjective relationship between two parties be instantiated. Specifically, it requires the speaker's manifestly intentional action of taking responsibility for the truth of her claim, and the hearer's recognition and uptake thereof.

I have then argued that a further distinction can be drawn amongst acts that require this intersubjective structure, namely between those wherein the performance of the act requires the active reciprocation of another –something that other can withhold at will– and those that merely require their understanding and awareness. The point I want to convince you of in this section is simply this: telling must belong to the Passive category. Its occurrence requires the understanding and awareness of the addressee, but no more.

[53] Different act-types will have different such invalidating conditions built into them, and practitioners of the convention will normally have, and expect others to have, sufficient grasp of these to make the outcome in a given circumstance tolerably clear.

[54] For discussions of forms of pernicious and motivated ignorance that are both caused by and collude in oppression and marginalisation, see e.g (Medina, 2013) and (Dotson, 2011).

To see that this is the case, consider the familiar phenomenon I will here label *the impossibility of testimonial resistance*: one simply cannot refuse an unwelcome telling. Consider an oversharing parent telling her teenage son about her recent visit to the doctors ("I am telling you, it was completely swollen and blue, my doctor says she's never seen anything like it!"), or being stuck on a train ride with a stranger who wants to describe his recent divorce. However strongly one wants to resist someone's attempt to put one in possession of such information, there is little one can do to avoid it.

This conclusion may sound overblown: surely, it may be objected, one can cut the speaker off, or ask to be left alone. This is true, but much like in the case of an unwelcome proposal or invitation, even this will not prevent a sufficiently determined speaker. If a (would-be) hearer is sufficiently fast, or the utterance sufficiently long, one can cover one's ears, drown the telling out by loudly shouting over it, or quickly remove oneself from earshot, but in practice this will only work in a subset of cases[55].

Although Moran explicitly recognises the reality of this kind of communicative situation (Moran, 2019), and at various times appears to use 'recognition and uptake as synonymous with 'understanding and awareness', he also describes telling as involving "two distinct freedoms" (Moran, 2018, p. 133), and cites with approval Grice's claim that "the intended effect [of a telling] must be something that in some sense is within the control of the audience" (Grice, 1957, p. 385). In elaborating his own account of telling, Moran writes:

> "It is not the speaker's aim that the belief in question be produced by the audience's simply being so constituted that his awareness of the speaker's complex self-referential intention somehow produces the belief in him. (Moran, 2018, p. 60).

[55] This impossibility of resistance is precisely what makes some communicative environments so invidious. If ill-intentioned speakers position themselves on platforms or in locations the rest of us have to pass through, we cannot help ourselves from taking their contributions in and thereby affording a kind of uptake. See e.g. (McDonald, 2021) on cat-calling. See also Thomas Schelling's classic discussion of the strategic resorts to communicative inaccessibility (Schelling, 1980).

> "the mutual recognition of [the speaker's] intention can play the role for the audience of providing him with a reason for belief, because he sees the speaker as presenting herself as accountable for the truth of P, and asking, through the recognition of her intention, that this offer of assurance be accepted. *And it is understood by both parties that this acceptance is something which the audience is free to give or refuse.*" (Ibid.,61, emphasis added).

If 'accepting the speaker's offer of assurance' means utilising the entitlements that one is granted in being told, then Moran is surely right that this is something the audience is free to give or refuse. One can decline to hold a speaker to her claims and apply one's right of complaint if what she tells turns out to be false. But one cannot prevent oneself from being assured of P in the first place, and this already makes a significant difference, both to her normative[56] and her epistemic position. Focussing for the moment on the latter, it should be recognised that whether a hearer finds a speaker to be credible and competent as a testifier (about P, or in general), is of course not itself directly under a speaker's control; she is therefore not entirely free to implant a belief in her hearer's mind simply by telling him. But neither is it under the hearer's control. A particular attempt to tell may draw a speaker's credibility into question if what says is sufficiently astonishing, but barring such circumstances, the antecedent standing assessment of the speaker's credibility will settle in advance whether a speaker succeeds in transmitting a reason for belief, even if the hearer would rather she had not.

Finally, an unwilling hearer can, depending on his skill, resort to strategic dissimulation – he can pretend to be asleep, or unable to hear or understand, or treat the hearer as someone lacking in the authority to make assurances. But while such uncooperativeness may prevent the exchange from progressing any further (unless the speaker suspects his duplicity), it will not wipe-out the transmissions of information that have already come through.

[56] Though Moran does not devote much attention to this, the recipient of a telling plausibly incurs obligations, and not just entitlements, of her own. See e.g. (Watson, 2004), and (Kukla and Lance, 2009; Lance and Kukla, 2013).

In short then, telling qua social act is more like proposing, inviting, and overdrafting than marrying, agreeing, or contracting[57]. In telling and being told, the hearer and speaker are not partners in a common activity, but rather agent and patient in the speaker's singlehanded exercise of unidirectional normative power to bind. The speaker's act of telling singlehandedly causes the normative and epistemic ground to shift beneath the hearer's feet; and while he has choices to make about how to conduct himself in this new landscape (whether and how to hold the speaker to her commitments and exploit the entitlements he has been given; whether to (attempt to) conceal his receipt of the telling, or openly and actively acknowledge it), he is not there of his own free will.

The threat of unilateralism

So far, I have argued that Moran's characterisation of social acts as act-types the performance of which requires the recognition and uptake of another admits of a further distinction, namely between what I termed *Active* and *Passive* social act types. Whereas the act-types in the former category require active reciprocation from another, the act-types in the latter category do not, and their performance can thus not be avoided at will, save by disrupting or exiting the communicative situation entirely before the speaker's performance is completed.

I then argued that the social act of telling must belong to the passive category. While its possibility clearly requires others in a general sense in that it trades on interpersonal practices and conventions, and while its performance in any given case depends on another person's awareness and understanding, against this background a person can indeed perform a telling on her own – and, if she is credible, offer her hearer a reason for belief. In this, I suggested, telling is like proposing, inviting, offering, or going into debt. Indeed, the analogy runs deep, for on Moran's view the provision of a reason for belief for one's interlocutor works precisely by providing that

[57] Moran frequently compares telling to both contracting and promising. Here, I have put less weight on the second category, since it is itself a controversial question whether one can make someone a promise against their will. See e.g (Scanlon, 1990)

person with a guarantee – a promise of a kind of warranty in the form of a right to complaint, should things not be as one claims them to be.

This matters for Moran's project of vindicating telling as a genuinely distinctive way of providing others with reasons for belief, something categorically different from the provision and interpretation of indicative signs. For Moran's point was precisely that telling is distinctive in that its manner of reason provision *takes two*. The speaker has to manifestly intentionally present herself as taking responsibility for the truth of her claim in assuring the hearer of it; and in addition, the hearer has to do something, and *does* do something when he takes her as a conversation partner rather than a mere source of signs, namely: accept her assurance and thereby take her act of speech as she intended it.

But if this role for the hearer is not one he can freely accept or decline, there is little distance between the *indicative* or *evidentialist*[58] construal of testimony Moran sets out to oppose and the alternative intersubjective one he defends. In both cases, the hearer is a mere recipient rather than a cooperative partner, and in both cases the speaker impinges on the hearer's state for the production of a result, providing him with reasons for belief unilaterally.

To be clear, my argument is not that the difference between testimony and evidence disappears altogether when we recognise that testimony is a passive social act. Rather, the character of the difference changes. We can distinguish three different models of telling that have so far come up:

The initial conception of testimony (and Moran's principal target) was what we could call an *emission model*: telling requires just that the speaker emit a meaningful signal, whether or not anyone is there to see it and capable of decoding and understanding it. Testimony is on this model like the flashing beacon of a lighthouse.

The model Moran proposes is what we could call a *transaction model*. On this model, testimony requires two parties, each of whom is openly and willingly participating in

[58] Moran seems to be using these labels interchangeably for the family of accounts of testimony he means to reject, viz. those that conceive of telling as one more indicative phenomenon and ignore or overlook its interpersonal normative character.

the transfer of normative entitlements and obligations. Testimony is here likened to striking a deal and shaking hands – it requires two parties actively participating. This model, I've argued, misrepresents the voluntariness of telling, and particularly the voluntary participation of the addressee.

At a mid-point between these two models is a third, which we could call a *transmission model.* On this model, telling requires the presence, awareness and understanding of two parties, but only activity of one. Like inviting, waving, or proposing, the act cannot be completed if no recipient is present, or if the recipient does not comprehend the action, but no further contribution beyond understanding and awareness is required. Provided the message is easy to understand and the telling loud enough for them to hear, etc., all you need to do to tell, on this model, is deposit your message with its addressee, like placing a letter in a mailbox.

By highlighting the impossibility of testimonial resistance, I have argued that the transmission model is the correct construal of telling. While the transmission model, too, affirms the notion that *telling takes two*, it does so in a manner that is relevantly different from Moran's intended transactional construal.

The unilateral character of acts of telling also causes trouble for an otherwise natural interpretation of Moran's notion of social acts, namely that they are a species of what philosophers, psychologists, and developmentalists have otherwise referred to as *joint action*. Several commentators on Moran (Fricker, 2019, 2021; Lawlor, 2021) have drawn this connection, but if I am right that a hearer can 'fulfil their role' while remaining not just agentially passive but also psychologically unwilling, then the alignment of the two notions will be difficult to defend. As Lawlor notes, even minimally demanding conceptions of joint action tend to attribute to the parties something like shared goals (Lawlor, 2021),[59] and in the kinds of cases under consideration no such goals can plausibly be ascribed.

[59] While Moran appears willing to concede this ("it would strike me as misleading to call this "joint action", especially in the sense of the speaker proposing that they do something together" (2021)) it is somewhat difficult to reconcile this claim with his repeated insistence across (Moran 2018) that *doing something together* is precisely the right way to think of testimony.

Chapter Four

Complex testimony and the social character of speech in interaction

> "The solution surely is provided for by a resource that is itself built into the fabric of social conduct, into the procedural infrastructure of interaction".
>
> Emmanuel Schegloff, 1992

In the previous chapter, I considered a philosophical debate over the epistemic character of testimony. A number of philosophers have sought to defend the notion that there is a categorical distinction between intentional communication and the emission of decodable signals. Amongst these, I focussed my attention on the proposal put forward by Richard Moran. Intentional communication, Moran argues, acquires its distinctive epistemic significance by establishing a normative relationship between the parties – a relationship akin to the one found between promisers and promisees, or the parties to a (bilateral) contract. In intentionally communicating, a speaker does not merely provide evidence for some fact, but takes responsibility for its truth, and in so doing confers a set of normative entitlements on her addressee.

Like in the case of promises or contracts, therefore, the receiving party has a kind of security against non-fulfilment: If what he has been told turns out to be false, he has the right to complain and criticise the teller for her duplicitousness (if she purposefully deceived him), or for her irresponsibility or incompetency as an informant (if she believed what she said was true, but ought not to have done so). And like in the case of promises or contracts, an act of the relevant kind only exists if the speaker manifestly intentionally stakes herself in this way, and her interlocutor gives recognition and uptake to this self-presentation. A reason to rely on the speaker's word is not granted if she utters the right string of sounds but only in her sleep, or as a joke, or in rehearsing her lines for a play; nor is it given if she really means to tell, or promise, or to enter a contract, but receives no reciprocation from her addressee – e.g. if her speech falls on deaf ears, or on no ears at all. The distinctive kind of reason-provision characteristic of testimony (and of promising and contracting) in a phrase, *takes two*; the speaker and hearer must cooperate in the exchange of words. Testimony is a social act.

Though sympathetic to the underlying idea that the epistemic significance of testimony involves the establishment of a normative relationship between the parties, I nonetheless argued that testimony is importantly different to contractual agreement. While a speaker only takes responsibility for the truth of some fact if she speaks intentionally, she does not need her interlocutor's cooperation for the reason-provision distinctive of testimony to occur. Social acts, I argued, fall into two categories: those that require the other party's active and voluntary reciprocation (*Active* social acts) and those that require only their awareness and understanding (*Passive* social acts). Contracting, I argued, sits in the former category, alongside marrying, hand-shaking, or high-fiving; but telling sits in the latter, alongside proposing marriage, inviting, waving, or slapping someone's back. When it comes to passive social acts, one party can accomplish the relevant performance (make it the case that they have proposed, invited, greeted, or congratulated) unilaterally, and altogether without or even *against* the other party's will.

This complicates Moran's story of a categorical difference between telling someone that P and emitting a signal that *indicates* or *is evidence of* P; though it is still correct to say that *telling takes two*, telling is, I argued, more akin to a transmission than a transaction: The addressee figures in the interaction as a passive witness to and target of the speaker's unilateral exercise of a normative power to bind, not as a collaborative partner in her transfer of information.

This chapter makes a perhaps surprising turn of direction. Having just argued *against* the idea of the hearer as a collaborative partner in communication, I now wish to revive it, albeit in a different guise and setting. For the line of argument pursued so far has been operating with a number of implicit restrictions in scope that we now need to cast off. Once we do, I shall propose, a whole new prospect on testimony, agency, and communication opens. This perspective provides a kind of vindication for the spirit of Moran's proposal, but also demonstrates where Moran himself fails to pursuit that spirit sufficiently far.

The key insight is the following: the argument of the last chapter implicitly envisioned the *objects* of testimony as individual, self-standing sentences (typically declaratives), or what Rachel Fraser (Fraser, 2021) has called *simple testimony.* Examples of simple testimony could include such statements as "I am considering becoming an osteopath", or "Your train leaves from platform 3" (Fraser, 4029), or, in Moran's own examples, 'I am travelling to Minsk' and 'It is raining in Spain'. However, as Fraser rightly notes, much, or even most, of our telling does not take this form. We also tell others what we know or believe in extended ordered sequences, sometimes proceeding over several conversational 'turns' at talk. Call this *complex testimony*[60].

Fraser argues that complex testimony has emergent epistemic properties that distinguish it from amalgamations of simple testimony; in particular, she argues that complex testimony increases a hearer's dependence on a speaker because of the way complex testimony can be laden with a speaker's content formatting choices and representational cues. However, while Fraser's attention to complex testimony is a useful corrective to the philosophical overemphasis on testimony offered in single self-standing utterances, I shall argue that she, too, underestimates its interactive

[60] The term 'complex testimony' is my own. Fraser instead contrasts simple with *narrative* testimony, but also notes that not all discourse is narratively structured (n2, p 4026), thus suggesting that narratives are a species of the broader genus *discourse*. Presumably, other such species could include a sequence of instructions, or extended lines of argumentation, say. Quite where to draw the line between these categories is unclear, but for present purposes also unimportant.

character. Complex testimony does increase the hearer's dependence on the speaker, but it also increases the speaker's dependence on the hearer in ways philosophers generally fail to notice.

To identify and remedy this deficit, I turn my attention to a set of different disciplinary perspectives, developed within psycho- and sociolinguistics and a sub-discipline known as conversation analysis. Attention to these bodies of research reveals that much of our communication is not just complex – delivered in extended structured discourses of interlocking claims – but *dialogical and collaborative*; a speaker's contributions are continuously facilitated and scaffolded by the active interventions of her hearer. Even when one person has the communicative initiative, i.e. one person is the *teller* and the other the *tellee*, the interlocutor is an active collaborator whose participation is giving shape and direction to the exchange.

Philosophers have failed to take adequate note of this fact, I propose, because they have assumed that spoken communication is relevantly similar to written communication, and therefore overlooked a number of crucial differences between the two. The difficulty here is in part a methodological one; without empirical observation of real world conversation, it is hard to put the evanescence and untidiness of actual talk before one's mind.

Drawing from empirical evidence from the communication sciences, I consider three phenomena that display the interactiveness of ordinary talk: backchanneling, communicative repair, and co-narration. Reckoning with the reality and role of these phenomena, I argue, requires us to rethink some of our ordinary assumptions about the epistemology of testimony, including pertaining to a speaker's autonomy over and responsibility for what she says, and the intersubstitutability of hearers. A hearer who is unwilling or unable to scaffold a speaker's communicative contributions (including her attempts to transfer what she knows and believes) can significantly curtail that speaker's epistemic agency, whereas a skilful and supportive hearer can extend it.

From simple to complex testimony

We sometimes transmit knowledge to other people by means of individual, isolable sentences. However, much of the time, we instead tell by means of *complex testimony* – "a structured discourse of interlocking claims" (Fraser, 4026). For instance, one might recount one's journey to work by means of a narrative composed of several ordered sentences, as in:

> "I arrived late at the station today and almost missed my train. As I was running, the guard saw me and held the door for a second, and I made it." (Fraser, 2021, p. 4030)

Complex testimony, Fraser argues, has emergent epistemic properties that render it irreducible to simple testimony; discourses are not just amalgamations of self-standing individual sentences. In particular, complex discourses organise and structure their encoded contents in ways that characteristically increase the hearer's dependence on the speaker: the hearer receives the information in a more pre-packaged form that compels him to accept more of the speaker's perspective on the world.

There are several features to this dependence. When content is presented piecemeal, self-standing sentence by self-standing sentence, as in simple testimony, the hearer has more opportunity to call out and challenge the speaker's implications and assumptions and make up his own mind about the subject matter; but when content is pre-packaged into a structured discourse, he loses some of this autonomy. Elements embedded within the narrative are less accessible to further probing or critique - they are not marked as at-issue or under discussion in the speaker's testimony. Moreover, the hearer needs to adopt much more of the speaker's representational formatting to even consider the information he has been offered. This in turn affects the hearer's attentional, interpretive, and inquisitive dispositions. For instance, "[w]hen a narrative—fictional or not—coaxes us into representing its protagonists [in one way rather than another], it 'selects' a representational format for us"(Fraser, 2021, p. 4036), Fraser argues:

> “Narratives which cue a simulationist representation of a given agent [viz a first personal perspective on their experiences] incline us to adopt the reactive stance towards said agent, and so to adopt a certain suite of interpretive and inquisitive dispositions. Conversely, manipulationist representations [viz a perspective on them as an object we can utilise and regulate in specific ways] incline us towards the objective stance, with its attendant suite of dispositions. Thus representational format influences inquisitive and interpretive dispositions by shaping (i) which kinds of question naturally occur to us concerning an agent, (ii) which sorts of answers close our inquiry, and (iii) which inferences we are inclined to draw when we get new information”. (Fraser, 2021, p. 4037)

Importantly, it is not *impossible* to trigger these perspectival and representational effects in simple testimony; it is just easier to do so in an extended discourse. A single sentence provides only a small number of opportunities to plant representational cues, whereas a protracted narrative provides many; indeed an extended narrative (such as a novel, say) can only be comprehended by gradually taking on board more and more of the world one is being offered, at least temporarily.

It is also not *impossible* to resist the representational formatting one is being offered in complex testimony. Clearly, one need not actually believe everything (or anything) the narrative tells one – e.g. that only some of its characters have relevant first-personal perspectives, whereas others are appropriately viewed as objects to be manipulated. One can ‘accept’ this as true only within the fictional world of the narrator, for instance, while remaining internally recalcitrant, or consciously posing the kinds of questions the narrative structure tries to occlude. The point is simply that a complex testifier’s pre-selected representational structure makes certain modes of response more or less accessible to the hearer, and that her ability to exercise this power increases with the complexity of her testimonial contribution. Narrative and other forms of complex formatting can impose specific interpretive scripts, make certain inferences salient or plausible, and encourage the hearer to inhabit a specific world-view.

On the face of it, then, complex testimony only exacerbates the agent-patient structure I have previously argued is characteristic of simple testimony; in receiving complex testimony, the hearer is *even more* dependent on and susceptible to the speaker's exercise of epistemic agency and so even more passive. Not only can the speaker force the hearer into a normative interpersonal relationship at will or put him in possession of information he prefers not to receive; she can also coax him into a particular representational stance and way of seeing the world, depriving him of opportunities to structure contents for himself. Accordingly, Fraser cautions against an unquestioning reliance on complex testimony. While it is fairly uncontroversial that we cannot get by without relying on knowledge transfers from other testifiers altogether, it is far less obvious, she suggests, that we ever need to rely on their complex testimony in particular. Perhaps complex testimony and the more radical dependence on other epistemic agents it involves should be treated with considerably more care, or even avoided wherever possible.

Fraser's attention to complex testimony, and to narrative testimony in particular, is certainly a welcome corrective to the peculiar philosophical over-emphasis on single-sentence telling, and her warnings about its distinctive modes of cognitive influence are well taken. However, the truth of the matter is, I shall argue, more complex. For when we turn our attention to speech delivered in ordered discourses encompassing several claims, we also find a place for speaker's increased dependence on her hearer.

To see this, recall the argument I made in the previous chapter concerning a phenomenon I termed *the impossibility of testimonial resistance*. The imposition of testimonial burdens and entitlements is inescapable, I argued, precisely because it occurs too quickly and flexibly for effective prevention or avoidance. There are two sides to this: On the *recipient side*, you might not be quick enough to prevent a determined speaker from completing a proposal of marriage or the disclosure of some intimate secret; given that you are in the same physical environment, and both capable of hearing and seeing each other, it will often be obvious that a message has been received and understood. Equally, however, on the *sender side*, you are yourself at risk from blurting out something you had not intended to say.

But as testimony expands in duration, an unwilling interlocutor gains more opportunities to exit or interrupt the speaker's flow of talk. I can tell an uninterested fellow train traveller *that I am going through a divorce* faster than he can leave the carriage or cover his ears; but I cannot compel him to remain present and attentive for the duration it would take to tell him how and why my marriage broke down. Equally, a speaker who has started to let slip something she did not mean to say has the time to stop and change course if the secret or story is long enough.

In complex testimony, then, the direction of dependence goes both ways: if Fraser is right, in entertaining one what is being told, a recipient of complex testimony is made to rely (more) on the speaker's representational pre-packaging. By contrast, a recipient of simple testimony retains (more of) his representational autonomy. To this extent, complex testimony involves a hearer's increased dependence on the speaker.

However, whereas simple testimony can impart its content in ways that by-pass the recipient's will, complex testimony provides more opportunities to withdraw one's participation and thereby curtail the speaker's effort. For the transmission of complex testimony to be completed, the recipient *at a minimum* needs to remain within earshot for longer and continue to decline to interrupt. To this extent, complex testimony involves a speaker's increased dependence on the hearer.

In fact, once we change our focus from single sentences to extended sequences of speech, we find that hearers typically do much more than simply remain passively present and declining to interrupt. Since Victor Yngve's influential paper "On Getting a Word in Edgewise" (Yngve 1970), linguists have noted a phenomenon labelled *backchannel communication* or simply *backchannelling*: while a speaker is talking, listeners regularly make little responses such as saying 'yes', 'uh-huh', 'oh', or the like. Backchannelling can also be non-verbal: a listener may nod, shake his head, or make facial gestures.

Backchannels allow a listener to signal his continued comprehension and attention to what he is being told. (Bavelas et al., 2000) developed this notion further by proposing a distinction between *generic* and *specific* backchannel communication. Whereas generic backchannel responses (such as nodding, saying 'yeah' or 'uh-huh') simply signal continued comprehension and attention, specific backchannel responses give a more tailored acknowledgement to the content of the speech; for example, a look of concern, raised eyebrows, or "oh no!" or "really?" signal that what has been said is either worrying or surprising.

Because backchannelling can happen simultaneously with the primary speaker's testimony, it does not require her to pause or cede the conversational turn the way a more extended response from the hearer would. A speaker can therefore continue to talk all the while she is monitoring the hearer's uptake through his backchannelling responses.

Backchannel communication is an incredibly robust finding in the empirical communication sciences. An extended recent corpus linguistic study (Reece et al., 2023) found backchannelling to occur at a rate of approximately 1000 words per hour of spoken conversation; 33.7% of speaker contributions elicited a backchanneling response, rising to 65.5 % when the speaker's contribution was five words long or longer[61].

The prevalence of backchannelling and other forms of active lister responsiveness is not merely a theoretical curiosity. Instead, empirical research on communication suggests that they serve a variety of communicative functions without which much of our testimony would be at best impoverished and at worst outright impossible. For instance, the sociologist John Heritage argued that backchannel responses can be *epistemic state markers* – when a listener responds to what a speaker has said with "oh", for instance, he thereby expresses that what the speaker said was news to him; "oh"

[61] This final datapoint it itself suggestive of a difference between simple and complex testimony; the need to express that one is still following along seemingly increases there more opportunity there is to stop following.

signals that knowledge has been acquired[62]. In the next few sections, I shall consider two additional such functions: communicative repair and co-narration.

Communicative repair

Communicative repair names the range of responses and interactional strategies ordinary interlocutors employ to address difficulties like mishearing, misspeaking, or misunderstanding one another; these difficulties are typically labelled as the 'trouble sources' or the 'repairables'. A few examples will help illustrate the idea[63].

1)

L: I read a very interesting story today

M: uhm, what's that.

L: w'll not today, maybe yesterday, aw who knows when, huh, it's called Dragon Stew

2)

Bea: Was last night the first time you met Missiz Kelly?

Marge Met whom?

Bea: Missiz Kelly

Marge: Yes

3)

Caller: … but- hh lately? I have fears a' driving over a bridge. ((pause))

Caller: And uh seems I uh- just can't uh (sit)- if I hevuh haftuh cross a bridge I jus', don't (goan' make-uh- do the) trip at all.

[62] Heritage's work on epistemic state marking function of backchannels is taken up in Jennifer Nagel (2019). Heritage's own later work further develops these insights by proposing a principle that conversation can itself be motivated and driven forwards by a manifest knowledge disparity between the parties over some subject matter: If I say make a claim and you respond with "oh", my next move will be to tell you more

[63] Examples from (Schegloff, 1992; Schegloff et al., 1977). Some transcription details have been omitted for clarity.

Host: Whaddiyuh afraid of.

Caller: I dun' know, see uh

Host: Well I mean waitam'n. What kind of fear izzit. 'R you afraid yer gunnuh drive off the edge? 'R you afraid thet uh yer gonnuh get hit while yer on it?

Caller: Off the edge or something

In each of these cases, a trouble source opens the possibility of a gap in understanding between the communicating parties. The presence of trouble may be caught immediately by the original speaker, as in excerpt 1), or its presence might be signalled by their interlocutor in the next turn, as in 2). In excerpt 3), a conversation between a radio host and caller on the subject of phobias, the trouble source is the host’s question (What are you afraid of?), with caller’s response to the question signalling that it has been misunderstood; the host is initially taken to have challenged the rationality of the caller’s fear of driving over bridges, prompting the caller to try justify further; when the host interrupts, it becomes clear that the question is actually asking simply for the content of the phobia: what is it about driving over bridges that seems scary when you are in the grip of this fear?

However, whether the original speaker or the hearer initiates the pursuit of repair, they will be looking to the other person to establish that the amendment, correction, or clarification subsequently offered has been duly understood and taken on board. Communicative repairs are therefore typically not single conversational turns, but rather little sequences of interaction stretching across a few turns(Schegloff, 1992; Schegloff et al., 1977).

The examples cited above are snippets from recorded and transcribed real conversations collected by Harvey Sacks, Emmanuel Schegloff, and Gail Jefferson, who founded conversation analysis as a field of research in California in the 1960s and ‘70s. An organising principle of conversational analysis is the commitment to studying real occurrences of communication in live interaction, rather than examples of speech or dialogue thought up by researchers, or taken from the literary canon. Imagined speech is liable to contain a number of distortions that set it apart from real speech; as one theorist puts it, our intuitions about language are *strong but wrong* (Enfield, 2017). For one thing, imagined speech tends to be 'cleaned up' and 'clarified' compared to real speech. The former does not contain all the little false starts, the hesitations, and apparently irrelevant tangents that occur in real speech. In real speech, we often struggle for the right word or interpretation before landing on it; both memory and imagination generally filters this out. But removing all of this untidiness also removes much of the communicative repair we perform, and so misleads us into thinking it is rarer than it is.

This challenge is ramified further when we note that the presentation and transmission of research occurs to a large extent by means of written media – articles and books. This presents a difficulty for representing all the relevant facets of instances of real communication in interaction; how do you capture paralinguistic cues like pitch, volume, or intonation when rendering a conversation on the page? What about bodily positioning, gesticulation, or facial expression? Even transcribing only the verbal content of conversation poses a significant methodological challenge.

Altogether, then, when it comes to communication in interaction, and communicative repair as an aspect of it, there is a serious difficulty involved in 'drawing the coverts of the microglot' in J L Austin's phrase (Austin, 1957) and even putting the object of study before one's mind or the minds of one's research community. When we manage to do so, however, we discover the extent to which partners in a conversation are continuously working to sustain mutual understanding and attunement and to signal whether they are still 'on the same page': hearers are giving off backchannel signals that they are still attentive and following along, like saying "mm-hm", "right", making facial or bodily gestures, or giving appropriate verbal responses; for their part, speakers are continuously monitoring these indicators to determine whether something requires clarification or amendment. . If one of them is not doing their part - if a hearer goes passive, or a speaker stops tailoring their contributions and monitoring the audience's uptake, say- mutual understanding will be derailed, and with it whatever other communicative project the parties are engaged in – discussing a story one has recently read, describing one's phobic fears, etc.

Accordingly, a key idea from the socio- linguistics and conversation analysis literature is the notion that mutual understanding should be understood in procedural, rather than substantive terms. What is meant by this distinction is that mutual understanding is not a matter of the parties statically affirming overlapping or coinciding sets of propositions at a time or over a stretch of time, as two folders may contain identical sets of documents from 1 pm to 2 pm. Instead, what makes understanding mutual between us is its aetiology and ongoing status: the fact that it arises from and is sustained by meaning-making procedures within which we collectively participate. This idea is particularly associated with the ethnomethdology of Harold Garfinkel, who writes that 'shared agreement', 'intersubjectivity', or what I am calling mutual understanding:

> "refers to various social methods for accomplishing the member's recognition that something was said-according-to-a-rule and not the demonstrable matching of substantive matters. The appropriate image of a common understanding is therefore an operation rather than a common intersection of overlapping sets". (1967, p. 30)

Being said-according-to-a-rule must be taken in a fairly informal sense. Interlocutors are not holding conversational contributions up against determinate manuals for proper speech, but rather registering inputs that are surprising, discordant, or otherwise ill-fitted to their own understanding of the ongoing activity.

The idea, then, is that we have procedures for collectively generating and collectively interpreting the social world, and it is our participation in those procedures that secures shared understandings and therefore ultimately secures the possibility of a world in common. As Schegloff writes:

> "without systematic provision for a world known and held in common by some collectivity of persons, one has not a misunderstood world, but no conjoint reality at all." (Schegloff, 1992)

Communicative repair is a defence of this kind of mutual understanding (Schegloff, 1992). It allows us to put a derailed conversation back on track, and reestablish the intersubjective attunement that facilitates seamless cooperation on our other projects.

A manifest gap in mutual understanding is not always and unfailingly repaired, however. If the rift is sufficiently small or insignificant, it can sometimes be ignored, reinterpreted, papered over unilaterally, or worked around without causing too much disruption. Or, by contrast, if the initial gap exposes a much wider gulf beneath, either one or both of us might decide to forego the attempt to repair and instead abandon the exchange altogether. The variety of circumstances in which these alternatives are available and/or attractive are instructive to consider.

The choice to gloss over the gap and evade the pursuit of repair can make sense when the gap is sufficiently small not to disrupt the discussion principally at issue between the parties – for instance, suppose that in discussing a visit to a mutual acquaintance C, speaker A gets the name of C's hometown wrong. Both A and B privately realise this, but since the name of C's hometown is not at issue in the story A is telling, neither A nor B bothers to point out or correct the mistake before proceeding in their discussion.

Alternatively, sometimes the gap is significant, but the cost of repairing felt to be greater still. For instance, suppose D cryptically alludes to a difficulty at work in conversation with E without providing more details, and that E does not know what work troubles D is alluding to. If E worries that she *ought* to have known (perhaps she has forgotten, or hasn't been paying proper attention when E has brought it up before?), she might decide it would be too embarrassing to reveal this by asking for clarification. Or perhaps E suspects, but does not know, that the work matter is delicate, and that D has declined to provide details on purpose such that it would be indecorous to press it further. Yet again, perhaps E feels that although understanding the work issue is important, it is more important still not to interrupt D as he is (finally!) opening up to her about difficulties of even greater urgency – a serious health issue, say.

Finally, suppose F and G are talking about a book they have both read over lunch. As F is excitedly trying to explain her great frustration with the lack of character development, she suddenly stops mid-sentence, choking on a piece of radish. It is now clearly of much greater urgency to remove the radish and ensure that F is okay. But even when this is determinably the case, it might seem absurd for either F or G to insist on returning to the topic of character development to clarify F's position, at least for a while, and perhaps for the remainder of the lunch; the matter has been dwarfed and rendered temporarily irrelevant by F's brush with death.

It should be noted that in each set of circumstances I have described, although the avoidance of repair may have been the 'right choice' on balance, failing to repair also has a cost. Even in the first case, where the scope for a gap in understandings between the parties is small and apparently insignificant, unrepaired conversational trouble can have unforeseen future ramifications. For instance, if the name of C's hometown suddenly *does* become a live issue in the conversation (if A and B are arranging to meet at the Starbucks local to C, say) either A or B might feel that it is necessary to backtrack and clarify the name of the town 'on the record', ensuring that they really are on the same page about the location of their appointment. More obviously still, in the second and third cases I considered, an opportunity to reveal something of oneself to the other has been lost, and might be difficult to recreate; perhaps the parties do not get together often; perhaps each will find it awkward to try to return to the topic.

Conversations are dynamic and each temporal stage is the product of a unique combination of factors (environmental, informational, relational, attitudinal, atmospheric, etc) that cannot simply be frozen in place and then switched back on when a more convenient time arises. Between intimates, there is usually more latitude to try to recreate the situation, and a higher chance of success at doing so, but even here, other factors can get in the way. Though E might sincerely want to know about

D's work troubles, for instance, and though D sincerely wants to tell, it can happen that the moment for having the conversation passes[64].

Consider now the circumstances in which repair is not pursued and the conversational project is instead abandoned altogether. In communicative contexts, this is something of a nuclear option. Even if one knows or strongly suspects that one's conversation partner is quite unlikely to be able or willing to cooperate in continuing the shared project (when it seems one's interlocutor cannot hear what is being said, or does not understand one's language; or when they are being intentionally recalcitrant, belligerent, or rude), speakers typically at least try to repair before abandoning the attempt altogether - repeatedly saying "hallo?" a few times into a phone connection that has seemingly gone dead, for instance, or rephrasing one's request to ascertain if the interlocutor really is being rude, or just misunderstood.

In fact, it is extremely difficult to withdraw from conversation without another 'move', and generally requires having managed to evade detection that the message has come through at all. If it is clear to both parties that B has been addressed by A and that B realises this, stonewalling A will itself amount to a kind of response on B's part[65].

Finally, even when interlocutors do choose to repair, it is important to note that they will not always be successful. They parties may try to repair, but misunderstand one another again, and accordingly not manage to coordinate on an accurate understanding of one another.

All this being said, however, it should also be noted that most of the time the pursuit of repair does succeed, and typically so smoothly that it comes and goes entirely unnoticed. Most of the time and in the majority of ordinary situations in which we interact with one another, we are well-attuned to what other people are up to, and both willing and able to meet them halfway if they struggle for the right way to proceed or express themselves. Even a pair of people engaged in heated argument

[64] In conversation analysis, it has been noted that the task of repair grows significantly more difficult with the passage of time (Schegloff, 1992).

[65]See also (Kukla and Lance, 2009; Wanderer, 2010) for a discussion of the vocative function of addressed speech makes neutral passivity unavailable for the hearer.

tacitly abide by norms of cooperativeness such as orderly turn-taking, limiting crosstalk, or moderating one's speaking volume and choice of words in accordance with expectations about what the other will be able to follow and make sense of; indeed, there is little point in arguing with someone at all if one does not try on some level to get one's own point across and understand theirs in turn.

Co-narration

So far, I have noted the prevalence of backchannelling and its use in signalling a need for communicative repair to the partners in a conversation. I also took note of the socio-pragmatic idea that such repair is itself the defence of a collaboratively produced and sustained mutual understanding between the parties. These findings all suggest that verbal communication in interaction is far more interactive and collaborative than philosophers typically suppose and account for.

However, one might suspect that the interactiveness and mutual dependence observed in communicative repair is nonetheless atypical of much of our communication; perhaps speech is significantly more interactive than we typically think, but only *when* and *because* something goes awry. Much of the time, however, things do not go awry, and in those cases speakers really are communicatively autonomous and not dependent upon the collaboration of their hearers. The corpus data already cited belies the notion that repair is exceptional, rather than run of the mill; but it remains an option that communication in interaction *could be* autonomous, even if it rarely is in practice.

Studies on co-narration suggest otherwise, however: they seem to reveal that when hearers are not backchannelling in the expected manner, even apparently monological speech breaks down or degrades in quality.

In a study by Janet Bavelas, Linda Coates, and Trudy Johnson (Bavelas et al., 2000), test subjects were placed in pairs with one member of each pair instructed to tell the other a close-call story – a situation they had experienced, where they narrowly avoided some dangerous outcome. For instance, one participant recounts a story where they fall asleep in bed with a new reading light still on. The light ignites their pillow, but the test participant wakes up in time to put out the flame and suffers no injury. The stories should take a few minutes to tell and have as much detail as possible. In half the pairs (the experimental condition), the listener is given a task to carry out while the story is being told: they are to count the words beginning with the letter ‘t’ in the narrators story, and press a button every time such a word is detected. The narrators are told that their listeners are “listening for something in the narrative”(Bavelas et al., 2000, p. 948), but are not informed of the precise instructions given to the listener, and cannot see the listener pressing the button, which is concealed under a desk between them.

The narratives were video-recorded and scored for quality by analysts who were unaware of the experimental conditions, and of the study hypotheses. For instance, analysts would rate how fluently the narrators described their story’s close-call climax (whether there would be gaps or uneven pacing in the story), and whether they would draw the story to a fitting close or keep talking on and on. The recordings were also be analysed for the occurrence of generic or specific listener responses. A generic response would be something like nodding, saying “yeah” or “um-hm”, whereas a specific response would reflect the specific content of the story, like a look of concern, raised eyebrows, words like “oh no!”, or supplying phrases that fill in the story, as in the excerpt below (Bavelas et al., 2000):

Narrator: I, like an idiot, decide to climb up the cliff instead of…

Listener: …going up the road

Narrator: …taking the easy way out and going up the road

In this excerpt, the listener supplies the phrase ‘going up the road’ which the narrator immediately accepts and incorporates into their narrative.

The study finds that narratives told in the experimental condition were significantly worse; narrators would more often repeat themselves, make sudden changes of pace, pause, trail off in unfinished sentences, use filler words like “”um”, “ah”, or “I mean”, or even comment on the low quality of their own story (“I don’t know how exciting that is”) and supplying further justifications or excuses for the close call.

It was also found that that the experimental condition reduced the occurrence of generic listener responses to 80% of the level found in control narratives, whereas the occurrence of specific listener responses dropped to just 5% of the control. Specific responses also occurred significantly later in the telling of the narrative.

Bavelas and colleagues propose that these two findings are directly connected: when listeners are distracted by another task, they become unable to provide specific responses and so "were not making their contribution to the narrative" and "helping [narrators] moment-by-moment to finish the story smoothly and effectively"(Bavelas et al., 2000, p. 950) the way listeners ordinarily would. This effect becomes self-reinforcing:

> "ordinarily the narrator and listener work together moment by moment to produce a good story. The presence or absence of appropriate listener responses would affect the quality of narration, but the quality of narration would also affect the quality of listener responses. That is, a responsive listener would improve the narrative, which would increase the likelihood that the listener would continue to respond, and so forth. An unresponsive listener would dampen the narrative, which would make the narrative less likely to elicit responses, and so forth". (Bavelas et al., 2000, p. 950)

The findings, Bavelas and colleagues note, oppose orthodox autonomous conceptions of narration, according to which only the narrator's own skill (and the interest of their experience) would matter, and instead support a collaborative model, according to which narrators and listeners exert constant reciprocal influence on one another.

The particularly interesting finding is that what obstructs narrators in this experiment is not the listeners inattention as such; on the contrary, the listeners in the experimental condition are tracking their narrators' speech very carefully indeed[66]. They are just not following or relating to the unfolding story *as a story*. Their (still frequent) generic responses signal that the speaker's individual words and phrases are coming through

[66] In an earlier study, the experimental condition involved a different listener instruction, namely to count the number of public holidays between two date points, but a later iteration changed this to ensure hearers were not simply tuning the speakers out altogether.

the channel clearly enough and are comprehensible as speech; what is missing is markers of the listeners uptake of their narrative significance. Some story parts are expected to be surprising, or funny, or unsettling; when these meta-communicative expectations are not met, it disrupts steady and confident narrative flow.

Naturally, it behoves us to take such experimental findings with a grain of salt, and to bear in mind the limitation of this study and others like it[67]. For instance, qualitative research of this kind usually has small and fairly homogenous participant samples (typically made up of undergraduate students in the psychological sciences), raising questions about the statistical significance and generalisability of the data[68]. A smaller number of test subjects also increases the risk of undetected confounding factors distorting the outcomes; perhaps the study happened by chance to bundle some especially able or especially poor story tellers into the respective test conditions, or happened to set up some particularly well or badly matched pairs. Finally, there is clearly some amount of difficulty involved in setting out a scoring system for narrative quality.

However, even if we find the experimental data unreliable, the notion that narrative quality suffers when listeners are insufficiently engaged also makes a lot of intuitive sense[69]. It is very natural to think that in taking the trouble to tell someone something, a speaker cares not just that each component of the story comes through, but also that the right kind of reaction is produced in the hearer. Story features are often tailored to these purposes – if the story is intended to amuse, some details may be exaggerated for comic effect, where others are skated over or omitted; if it should explain, causal connections are dwelt upon and elucidated with extra diligence. A hearer's neutral responsiveness would signal that these purposes have been frustrated, and it stands

[67] It bears noting that similar findings have been sustained in subsequent studies and analyses, both by Bavelas and colleagues (Bavelas et al., 2002; Bavelas and Gerwing, 2011) and by other researchers, such as (Tolins and Fox Tree, 2014) or (Bertrand et al., 2007)

[68] Bavelas and colleagues consider some of these limitations directly in their discussion, noting for instance how the how the reciprocal influence on narrative quality they hypothesise makes it very difficult to test for the causal influence of listener responsiveness on narrative quality. Since the causal influence is not linear but reciprocal, it violates the assumptions of some forms of statistical analysis.

[69] Anecdotally, I have found that a lot of people upon reflection recognise this phenomenon from their own lives.

to reason that this would matter for speakers: that it would first generate confusion, then encourage doubling back on already communicated parts of the story, and ultimately lead to explicit self-questioning about whether one's story makes sense or fits the brief. Placing oneself in the shoes of a narrator whose listener manifestly pays close attention but does not react as expected, these are the patterns of behaviour one would naturally expect. They are not random forms of spiralling, but the predictable and rational strategies a speaker might employ to regain a common footing with her interlocutor.

Speech in interaction as a collaborative undertaking

So far, I have considered two different contexts in which a speaker depends upon the cooperation of her interlocutor to exercise her communicative agency in testifying what she knows: communicative repair and co-narration. In each of these contexts, a speaker who is trying to get information across to a recipient will be hampered in the attempt if her interlocutor fails to play his part – if he does not scaffold and facilitate the attempt by (at least) indicating whether and when her message is received and understood as she intends and expects.

But a listener can also more actively and constructively shape the direction and content of the conversation. This possibility has already manifested itself in consideration of listener scaffolding through communicative repair or co-narration; in a typical format for repair, a hearer immediately proposes an interpretation of the speech that is the trouble source ("do you mean…?"); the interpretation the hearer offers may provide a better, more accurate or complete statement of the view the speaker meant to put forward. Similarly, in Bavelas and colleagues' study of co-narration, listeners sometimes made 'continuation proposals', filling in the speaker's story for them as in the short segment cited above. Again, in doing so, the hearer may hit upon what is *by the speaker's own lights* a better formulation of what she was trying to convey.

In each of these cases, a hearer's response talk can help a speaker *refine*, *clarify,* or even *discover* and *determine* what she intends to tell. This does not mean the hearer is coopting the role of speaker; while he is responsively scaffolding, moulding, and translating her contributions, leading her to adjust and amend what she would otherwise have taken herself to be telling him, she still has the communicative initiative and the prerogative to decline his proposed interventions and interpretations as inaccurate or irrelevant.

Notice, however, that this interplay may also allow a hearer to sneak something of his own view of the world into what A is saying[70]. Take the following imagined dialogue:

> A: That opening batsman is no good
>
> B: The one who just got dismissed? The tall one with the big nose?
>
> A: Right, Harris.
>
> B: That's Norris.
>
> A: Oh right, Norris then.

Perhaps A has never noticed that Norris has a big nose before, and perhaps she would find it indecorous to bring it up if she did; it could thus not have been a part of what she originally intended to tell. But when B appends it to his clarificatory or elaborative contribution, B would have to take on the cost of side-tracking the primary conversation to disavow it. If A says: "Yes, that is the opener I meant, but I wouldn't say Norris has a big nose, and I don't think you should either", the conversation has changed topic: it is now about noses or norms of politeness, not about opening batsmen. But if she does not disavow it, her next contribution ("right, Norris") will be read as a tacit assent to B's description.

A successful conversational exchange accordingly requires virtuous qualities in both parties. The more capable the hearer, the better he can fulfil his role. He will be more successful if he is attentive, patient, and imaginative, well-attuned to his conversation partner's specific goals and interests, and skilful as a social operator. Equally, if he is unable or unwilling to support the speaker's conversational project, it is within his power to disrupt and derail it. Her attempts to get a position across to him will be

[70] For an exploration of different way of sneaking more content into the conversation, see e.g. Mary Kate McGowan's work on covert exercitives e.g. (McGowan, 2019)

hindered if he is impatient, uncomprehending, or graceless in manoeuvring the conversational territory, or if he forcibly over-imposes his own version of events.

The observed prevalence and importance of backchanneling, communicative repair, and co-narration suggests that talk is much more interactive than philosophers typically suppose. The linguist Herbert Clark evocatively likens John Searle's neglect of the hearer's contribution to communication to a wedding photo cropped too closely, showing only the groom and not the bride (Clark, 1996, p. 137). A good share of the reason for this neglect, I have suggested, is the fact that much philosophical work on communication is implicitly modelled on written communication or on speech that is merely imagined rather than observed in real life; but focussing on the former is liable to systematically mislead. Speech is relevantly different to written communication, as the empirical work I have considered clearly demonstrates. Moran's approach to testimony, alongside the work of other philosophers who have taken note of its normative interpersonal character, arguably begin to redress this deficit, but, I argued, does not ultimately succeed. We can now diagnose some of the reasons why: when philosophers focus on simple testimony to the neglect of complex testimony, they implicitly restrict their attention to the form of testimony where speakers are communicatively and epistemically self-sufficient. But these are in fact the aberrant cases, rather than the typical ones.

In Chapter Three, I proposed that one of the contributions of Moran's project is a reorientation in the philosopher's method of cases in the epistemology of testimony. The relevant paradigm exemplars of telling include not just the emission of a meaningful signal, but also the intention with which it is produced and the understanding with which it is received. Philosophers have missed the epistemically distinctive character of testimony and failed to see how it differs from perceptual confrontation with evidence because they have zoomed in too closely to see all the relevant parts and the connections between them. When we remedy this mistake, we see that a more complex account of responsibility and agency is needed to understand testimony's epistemic significance.

What I am now suggesting is that we can apply the same argumentative move again. The relevant paradigm exemplars of testimony, I have argued, are not those delivered in singular utterances, but those that occur across an extended process of talk in interaction. To get the right phenomenon into view, we need to include not just the intentional transmission of a meaningful signal from one party to another, but also the place of that singular act of speech in a wider interactive sequence. When we do, we see that meaningful communication is actually a joint production in which both interlocutors are giving shape, content, and direction to what is said between them. Once again, a more complex account of responsibility and agency is needed to understand testimony's epistemic significance. Here, I have only made the first nascent efforts in this undertaking.

Conclusion to Part II: Exchange Fetishism and how to avoid it

In her now classic paper on trust, Annette Baier (1986) describes the tendency to interpret basic forms of interpersonal dependence in quasi-contractual terms as 'exchange fetishism'. Contracts, Baier writes, allow us to rely on others "enough for mutually profitable future-involving exchanges, without taking the risks trusters usually do take. They are designed for cooperation between mutually suspicious risk-averse strangers"(Baier, 1986, p. 251), such as well-to-do independent men striking deals in gentlemen's clubs.

Baier's point is that the model of contractual exchange is altogether too formal, too explicit, and too voluntaristic to explain fundamental forms of interpersonal dependency such as, in Baier's example, a small child's trust in its caregivers – ways of relating to others that are in turn necessary for explaining the possibility of the subsequent emergence of forms of trust that are limited, formal, and voluntary like contracts and promises:

> "It is plausible to construe the offer whose acceptance counts as acceptance of a contract or a promise as at least implicitly including an invitation to trust. Part of what makes promises the special thing they are, and the philosophically intriguing thing they are, is that we can at will accept this sort of invitation to trust, whereas in general we cannot trust at will. Promises are puzzling because they seem to have the power, by verbal magic, to initiate real voluntary short-term trusting. They not merely create obligations apparently at the will of the obligated, but they create trust at the will of the truster. They present a very fascinating case of trust and trustworthiness, but one which, because of those very intriguing features, is ill suited to the role of paradigm. Yet in as far as modern moral philosophers have attended at all to the morality of trust, it is trust in parties to an agreement that they have concentrated on, and it is into this very special and artificial mold that they have tried to force other cases of trust, when they notice them at all."(Baier, 1986, p. 245)

Although Baier's criticism is thus launched at broader accounts of interpersonal reliance, and not at accounts of testimonial or epistemic reliance specifically, I want to suggest that her point applies to Moran as well, and indeed that Moran falls victim to a kind of exchange fetishism of his own. His account of testimony, like the contractual conception of trust Baier critiques, imposes the ‘very special and artificial mold’ of explicit agreement – a mold characteristic of promise or contract – upon a phenomenon to which it is ill suited: the interpersonal transmissions of knowledge through testimony.

I think Stanley Cavell has a similar thought in mind when he writes, in the context of critiquing a utilitarian model of promising:

> “But my worry [about the Utilitarian model] is not just that it makes commitments more like explicit promises than they are but that it makes promises more like legal contracts than they are. About these everything Rawls says about offices, defenses, moves, etc., is true; the details of "offer", "acceptance", "consideration", "mis-representation", etc., are elaborately specified, the practice is definitive, and a given conflict can be adjudicated (umpired). This, however, involves a whole way of looking at society, one in which all human relationships are pictured as contractual rather than personal, within which one's commitments, liabilities, responsibilities are from the outset limited, and not total, or at any rate always in the course of being determined. We still relate to one another as persons, but only so far as we stand in certain socially defined roles with respect to one another. The picture is made clearer if we include the suggestion that the central idea underlying the English Law of Contract is that of a bargain.” (Cavell, 1999, p. 299)

That Moran’s model of the interpersonal character of testimony is one on which one’s responsibilities are ‘limited, and not total’ and capable of objective adjudication is clear from his discussion of Raz’s case of a driver, Harry, who insists that he is merely *advising about, and not promising* John a ride tomorrow (Moran, 2018, pp. 131–133). In expecting and intending his words to have the perlocutionary effect that John will indeed rely on him for a lift, Harry

> “undoubtedly incurs a certain responsibility for inducing this reliance, especially should he change his mind and leave John high and dry. But while

> incurring this responsibility, Harry presents himself to John as declining to assume another responsibility, one associated not only with creating reliance in John, but with binding himself and conferring a right on John, a specific right of complaint should he fail to come through with the ride. It is assuming or declining to assume *this* responsibility that is specifically an exercise of his normative powers as a speaker and moral agent."(Moran, 2018, p. 132)

What I want to dispute is not whether it is possible merely to advice without promising, or whether the latter will involve the conferral of a special right of complaint absent in the former; I am prepared to agree with Moran on both these matters. Moreover, Moran rightly notes that whether a promise occurred or not may make no difference to the likelihood of John's complaining if Harry fails to show.

What I want to question is rather whether it is right to draw the line between the two cases, i.e. of advising and of promising, as sharply as Moran apparently wishes to do. In particular, it strikes me as wrong to insist that the legitimacy of John's complaint if Harry fails to show is entirely conditional on Harry's intention. This, I suggest, is precisely to make the speaker's intentional actions *too special*, and to take our other ways of committing ourselves in speech too lightly[71].

My charge that Moran commits the mistake of buying into a kind of *exchange fetishism* of his own thus leads me to question his liability to another form of fetishism, namely what Bernard Williams (Williams, 2002, p. 100-109) described as 'fetishizing' assertion. Williams argues that it would be a mistake to claim that the demands of truthfulness have been observed by a speaker who says something literally true, all the while knowing that their audience will understand them to mean something else,

[71] Just before the passage quoted above, Cavell writes: "Where does the idea come from, which has had currency at least from Hume to J. L. Austin, that promising is so special an act, that the words "I promise" are a sort of ritual of high solemnity? There is nothing sacred about the act of promising which is not sacred about expressing an intention, or any other way of committing oneself. The words "I am going to . . ." or "I will . . ." do not in themselves indicate that you are "merely" expressing an intention and not promising. If it is important to be explicit then you may engage either in the "rituals" of saying "I really want to . . .", "I certainly in- tend, will try to . . .", or the ritual of saying "I promise". It is this importance which makes explicit promises important. But to take them more seriously than that, as the golden path to commitment, is to take our ordinary, non-explicit commitments too lightly."(Cavell, 1999, pp. 298–99)

which is false. Moran addresses this issue directly in his published response to critics, and states that he "fully agree[s]" (Moran, 2019, p. 789) with Williams on this point.

But there can surely be more than one way of fetishising assertion. Though Moran does not say *that the demands of truthfulness have been observed by a speaker who intentionally misleads*, he does promote a conception of speech that puts the speaker's explicit and deliberate assumptions of responsibilities on a pedestal. Even if Moran's account allows that liability to the demands of truthfulness extend beyond intentional assertions to cover e.g. false implicatures, liability to *interpersonal demands* apparently do not; a speaker who has spoken truthfully but also intentionally implicated a falsehood may fairly criticise *herself* for so doing, but her hearer may not – or if he may, his must be a lesser complaint than if he had been lied to directly.

In other words, Moran's account of testimony enforces a conception of its normative structure on which a distinction is drawn between *impersonal* and *interpersonal* demands. *Impersonal demands* may enforced and adjudicated by one's own conscience or perhaps by an all-seeing God[72]. By contrast, *interpersonal demands* may be enforced by one's communicative partners, but these apply *only if* they have been explicitly and voluntarily assumed by the speaker. This, I believe, is precisely to take, as Williams puts it "a rather narrow view of [one's] responsibilities"(Williams, 2002, p. 110), and to make the mistake of characterising the setting of ordinary speech as akin to the "quite special circumstances"(Williams, 2002, p. 109) of commercial activity or the courts of law: "at once adversarial and rule-governed" (Ibid.)

The suggestion I want to make is that philosophical work in both domains – repair and testimony – exhibit symptoms of the same syndrome: a web of conceptual connections that together encourage and enforce certain lines of thinking while precluding or occluding others. The syndrome includes an expectation that

[72] Williams explicitly considers how the distinction between direct lies and implicated falsehood he finds fetishistic has been endorsed by religious thinkers: "God knows what you are asserting even if the hearer does not, be-cause you speak always in the presence of God. He knows what you really assert, because he knows your intentions. But what intentions? Deceit, after all, is a relation between you and your earthly hearer, and the question of what you meant must be answered in terms of intentions directed toward that hearer. God may know my intentions in the sense of my good intentions, but the intentions that form my meanings cannot rest with him, independently of the uptake I aim to secure in the world." (Williams, 2002, p. 104)

interpersonal normativity between intimates will resemble the formal transactions typical of legally regulated commercial activity; that they will centre in particular on transfers of property, penalties for non-fulfilment, or regulation of encroachments on the autonomous domain of another; that the norms applicable between people are the timeless, perspectiveless norms of fairness and freedom from interference; that wherever further duties or entitlements arise, these can always be traced back to a voluntary agreement; and that voluntary agreements are themselves entered into only for the sake of an expectation of personal gain. The parties who populate this moral field are themselves fundamentally self-interested, relatively self-sufficient, and mutually suspicious strangers, seeking to limit mutual exposure without yet totally foregoing the benefits of cooperation; their obligations to one another are from the outset limited and precise, rather than total and diffuse, and can often be best fulfilled by staying out of each other's way; the parties are always interacting as if for the first time, and always both able and willing to walk away at a moment's notice if a sufficiently attractive exchange is not forthcoming.

This last point is, I have come to believe, particularly important: In offering a conception of interpersonal normativity, is *being left alone* the goal or the threat?[73] One can learn a lot about a moral outlook by asking this question. If other people are to fit within one's moral scheme, are they to be held at arm's length, viz. kept at bay and in control to limit their hazardousness? Are they fundamentally seen as threats to oneself – to one's possessions, security, or autonomy - who must earn their proximity through good performance? Or are they *already* close? Already so much a part of one's life that one cannot and would not wish to live without them, *even when* they do badly? Are they so close that insisting on boundary line between their interests and one's own, their projects and one's own, their agency and one's own, sometimes makes little sense?

[73] Much of the legal and political literature that has influenced the bookkeeping model takes for granted that it is the goal to keep the other from crossing the boundary into one's autonomous sphere; but there are equally natural and compelling ways of thinking about interpersonal relations that suggest it could be the threat. See e.g. (Gilligan, 2016, p. 63)

The resources gathered in Chapter Four have sought to develop a different way to think about interpersonal interaction, and particularly about communication in face-to-face encounters. Rather than the execution or adjudication of a contract, I suggest, the mutual dependence and interrelations manifest in talk in interaction better resemble those of the partners in a dance: Though one person may have taken the initiative to start or to lead, both have to willingly participate, and both can add their flourishes and initiate redirections. If one lacks the skills or the inclination, he can unilaterally make it cumbersome for the other to continue, or even bring the activity to a halt. But insofar as choices about how to proceed are being made, they are made spontaneously and continuously and under situational constraints; a partner can suddenly spin one into a position one did not intend or foresee or explicitly agree to; they can make one look ridiculous or gracious by the position they themselves take up with respect to one; and the objective of the activity is often enjoyment or intimacy, rather than personal gain; insofar as personal interests are at stake at all, they are intermingled and coinciding, rather than competitively opposed.

The pursuit of repair, I shall argue, is itself an expression of a continued commitment to the other person and to the idea of a valuable common project that expands beyond the confines of exchange.

Part III

Chapter Five

Repair as the Pursuit of Mutual Understanding

> "The possibilities that exist between two people, or among a group of people, are a kind of alchemy. They are the most interesting thing in life. […] It isn't that to have an honourable relationship with you, I have to understand everything, or tell you everything at once, or that I can know, beforehand, everything I need to tell you. It means that most of the time I am eager, longing for the possibility of telling you. That these possibilities may seem frightening, but not destructive, to me. That I feel strong enough to hear your tentative and groping words. That we both know we are trying, all the time, to extend the possibilities of truth between us. The possibility of life between us."
>
> Adrienne Rich, "Women and Honor: Some notes on Lying"

The thesis of this chapter is as follows: the pursuit of repair is the pursuit of mutual understanding; it is an extended and interactive project of joint inquiry and self-disclosure. What motivates us to repair is not a striving to meet some impersonal standard of justice (retributive, distributive, or otherwise corrective), but rather a desire to understand the other person and be understood by them.

The communicative character of repair is not accidental or dispensable, but essential to the fulfilment of its interpersonal function. Repair, like the bulk of our communicative activity, is cooperative: It is not a matter of one party restoring the status quo ante, but about both of us paving a way forward together.

Because the parties themselves are not static but themselves liable to change and develop, the process of repair has no natural and determinate end point; understandings can shift, new modes of interpretation can become available and gain or lose intelligibility and aptness for the parties. The pursuit can therefore also be creative and productive: the antecedent norms of the relationship may be clarified, renewed, and reaffirmed, but may equally well be replaced by others found to be more adequate.

In developing and defending this proposal, I shall have occasion to revisit several of the ideas introduced in earlier chapters. As such, a brief recap of how we got to this point is in order to situate the next stage of the inquiry.

In chapter two, I proposed that many prominent accounts of repair, despite significant differences between them, follow the contours of the same over-arching model, a model I labelled the Bookkeeping model of repair. According to the Bookkeeping model, recall, the project of repair essentially involves tracing the impacts of a prior act of wrongdoing (whether those impacts concern the victim, the wrongdoer, or the relationship between them)[74], and then identifying responses that can redress and counterbalance each of these. Thus, though some of the authors I discussed explicitly shy away from this language in formulating their positions, I argued that the model centrally invokes the idea of a moral ledger upon which debts incurred through wrongful conduct ought to be matched by subsequent repayment in kind or by compensation of corresponding value.

[74] (Radzik, 2009c) argues that wrongful actions can also leave a normative mark on third parties or on the wider moral community, and so that these relationships too can require repair. In this dissertation, I have mostly set the complication of third-party considerations aside, but it would be relevant to address in more detail in future work.

This model, I noted, has a number of apparent attractions – most notably, it makes the otherwise puzzling remedial potential of moral repair look both familiar and practically tractable by subsuming it to recognisable mechanisms of (re)distributive justice: the difficulty to be addressed is simply a division of illegitimate gains or losses, and the remedy against this is to balance the books by restoring to 'creditor' and 'debitor' what each is rightfully owed by the other. Accordingly, I described the model as taking a *top-down* approach to interpersonal moral repair, viz applying the conceptual tools, methods, and organising commitments of political or legal normativity to the smaller, more intimate scale of interpersonal morality. While admitting that this conception of repair still leaves us a number of complex moral tasks (identifying and typing each incurred cost or gain; accurately 'pricing' them in candidate restitutive actions; etc), this model promises the possibility of precision in moral assessments, clear and repeatable action-prescription for the repairing wrongdoer, and determinacy in the conditions of success.

However, I then argued that the model also confronts some serious and unacknowledged difficulties. Scrutinising three prominent variants of a bookkeeping conception of repair (the Penance view, the Self-denigration view, and the Reassurance view), I argued that the model is either explanatorily untenable, morally objectionable, or both at once.

I also raised doubts about the bookkeeping approach as such. In the best case, it is under-supported: it entrenches undefended normative priorities and substantive values that ought to be subjected to critical scrutiny, both in their own right, and for their fittingness within the interpersonal context; in the worst case, it is normatively suspect and practically self-undermining: it marginalises, distorts, or obscures vital aspects of the phenomena it sets out to explain, and casts repair as meanly fastidious: the preoccupation of a suspicious and self-concerned soul that squints. Repair so conceived, I suggested, has little to recommend it, and is significantly at odds with deeply held commitments to the value of interpersonal relationships, to generosity, and to mutual affection.

The critique of the bookkeeping model I developed in Chapter Two may not be sufficient for persuading its proponents that it ought to be abandoned wholesale. My primary objective in that chapter was simply to elucidate its underlying logic and thereby allow us a clearer-eyed appraisal of the associated costs and challenges. Since the model's organising assumptions are so often adopted by default and without much defence, this should be a worthwhile undertaking in its own right. However, to unseat the dominance of the bookkeeping framework, more is required; to complete my case that adoption of the bookkeeping model is *choice*, and an unattractive one at that, I will also need to demonstrate that we have genuine and preferable alternative available to us. Without this, the rejection of Bookkeeping would leave us falling back on one of the repair-skeptical options identified in the introductory chapter, and accept either that there is no such thing as repair, or that responding to wrongdoing is the prerogative and sole responsibility of the wronged party. Providing such an alternative conception of repair is the task of this chapter.

In doing so, I will also draw from the findings of chapters Three and Four. Across those chapters, I considered the proposal due to Richard Moran that the distinctive interpersonal character of testimony categorically distinguishes it from other sources of information, and explains the priority we afford to people's explicit and voluntary declarations of their attitudes. Telling someone that p, Moran argues, is not like providing indicative evidence for the truth of p, but rather a way of taking responsibility for it, and thereby establishing a normative interpersonal relationship between speaker and hearer. It therefore provides reasons for belief of a distinctive kind – reasons the existence of which depend on the participation and understanding of both parties: the speaker's manifestly intentional self-presentation, and the hearer's reciprocal recognition and uptake thereof. In this, Moran proposes, telling is like promising or entering a contract. The speaker stakes herself and her reputation, and in doing so grants the hearer a suite of normative entitlements – e.g. to justificatory buck passing or to complaint if p turns out to be false.

I argued, however, that Moran's proposal both under- and over-describes the social character of testimony. Adopting with minor revisions Rachel Fraser's distinction between simple and complex testimony, I argued that testimony of the former kind requires less of the hearer than Moran supposes, but also, and more interestingly, that the latter requires a good deal more. When testimony is offered in single self-standing utterances, a speaker can provide disclosures unilaterally. But when testimonial disclosures are offered across a series of interlocking claims (made sequentially, or across multiple conversational turns), as it more typically is, a speaker profoundly depends on her hearer's participatory agency.

Using findings from the empirical conversation sciences, I identified a number of roles a hearer ordinarily plays in scaffolding and facilitating a speaker's communication, including in providing backchannels and pursuing communicative repair and co-narration of a speaker's first personal experiences. A hearer who is unable or unwilling to collaborate with a speaker in these ways curtails that speaker's epistemic and communicative agency, hampering her ability to uncover, refine, and transmit what she knows, feels, or experiences through speech.

These findings seriously challenge the ideals of epistemic autonomy and agential self-sufficiency that have been typical of mainstream epistemology of testimony, and with them the assumption that hearers are passive and fully substitutable receivers of knowledge. These ideas resurface in the present chapter.

The what of repair: Repair as a defence of intersubjectivity

In chapter two, I argued that most philosophers have approached moral repair using what I called a top-down approach. The top-down approach was distinguished by a choice of conceptual framework and tool set, a choice of animating cases, or both elements in combination. In writing about interpersonal moral repair, I noted, many philosophers start from a set of legal or political ideas and concepts: a commitment to fair distribution and equality of resources, retributive justice, and a fundamental prioritisation of autonomy, protection of property, and freedom from the interference of others. They also often start from cases of large-scale and unusually egregious wrongdoing – genocide, violent political repression, or profound and widespread deprivation of basic rights and freedoms. These two starting points (starting with these principles, or starting with these cases) are mutually reinforcing, and jointly recommend a conception of interpersonal moral repair that makes it continuous with social or institutional conceptions of corrective justice: the bookkeeping model of repair.

My alternative explanatory template for moral repair has already been introduced: communicative repair. Communicative repair, I aim to show, provides an apt and illuminating model for its moral counterpart. In taking this notion as my point of departure, I am therefore also making the case for a bottom-up approach to moral repair. Instead of working downwards from theoretically stipulated set of abstract and general legal-political principles of justice, my account starts from the hyper-localised and empirically discovered micro-normativity of in-person interaction. Instead of centring grave and rare cases of transgression, my approach will be to work upwards from transgressions that are so minor and so commonplace that we typically repair them spontaneously and without even realising we are doing it: the everyday slips of the tongue, the small misinterpretations and disruptions to interactional fluency, and the cache of strategies we continuously use to correct and restore.

In introducing communicative repair in the previous chapter, I alluded to Emmanual Schegloff's description of repair as a "defense of intersubjectivity"(Schegloff, 1992) or of what I am here terming 'mutual understanding': whenever two or more people are engaged in an interaction of some kind – passing each other in traffic, carrying out a commercial transaction, getting acquainted, having an argument, etc – they operate with a set of underlying assumptions, expectations, and procedures that together secure the meaningfulness of their stock of available 'moves'. Whether natural or conventional, whether universal or local to a specific context and situation, we display our adherence to these norms of intelligibility by shaping our own contributions to be readily interpretable, usually without giving much, or even any, thought to the matter.

It is this background of shared procedures and resources that enables us to make sense of the behaviour of others around us, act in ways they will understand, and collaborate on chosen projects; the participation in these procedures is both an interpretive and a productive resource. We use them to sustain understanding, but also to develop new forms of meaning between us. As Schegloff memorably puts it:

> “without systematic provision for a world known and held in common by some collectivity of persons, one has not a misunderstood world, but no conjoint reality at all.” (1992, 1296).

Notably, while philosophical accounts of communication and cooperation often include superficially similar notions like mutual knowledge or common ground[75], what sets the socio-pragmatic idea of mutual understanding apart its explicitly dynamic and procedural character. Mutual understanding between two parties is not just a matter of their affirming certain identical or overlapping contents: it does not suffice that each knows (believes, accepts, or expects) what the other knows (believes, accepts, or expects), knows *that the other knows this*, and so on. It is also a matter of these understandings themselves being practically enacted and operationalised in their behaviour over time. We are liable to overlook this procedural aspect of intersubjective understanding when we crop interactive phenomena too closely, looking at just one moment or move while neglecting its place in an extended interactive sequence. Mutual understanding, then, is not just an interpretive resource, but also a generative one: operative understandings are themselves in flux and being collectively created and revised as much as they are being discovered.

We resort to repair, Schegloff proposes, when this ongoing interpersonal attunement breaks down or is at risk of doing so; when errors in execution or interpretation (“trouble sources”) give rise to divergent understandings that destabilise or obstruct whatever specific interactive project the parties are engaged in.

Though I focussed in the last chapter on the use of repair in *conversation*, the underlying phenomenon clearly has a much wider scope. If linguistic communication is a natural starting point for an investigation of repair, this may simply be because so many other forms of interaction *involve* communication, or are themselves incipiently communicative: to coordinate with other drivers on the scene of a road traffic accident, for instance, one uses one’s lights, horn, or arm movements to express intentions or understanding in the place of words; in making a purchase, placing a product on the

[75] E.g. in (Lewis, 1969) or (Stalnaker, 2014)

counter and waving one's credit card can take the place of saying "I'd like to buy this pint of milk and pay by card", and so on. However, in either case, it can happen that either party's conventional non-linguistic signalling is not understood, and the coordinated activity is at risk of derailment; in these cases, we generally resort to linguistic resources to clarify and elaborate: we identify the locus of intelligibility failures, and set out to reestablish a common footing. Communicative repair, I noted, occurs with remarkable regularity in ordinary talk; though we frequently slip up or confront minor failures of alignment, we are also typically so well-attuned to our interactive partners and activities that we quickly understand what and how to repair.

My proposal in this chapter, then, is as follows: moral repair is of a piece with communicative repair. It is a response to the same kind of problem, viz. a manifest lack of, gap in, or threat to mutual understanding; and it is remedied by a similar process: collectively identifying the locus of the failure of alignment and setting out to reestablishing a common footing capable of sustaining future engagement.

Notice how easily this proposal explains why, and when, the pursuit of moral repair would matter to us. Our commitment to moral repair, on this picture, is simply an extension of a fundamental concern to preserve the reciprocal intelligibility that affords us a shared and shareable social world, and which enables us to coordinate and cooperate on specific projects within it. If two people cannot make themselves understood to one another, all but the most self-serving and antagonistic forms of interaction will be brought to a halt[76]; even arguing, scolding, threatening, and insulting lose their point if the other cannot understand; even wars and battles have common codes for combatants to avail themselves of, e.g. to indicate surrender, or suggest one's strategic posture to the enemy[77].

This account of the motivation for repair may immediately seem over-extensive; in the introductory chapter, I noted that not all cases of moral wrongdoing lead to pursuits of repair; some wrongful actions go unaddressed, and sometimes this is by choice. But

[76] Only in a fight to death – when the parties have explicitly abandoned all interest in a shared world – will there be no scope for an interest in understanding the other party and making oneself understood.
[77] (Schelling, 1980)

if wrongdoing is a breach of intersubjective understanding, and intersubjective understanding sustains a world in common, would we not always and unfailingly pursue repair?

But we also need to appreciate that not all gaps in intersubjectivity are equally threatening, that repair is not our only mode of response to them, and that the reparative pursuit of mutual understanding requires different investments of effort in different contexts. As discussed in the prior chapter, some gaps to understanding are small, and can be evaded or ignored with little difficulty; others can be unilaterally reinterpreted, or set aside for a more opportune moment. Still other cases suggest one is better off jettisoning a common project with the other party altogether: some gaps are so vast that closing them would be a disproportionately demanding undertaking; not all interactive projects or partnerships are equally worth retaining.

Similar dynamics can arise, I suggest, in the case of moral transgressions. A small violation of normative expectations can be ignored, set aside, or privately given a more favourable reading ("he probably didn't mean it like that", "she must have thought I was only joking", "it was a bit insensitive, but it's not worth a big fuss" etc). By contrast, a manifestly major gap in understanding might signal such lack of concern or fundamental divergence in our values that mending the relationship is too large a task to be worth our while, and instead motivate one or both of us to avoid the other person immediately and for good.

I shortly will say more about how moral repair is actually pursued in practice, but the comparison with communicative repair should already highlight its collaborative character. If repair (communicative *and* moral) is an interactive procedure in which the parties need to coordinate on a trouble resource and invest some effort in resolving their interpretive differences, then the decision to repair will also depend on one's assessment of the other party: some partners in repair are more able, and some relationships more valuable to retain.

In moral as in communicative repair, then, whether repair is pursued or not is a product of a few different factors: how disruptive is the divergence in understanding to the common project? What degree of cooperation or recalcitrance can be expected from the other party? And how valuable is the shared project and the relationship one is engaged in?

Whether a conflict occurs *outside* or *within* an existing relationship (platonic, romantic, familial, or otherwise) generally shifts our answers to all these questions; we are more likely to discover a need for or interest in pursuing repair in cases where the parties are committed to one another and to maintaining, if only for a short while longer, a relationship: between friends, lovers, family members, co-workers, or others who desire or require the continuation of a stable partnership of some kind. In addition, though intimates generally have more robust mutual understandings in place, they also have far more occasions for significant *mis*understanding, and will generally be more motivated to avoid the risk of a deepening conflict.

By contrast, in one-off interactions with strangers – in moving through the morning traffic, say - I can much more easily afford to handle a wrong done to me with *either* incivility, magnanimity, or indifference; since I am unlikely to knowingly encounter the same person ever again, my behaviour on this occasion is not setting expectations I will be held to in the future, and it ultimately matters relatively little whether the incident is repaired or not[78].

These considerations provide a valuable lesson: Wherever we find a reserve of common interests between the disputing parties, however small, we find a motivation to retain some measure of mutual understanding and ipso facto a motivation to pursue repair when we encounter a conflict. This interest may initially be one-sided; whether you are the wrongdoer or the wronged, your initiative to pursue repair could be the sign of good will that sparks my willingness to collaborate in working through our differences. By contrast, if I show myself recalcitrantly unwilling to reconcile, your initial desire for repair may be extinguished. Our respective interests in repair can also increase or decrease as the process unfolds; but if one of both of us remains indifferent to the other, or would rather avoid any further interaction, the attempt to repair cannot get off the ground.

[78] Goffman notes the willingness to accept responsibility for small infractions in (Goffman, 2009)

Some philosophers writing on repair and forgiveness find this an unacceptable consequence; a remorseful wrongdoer may find herself unable to discharge her responsibility to set things right if the victim wants nothing to do with her. If one is thinking of repair as the repayment of a debt, this will indeed seem unfair or even unintelligible; the wrongdoer's steadfast refusal would make the deficit effectively unpayable, leaving one forever in the other's debts. But I have sought to argue that these are not the only terms in which to see the demands of interpersonal normativity; we sometimes do incur duties and responsibilities without actively seeking or willing it; and requiring other people's participation and support to discharge them is not aberrant but typical. Moreover, the notion that we can sometimes find ourselves practically unable to repair strikes me as much more accurate to our ordinary experiences – a point I return to later in this chapter.

In short, the idea that moral repair is of a piece with communicative repair connects the motivation to pursue repair with a more foundational interest in retaining mutual intelligibility, but also explains when this interest can itself be outweighed or set aside - when the rupture is too small to be disruptive or too vast to be worth the effort, for instance. Moreover, since restoring mutual understanding requires effort and collaboration, prospects for and interest in its success also vary with attachment to the other party and with our expectations of their participation.

Notice that these variations in whether and when repair is sought are much harder to explain on a bookkeeping conception of repair. Firstly, if repair is conceived as a species of justice (distributive, retributive, otherwise corrective), then the avoidance of repair amounts to justice delayed or justice denied – a consequence that should surely be avoided wherever possible. Whether one has a personal relationship with the other party should in itself matter little. Moreover, because the bookkeeping approach frames repair as a task for the wrongdoer, rather than as a collaborative undertaking requiring the willing participation of both, it becomes much harder to understand why any victim of wrong would willingly forego it; why would you not want to receive what is duly yours?

Wrongdoing as an epistemic challenge

So far, I have introduced the idea that moral repair is of a piece with communicative repair: Both phenomena, I proposed, arise when mutual understanding is under threat or facing breakdown; and in both cases, the objective of repair is to reestablish the parties' reciprocal understanding of one another, allowing them to regain a common footing.

This conception of the problem of repair, I suggested, helps us to understand why repair matters to us, and why the motivation to pursue it (whether as the wronged or as the wrongdoer) can vary significantly with context, and not least with one's expectations of and investment in the other person.

The picture so far introduced may immediately give rise to the following line of objection: Perhaps it should be granted that moral wrongdoing *can* involve a failure of mutual understanding and intelligibility; but the instances where this is true are fringe rather than standard.

For instance, it might be granted that "a failure of understanding" is an apt description of a case wherein one person acts in a way that is hurtful to another, and does not realise this, but ought to have known. In cases of *unwitting but negligent wrongdoing*, the wrongdoer culpably fails to understand what is required of her; perhaps she is operating with an outdated or alien moral code which she could and should have revised, but has not. Until the nature of the situation has been explained to her, she will not see why the victim finds her behaviour to be upsetting. Though she has plenty of good will and concern for him, she fails to appreciate the true significance of her own action.

In cases of this kind, it may be agreed, the source of the issue is a set of diverging normative understandings which need remedying. However, the objection goes, when it comes to wrongdoing, such cases are the exception and not the rule. In the typical case, the problem is the wrongdoer's quality of will, and failures of understanding are not at issue. Moreover, even when a failure of understanding *is* at issue, e.g. in cases of unwitting but negligent wrongdoing, the failure of understanding itself is not the whole or the heart of the problem – it is the hurtful action, and perhaps the negligent ignorance that makes her blameworthy and which repair must address. The relevant norms a wrongdoer violates are not mere intelligibility-norms, but norms of moral acceptability.

In short, then, the objection presently under consideration holds that the conception of the problem and pursuit of repair I have offered at best applies to only a subset of the relevant cases; and even where it *does* apply, its characterisation of the animating issues is shallow and incomplete: Wrongdoing is not in general about a failure of interpersonal understanding, and accordingly, neither is repair.

This line of objection can be resisted, however, and doing so will help expand and clarify my proposal. Firstly, it is true that moral wrongdoing involves the violation of a variety of more specific norms – norms of honesty, of care and concern, of reciprocity, or whatever else may apply in the concrete situation. But when a violation of one of these more specific norms gives rise to a pursuit of repair, its violation is *also* a failure of mutual understanding.

Conformity to moral and interpersonal norms are a part of our ordinary expectations of other people. When we are wronged, we do not just react with frustrated resignation as we would when a calculated gamble does not go our way; we also feel shock and surprise, at least momentarily. Wrongdoing is confusing, disorienting, and destabilizing, particularly when it occurs within an intimate relationship. It raises doubts and questions that often manifest themselves in the paradigmatic ways we express our blame to its target. We say things like: "How could you?", "What were you thinking?" or "Do you have any idea how that made me feel?"

We should refuse to treat such questioning as purely rhetorical, and instead take it at face value[79]. Being wronged raises questions because it suggests or reveals that we are not in the kind of situation we thought; the other party is not abiding by the rules of conduct we believed were mutually accepted, or are not interpreting those as we are,

[79] I take the idea of refusing to treat questioning as rhetorical from the literary scholar and cultural theorist Lauren Berlant. See e.g. https://humanities.uchicago.edu/articles/2019/11/why-chasing-good-life-holding-us-back-lauren-berlant

or are not giving the same weight to the stakes and the risks as us. Wrongdoing is epistemically unsettling and calls for a kind of inquiry to help us establish where we have diverged and why. To proceed, we need to understand the other person, and we need them to understand us.

It is easy to underestimate how significant an undertaking it typically is to understand and come to terms with wrongdoing. Much of the philosophical literature on blame and repair is set up in ways that sidestep these questions. We are typically given cases to consider where it is artificially clear who is at fault and why: The wrongdoer has carelessly broken a precious vase, maliciously stepped on one's gouty toes, wilfully breached a promissory obligation, or failed to return a borrowed book. The brevity and concision of such God's-eye descriptions obscures and misrepresents much of the experience of the affected parties themselves.

As participants on the ground, we confront the situation with imperfect information. When we ask "What happened?" or "Why did you do that?", we genuinely want to know, and expect a real answer. In many real cases, it is not immediately obvious that the other's behaviour *was* indeed careless, malicious, or wilful, and even when it *is* obvious, that hardly closes the inquiry; if the other person is someone with whom I wish or need to maintain a relationship, I need to understand what lead them to adopt such an attitude to me, and I need them to understand why and how I find it unacceptable. Does the other person accept the same characterisation of their action, or do they continue to see it differently, and if so, why?[80] When we rush past this state of confusion and inquiry and stipulate a specific description of the wrongful action as accurate and mutually accepted, we bypass much of the work of repair[81].

The objector might persist as follows: surely it is possible be unsurprised by wrongdoing. Sometimes a victim *does* react with dejected resignation, or with a

[80] Though I focus here on the epistemically unsettling situation of a *victim* of wrongdoing, it is worth emphasising that it can be at least as destabilising to be its agent. Raimond Gaita describes the "lucid remorse"(Gaita, 2004, p. xxi) that can attend the horrible first personal question "My God what have I done?", as well as the difficulty many moral theories have in giving adequate sense to the distress it expresses. My thanks to Miranda Fricker for pressing me to include this point.

[81] For additional criticism of the tendency to assume a background of agreement on the moral norms between disputing parties, see also (Dover, 2019)

scoffing satisfaction: “I knew this would happen!” I grant that such cases do occur, but notice that these, too, open more avenues for questioning than they close: if you really felt certain the wrongdoer would injure or hurt you, why did you put yourself at their mercy as you did?

Sometimes we suspect that someone is liable to transgress again, but nonetheless adopt a resolution to treat them otherwise, hoping against hope that our well-founded pessimism will finally be disproven. In doing so, we are continuing to see the other person as an agent, capable of change and development. If we hold this attitude, our willingness and interest in questioning the other will persist, even if it has to sit ambivalently alongside a fear that we are simply being foolish or naïve.

By contrast, if all hope for the other person’s improvement has been lost, we are seeing them as someone with a psychological compulsion, mechanistically determined to fail us again and again. In such cases, we might truly say “I knew it, I predicted this all along!” But when this happens, we do not see a point in pursuing repair – we are responding to the wrongdoer with what Strawson memorably termed the objective attitude; she is someone to be managed and controlled, perhaps avoided or treated, but not someone with whom we can have a full relationship. As such, my proposal stands: Whenever repair is pursued, we find ourselves with an epistemic challenge of sorts: a lack of, gap in, or threat to mutual understanding.

This, then, should address the first part of the objector’s challenge, viz. that my proposal applies only to a subset of the relevant cases. To respond to the second part of the challenge, viz. that reestablishing understanding is not itself adequate for repair, we shall first have to take a closer look at the *how* of repair, and its connection to exculpation and condonation.

The *how* of moral repair and the role of narrative

So far, I have outlined and defended my starting conception of the problem of repair; however, more still needs to be said about how the pursuit of it proceeds in practice. If moral repair is a response to a lack of, gap in, or threat to mutual understanding, what do the parties do to reestablish it?

In considering this question, we encounter a couple of points of divergence between communicative and moral repair that reveal an explanatory gap that needs to be filled. Firstly, moral repair responds to a significantly different array of trouble sources. In the communicative case, what is repaired is always situated within the communicative interaction itself – what someone said, or did not say; what was heard or misheard; or the variety of para-linguistic and extra-linguistic cues that stand in for linguistic communication, such as gestures, facial expressions, intonation, or body language.

By contrast, in cases of moral repair the trouble is not always something that was said, heard, or expressed between the parties themselves; indeed, the occasion for moral repair can just as easily be some action or inaction undertaken behind the other's back, which one never intended or expected them to find out about – learning that a friend stayed silent when others badmouthed me at a party, say. Though I have argued that such violations of normative expectations are also failures of mutual understanding, the fact that my friend's conspicuous silence was not *directed to me* or *intentionally undertaken in my presence* nonetheless affects what redressing them involves. The expressive significance of my friend's action is left open in ways it would not have been if it were backed up by a reflexive communicative intention along the lines envisioned e.g. by H. P. Grice.

Secondly, and relatedly, whereas communicative repair happens either straightaway or not at all, moral repair often occurs much further in time from the occurrence of the trouble source. Researchers have argued that communicative repair rarely occurs more than three or four conversational turns after the trouble source it addresses[82]. A plausible reason for this might be that our working memory for conversation is relatively short; it is difficult to *replay* how a conversation unfolded in your mind even when you are strongly motivated to do so. We generally remember just the gist, and not the detail of what was said. This makes it much harder to 'redo' or 'revise' a troubled conversational move the further into the future you advance; securing the referent for 'what I meant by that' becomes harder.

[82] (Schegloff, 1992)

Though a non-trivial amount of moral repair happens with similar immediacy (saying sorry for bumping into people, etc), plenty of common cases involve a greater temporal remove from the instigating events. Even when a violation of normative expectations occurs in interaction between two parties, it could be hours, days, weeks, or even years before they take up the task of repairing it.

These differences – that wrongdoing is not always *intentional expression in interaction*, and that it is often repaired *at a temporal distance* – means our stock of common strategies for communicative repair are not up to the task of moral repair on their own; we cannot just say "huh?" or "what did you mean by that?" as we can and do in the flow of talk. In cases of moral repair, we need more stage-setting and context to coordinate on a mutually recognised trouble source for repair to address, and we have more work to do in uncovering the meaning of troublesome behaviour in the first place.

These, I propose, are themselves communicative tasks. So much is required, and achieved, simply by putting the wrongdoing into words, or *coming to terms* with it. Doing so requires me to formulate, at least provisionally, an interpretation of the significance your action has for me ("I saw that you …. and that was a really hurtful, because…"; "it makes me so angry that you would do that without…"). In doing this, at least two things happen: firstly, one gives a certain kind of shape to the experience, crystallising it in a specific form; secondly, one makes a portion of one's experience available for the inspection and consideration of the other[83].

Several theorists have remarked on the importance of public acknowledgement and truth-telling in the context of serious wrongdoing[84]. Practices like truth-commissions, victim impact statements, and oral histories reflect the notion that those who have suffered grave harms and borne witness to wrongdoing should be given a platform to speak directly to those who have committed wrongdoing or colluded in it. In these contexts, the first-personal accounts of the affected parties are not just one source of information among others; even when the historical record of facts and events is

[83] A point echoed in (Dover, 2019):

[84] (Govier, 1999; Walker, 2013, 2006a)

widely known, personal testimony gives a voice to the downtrodden and affords a distinctive kind of acknowledgement of their experiences[85]. Trudy Govier characterises acknowledgement as the "marking or spelling out or admitting" as significant that which is known (Govier, 1999); in acknowledging, one attends to some reality, as opposed to ignoring it, or allowing oneself or others to be deceived with respect to it, and places is in the public cognitive scene.

When the reality to which one is giving voice is complex, the account one gives of it will be as well. In the previous chapter, I noted that much of our testimony is offered not in individual, self-standing utterances, but in extended, ordered sequences, including personal narrative. Accounts of wrongdoing, too, typically involve narrative. Narrative adds representational structure and formatting, encouraging or enabling an audience to inhabit the narrator's experiential perspective[86]. But it also has morally significant functions of its own.

Many theorists of narrative have argued that placing an unexpected or unsettling occurrence into a story sequence helps reestablish its intelligibility. The psychologist Jerome Bruner writes that narratives are "our armamentarium for dealing with surprise"(Bruner, 2003, p. 29):

> "Narrative is a recounting of human plans gone off the track, expectations gone awry. It is a way to domesticate human error and surprise. It conventionalizes the common forms of human mishap into genres—comedy, tragedy, romance, irony, or whatever format may lessen the sting of our fortuity. Stories reassert a kind of conventional wisdom about what can be expected, even (or especially) what can be expected to go wrong and what might be done to restore or cope with the situation. Narrative achieves these prodigies not only because of its structure per se but because of its flexibility or malleability."

[85] On the idea and the importance of voice, see (Walker 1997), esp. chapters 5 and 9.

[86] (Fraser, 2021; Pettigrove, 2007)

Elizabeth Camp makes a similar proposal, arguing that storytelling supplies us with "wayfinding instructions"[87] for navigating life, particularly in confrontation with the unfamiliar.

Notice in this description an echo of sorts of a point I made earlier about the explanatory power of adopting a bookkeeping framework. The bookkeeping framework, I argued, seeks to domesticate the unfamiliar into a recognisable format, and to subsume one's own particular experience into a general pattern; a specific incident becomes a loss or debt of *this* type or *that* type, and *this* or *that* magnitude, and this categorisation gives us a set of expectations about how it might be offset or counteracted. The use of narrative can fulfil a similar categorising and familiarising role, but with considerably more flexibility and detail. By placing them in narrative structures, we can categorise the effects of wrongdoing as losses and debts, but also as misfortunes of other kinds: breaches of expectation; disappointed or misplaced trust; hurt feelings; falling victim to illness and disease; losing one's sense of reality, or one's way, or oneself. In capturing the impacts of wrongdoing in narrative form, we circumvent need to convert and reduce each of them to a concrete weight or currency. In addition, whereas the bookkeeping model 'types' transgressions in terms of an end state or bottom-line, a narrative contextualises it by placing it in a story sequence. Narrative allows wrongdoing to wear (some of) its causal antecedents and background enabling conditions on its sleeve, where the bookkeeping model conceals or obscures them.

These communicative resources – putting one's experience into words to give it shape and make it sharable; reclaiming a voice and giving public acknowledgement to difficult realities; narrating wrongdoing and its effects and placing these in a story sequence – allow us to bridge the temporal and expressive gaps that distinguish moral from communicative repair.

As I argued in the previous chapter, these are collaborative activities, not unilateral achievements by the speaker. Even efforts to narrate a personal experience fall flat if

[87] Unpublished manuscript, forthcoming as "Stories and Selves: A Twisted Love Story about the Meaning of Life" with *the Royal Institute of Philosophy Supplement*.

one's interlocutor does not scaffold and facilitate with participatory actions and interventions. Moreover, as a central 'character' in the story plot, the other person may challenge one's version of events by supplying their own ("that is not how it happened", "I couldn't possibly have known you would see it that way, I only meant to…"). Indeed, if their understanding is to be mutual and practically enacted, neither wrongdoer nor victim can remain merely an audience for the other's self-disclosures: each must also supply their own.

In her discussion of acknowledgement's opposites, Govier notes how both self-deception and the intentional ignoring of unpleasant realities can be assisted by the collusion or complicity of others willing to join in obscuring or turning away from the truth. In communicating about a past wrongdoing, each party's participation serves as a check on the other's blind spots, biases, and impetuous rushes to judgement.

Notice that this process leaves a significant amount of scope for contestation and negotiation. If things go well, their respective stories are not competing master narratives jostling for dominance, but complementary perspectives that can be fitted together into a sharable whole that is adequate and honest to them both. Moreover, in hearing each other's stories, the parties should be neither overly skeptical nor excessively deferential; taking me seriously as a story-teller means expecting my competency, but not my incorrigibility. Relevant challenges to my initial reconstructions of events should be welcome where they help me see things in a more satisfying light. The norms of the relationship are themselves liable to alteration in this process.

Success in this activity requires various virtuous qualities – patience, attentiveness, imaginativeness – but also real skill. Some of us are more naturally adept at story-telling, and some of us have had more experience to hone our abilities. Appreciating these differences helps us to understand why the pursuit of repair is sometimes so easy, even when the infraction is grave, and at other times impossibly hard although it was not. Even sincerely committed parties can be hampered in their pursuit of repair if they simply cannot find sharable terms and understandings that allow them to coordinate on a story. Lacking hermeneutical resources, then, can be an obstacle to repair[88].

Exculpation, Condonation, and the threat of collapse

To this point, I have presented the contours of my account of repair as the pursuit of mutual understanding. Like communicative repair, I argued, moral repair is a response to a lack of, gap in, or threat to intersubjective intelligibility and attunement that threatens the cooperative project between two people. The problem posed by violations of normative expectations is accordingly in part an epistemological one: the parties need to understand one another and to feel understood. The resources that facilitate this are themselves communicative practices and therefore require their interaction and collaboration. The parties are not just verbalising their own side of the story in each other's presence, but actively relying on their interlocutor's communicative scaffolding and facilitation to arrive at understandings that are accurate and acceptable to them both and capable of being shared between them.

At this stage, I want to consider an extended challenge to the proposal thus far developed. According to the challenge, the project of jointly reconstructing mutual understanding may be morally valuable in either of two ways; by bringing new facts of the case to light, or by providing prudential or pragmatic grounds to set the matter aside. In neither case, however, does it provide repair.

First, a fuller understanding of the situation may bring to light some relevant facts of the matter that cast the agents or their actions in a new light, and thereby undermine the notion that there was anything to repair in the first place. For instance, perhaps it is revealed that the wrongdoer acted in ignorance, or under duress, or that they are deficient in some capacity that diminishes their ability to participate in repair. In this case, however, understanding has not provided reparation, but *exculpation*[89]; though the action was wrong, the agent is not blameworthy or accountable for its repair.

[88] For the idea of hermeneutical resources, see (Fricker, 2007)

[89] Of course, a closer inspection of the situation could also reveal that the action was *justified,* and so not wrongful at all. There is a lively debate in philosophy about which considerations can justify, which excuse, and what makes it so. For present purposes, since my interest is in responses to

Alternatively, if the pursuit of mutual understanding has not hit upon any such appropriateness-undermining facts, it may instead have helped the parties to contextualise the wrongful action, or somehow made it easier to bear – perhaps by reminding them that failure is human, that the wrongdoer has important merits that outweigh her flaws, or that they share a prudential interest in putting the matter behind them. In this case, what takes the place of reparation is a kind of *condonation* or *wilful redirection of attention*: resolving, and becoming psychologically able, to treat an unjustified, unexcused, and unrepaired act of wrongdoing as something one can simply set aside.

Each of these functions have their proper place in the aftermath of wrongdoing, the objector will say, and we can therefore agree that the pursuit of mutual understanding is often a good thing. However, in neither of its roles does the pursuit of mutual understanding supply *repair*. Moreover, an account that collapses repair to exculpation or condonation commits a serious mistake. Adopting my account would have us cultivating the search for excuses in far too many cases, and to settle for condonation whenever that search is unsuccessful.

Why would this latter be a bad outcome? Clearly it is sometimes (perhaps even often) the case that our initial assessments of a situation casts it in an unduly negative light, and that gaining more insight and understanding of the circumstances can help us see that there is less to overcome, or even that the person who committed the transgression really has nothing to apologise for[90]. But, says the objector, it is equally clear that this is not *always* the case, indeed that it must not be: if *all* wrongful conduct is ultimately to be excused, then too little remains of our responsibility practices as we know them.

This line of the objector's worry could be cached out in terms of what we might call epistemic anxieties about moral responsibility and accountability for wrongdoing.

wrongdoing, I am setting justification aside, and with it these difficult questions about its difference from exculpation. For another account of the ways in which a pursuit of understanding can supply or promote either justification, exculpation, or forgiveness, see (Pettigrove, 2007)

[90] (Pettigrove, 2007) makes a number of interesting proposals along these lines, including in suggesting that understanding can mitigate the apparent wrongfulness of someone's conduct, and compel one to retreat from the least charitable interpretation of them to one that is more generous.

The underlying concern is that our initial ascriptions of responsibility for some action or omission can always be undermined by discovering more information about the agent and how they came to act as they did. Arguments about the significance of causal determinism to moral responsibility arguably fit into this structure (Pereboom, 2011); however, even some who profess to being unmoved by this kind of scepticism about responsibility exhibit concern about what a more complete psychological science might uncover about the sources of our actions, attitudes, and patterns of response (Strawson, 1962; Watson, 1993). Yet another version concerns the role of personal histories ((Watson, 1993; Wolf, 1990) but see also (Ebels-Duggan, 2013)). To allow repair to collapse into exculpation, holds the objector, is to open oneself to a slippery slope that will threaten to swallow up, undermine, or replace the very accountability practices we are trying to explain and vindicate.

Equally, some amount of *condonation* is perhaps necessary for living with others or even with oneself; it is plausibly better to condone wrongdoing once in a while than to remain forever burdened and bitter by what has been. But though it may be the lesser of two evils, condonation of wrongdoing is always regrettable. Kolnai (Kolnai, 1974) accordingly warns against too lax an attitude towards condonation in strong terms: it is "conniving", "immoralistic", "unfair", "undignified and self-soiling", "spineless accompliceship", and "submissive meekness before evil" (Ibid). In direct opposition to true repair and forgiveness, which involves the *eradication* of wrongdoing, condonation risks fostering it.

This, then, is the upshot of the challenge: the pursuit of mutual understanding risks either encouraging, enabling, or even pressuring the parties into overextending exculpation, displacing the scope for repair altogether; or else encouraging, enabling, or even pressuring the parties into treating the matter between them as resolved when it is not, providing at best the illusion or repair as a cover for what is really widespread condonation of wrongdoing. This is worrisome not just because it gives a pass to wrongful action and thereby encourages more wrongdoing in the future, *but also* because it taints the parties as insufficiently disciplined or discerning, or as enthralled to bad faith. The victim is not taking himself or the moral norms applicable to him sufficiently seriously; the wrongdoer is manipulatively or self-servingly obfuscating, spinning a story to evade the responsibilities that are appropriately hers to discharge; both are allowing attention to be diverted away from the moral heart of the matter and the real work of repairing.

Like the critiques I raised to the bookkeeping model of repair in the second chapter, this extended challenge strikes at both the explanatory adequacy and the ethical viability of the proposal I have put forward. The explanatory face of the challenge holds that my conception of repair supresses or undermines an explanatorily important distinction between repair and relevant cognate notions; the ethical face holds that this collapse of distinctions has morally onerous consequences and encourages vicious attitudes and responses to wrongdoing.

But the challenge can be met. First, we should not accept the notion that the discovery of exculpatory considerations invalidates responsibility for wrongdoing altogether; more plausibly, it moderates it, or affects how we ought to interact with the wrongdoer in pursuing repair (Mason, 2019; Sliwa, 2019, 2010). Accordingly, the challenge that my account collapses a morally and explanatorily important distinction quickly loses a good deal of its alarmist sting. If having an excuse does not remove the conceptual foothold for responsibility and repair altogether, but rather shrinks or shifts it, we need not be so concerned about drawing and sustaining a line between them. But secondly, and more importantly, we should also question what motivates us to care about the line-drawing exercise at all.

It may initially sound strange to raise this question. The notion that there ought to be a clear boundary line between repairing on the one hand and excusing on the other is almost axiomatic in the literature on apology – both within philosophy, within other academic disciplines, and even in therapeutic and self-help literatures on the subject[91]. But despite the reverence with which the distinction is commonly treated, it is, I suggest, far from obvious that we generally observe or enforce it as participants in moral conflicts. In many of our everyday pursuits of repair – those pedestrian cases

[91] For examples from philosophy, see e.g. (Hieronymi, 2001, p. 530; Hughes and Warmke, 2024); for self-help, see e.g. (Lerner, 2017, pp. 68–69).

that are so common they generally fly below the radar of conscious attention, reflection, and effort – 'accounts' and 'apologies' often occur not only in tandem but profoundly entangled with one another (Austin, 1957; Goffman, 2009). This puts some pressure on my opponent's claim that marking this distinction is theoretically important, and that a failure to do so is normatively damaging.

Confronted with this response, my opponent might entrench her critique: if indeed it is correct that accounts and apologies often mix 'in the wild', particularly when we are not watching closely or reasoning carefully, we have all the more reason to insist upon their separation in principle. A more clear-eyed theoretical appraisal will help to train and discipline our unruly (and, it would be argued, *immoral*) instinctual responses. Further support may be added to my opponent's position on this point by noting that even in the wild, we sometimes *do* scrutinise and challenge our intuitive responses more closely, namely when the moral stakes are higher and/or the transgression more important to us; and when we do, the line between excuse and exculpation on the one hand and genuine apology and repair on the other is one we evidently do care about. We should learn from *these* cases, my opponent would hold, rather than the commonplace and more informal ones.

In keeping with my bottom-up explanatory strategy, I question the notion that we should give explanatory priority to the cases that are harder, rarer, and which involve more conscious reflection. I believe philosophical accounts should track the practice we actually engage in most of the time – not the just practice we believe we ought to have when we pose the question in moments of cool reflection. If our careful reasoning is beholden to false simplifications and distortions, it will lead us astray as a source of practical guidance or else turn out to lack application to lived experience altogether.

In addition, as it is in other contexts, so too in moral theory: hard cases make for bad law. Hard cases typically call for reflection because they involve additional factors or complexities that put our ordinary ways of responding out of force or under unusual strain. The demands of the extraordinary should not impugn the adequacy of our ways of handling the ordinary. However, since it is the adequacy of the bottom-up strategy that is in question, it will not do to rest my case solely on this thought, so let me offer a different line of response.

The principal reason we care about separating repair from exculpation and condonation, I suggest, is that failing to do so can seem profoundly unfair. The distinction between repair and exculpation in particular becomes important, one might think, because *not* being excused makes one liable to a sanction or loss, and it is unfair to impose such costs on those who do not deserve to bear them. By contrast, a wrongdoer who is *not* excused, and so *does* deserve to bear the cost of her wrongdoing, is disrespected and underestimated when the victim falsely treats her as non-liable; it is condescending or offensive to be treated as akin to a child without the requisite capacities for mature responsiveness, or as lacking in self-control or moral knowledge when one is in fact fully responsible for one's wrongful conduct.

But if this is the real motive for my opponent's critique, it need not trouble us; for this understanding of exculpation takes for granted the bookkeeping conception of repair as liability to compensate that I have criticised and rejected. On the bookkeeping model, is it correct to say that a wrongdoer owes restitution to her victim if and only if she is not excused. But on my conception of repair, no such connection between liability, desert, and repair follows.

A similar line of response can be made concerning the charge of a collapse between repair and condonation. If we take for granted that condonation involves a failure to exact restitution as a condition for reconciliation, then my alternative conception of repair looks like it fails to mark the distinction; for I allow that the parties can restore their bond and reconcile even when neither has repaid or compensated the other. But if we jettison the bookkeeping assumption that motivates this account of condonation, no such conclusion follows.

The crucial point, then, is that much of our interest in separating excuse and condonation from apology and repair stems from an implicit acceptance of a bookkeeping framework; but when that framework is jettisoned, we also lose much of our reason for closely guarding the boundary lines. This consequence should be welcomed; exculpation and condonation are both contested and contestable notions, and we should leave open the possibility that there are no stable and context-invariant conceptions out there for us to discover. If the objector is to sustain her charge, she needs to argue i) that my conception of repair cannot make sense of *any viable way* of distinguishing between repair and exculpation or condonation, and ii) demonstrate why this is a cost – even for someone who rejects the idea that repair must be earned through restitutive repayment.

The alchemy of apology revisited

The question of the distinction between exculpation and condonation notwithstanding, an objector might persist in worrying that my account of the pursuit of repair nonetheless leaves out something of profound importance. Indeed, this returns us to the challenge canvassed in a previous section of the chapter: That even if violations of moral norms involve a failure of mutual understanding, remedying that deficit does not itself suffice for repair. Having taken on board the lesson from the previous section, we are now conscious of the fact that the sufficiency challenge must not sneakily reintroduce the assumption that repair resemble what the bookkeeping model describes. But all the same, a question may remain: Can it really be enough for repair that the parties talk it out, and arrive at some understanding of one another that is accurate and acceptable to them both?

To answer this question, it must first be recalled that the notion of mutual understanding I am proposing characterises it as an epistemic project with a distinctive agential and interactional profile. The parties' understandings needs to be *shared*, rather than simply symmetrical and mutually known. For this reason, it does not suffice that each has arrived at an understanding of the other, nor that each knows of the interpretation the other applies; if I have concluded your behaviour evinces a selfish lack of concern for my interests, and you have concluded my hostility expresses a sentimental oversensitivity, say, and both of us understand that this is how the other sees the situation, then we do not yet have an understanding both of us can share.

The understandings we have must also be *practically enacted*, rather than merely theoretical; what is needed is not just possession by each of us of the right constellation of attitudes at a time, but the manifestation and application of these understandings in action over time. It is not sufficient for mutual understanding that each of us has *read the other right*, but keeps this to herself. The understandings we develop need to find expression in our interaction. Moreover, if the appraisals we make of the past transgression are themselves affectively toned, then sharing them will require not just adopting and acting on beliefs about the other, but also relating to and validating his emotional reactions.

Part of the reason for this is that mutual understanding is fundamentally *open-ended*, rather than finite and fixable; since neither party is static, the 'target' for understanding itself shifts and develops over time, and this itself can affect what narrative construals we find to be fitting and informative. Moreover, it is not just the passage of time that changes us, but also the pursuit of mutual understanding itself; the project of arriving at an adequate story can itself supply and unlock new insights and provide opportunities to reposition or develop ourselves[92].

Finally, it must be noted that the objective of the pursuit is not just adequate interpretation of how a shared past has brought us to the present moment; we may be, and typically are, interested in understanding the past for its own sake, but provided we have not been dissuaded by what has already been uncovered, we also pursue understanding to set a viable direction for a future we can share. Narratives, as Camp proposes, can supply 'wayfinding instructions'; knowing where we are and how we got here is the first task to undertake in order to pave a path forward together.

To see the significance of these final two points, consider a pair of people who have attempted to repair, but have ultimately abandoned the pursuit. "I really tried to understand her position, and to get her to see what I see", one might say, "but we just couldn't see eye-to-eye". Or: "We ultimately agreed that we are just too different to work this out". In a case like this, the parties have found themselves stumped by the difficulty of establishing mutually acceptable terms for their future engagement. They may have a sharable understanding of their past, but cannot agree on where this

[92] Annette Baier, Victoria McGeer, and Maria Lugones have each described the possibility of personal transformation in confrontation with another's world view in terms of a metaphor of travel (Baier, 1997; Lugones, 1987; McGeer, 2022). See also (Dover, 2022, 2023a) for accounts of the possibility of personal change and development in the encounter with another.

leaves them *now* and what should happen next. Earlier in this chapter, I suggested that one source of this stalemate could be a lack of adequate hermeneutical resources; but it could also result from a failure to see either oneself or the other as still capable of change. It can be tempting and comforting to withdraw from engagement with the conviction that "I am what I am, and I shouldn't have to change myself for his sake!" – and it can be easier still to worry that one will somehow be *changed by the other*.

Philosophy often cultivates and celebrates this position[93]. In describing the 'therapeutic' worldview he considers the only alternative to responding to others' wrongdoing with punishment, Herbert Morris writes that its "logic of cure will push us toward forms of therapy that inevitably involve changes in the person made against his will"(Morris, 1968, p. 487), and further that:

> "In doing this we display a lack of respect for the moral status of individuals, that is, a lack of respect for the reasoning and choices of individuals. They are but animals who must be conditioned. I think we can understand and, indeed, sympathize with a man's preferring death to being forcibly turned into what he is not."(Morris, 1968, p. 487)

In far less hyperbolic terms, Bennett, too, worries that whereas punishment (including that which is self-imposed) respects a person's status as a mature agent, attempts to educate or instruct her "would suggest that it was not the wrongdoer's responsibility to discover and decide for [herself] how to meet [her] responsibilities within the practice." (97) An rational adult member of the moral community, he argues, "can be left alone to get on with things – and can expect as a matter of right to be left alone to get on with things" (96)[94].

[93] For a discipline officially committed to the value of persuasion and learning, philosophers often seem surprisingly uncomfortable with the idea of letting oneself be influenced or changed by another See esp. (Dover, 2023a; Springer, 2013)

[94] Even Dover (Dover, 2019) raises the concern that communicative responses to wrongdoing can be condescendingly didactic and lecturing, amounting to a moral pedagogy unfit for adults. However, Dover's scepticism is significantly tempered later in the paper: "even an initially univocal, peremptory intervention can end up resulting in a much more substantive dialogue than the initial speaker bargained for—so long as the initial speaker is willing to hear what her addressee has to say in response, and to respond in kind. What ultimately matters most is not how a conversation starts

Alongside the challenge with which I started this section, the concerns raised by these authors exhibit the twin hazards of both over- and underestimating the normative potency of talk. One the one hand is the worry that merely talking cannot possibly be up to the task of restoring what wrongdoing has wrecked; on the other is the fear that it will disrespectfully or dangerously attempt to undermine, subvert, or discipline autonomous adult agency.

The suggestion with which I wish to end this section is that we give ourselves over to these forms of change and development, and embrace the creative and recreative possibilities they hold. Our need for repair is itself a manifestation of our vulnerability to other people; the pursuit of repair, as I have characterised it here, shows that this pursuit can also be a source of value: a kind of alchemy, then, after all – securing the possibility of life between us.

but where it goes." (fn omitted). As previously noted, Dover's other work (Dover, 2023b, 2023a, 2022) also takes a far more favourable position on the role and value of interpersonal influence.

Chapter Six

Standing up for yourself: Protest as self-creation and self-discovery

In the past five chapters, I have focussed on the problem of repair at it arises in an interpersonal setting. Indeed, I have argued that we are liable to misrepresent and distort that problem and our resources and procedures for resolving it by treating it as a problem of legal or political justice. I have also argued that the pursuit of repair is fundamentally a joint project – something that requires the interaction and cooperation of both parties. If one is unable or unwilling to participate, repair becomes impossible.

This final chapter to some extent departs from the rest of the project. Here, I set out to consider how responses to wrongdoing change when the wrongful treatment is also an injustice, and when the victim cannot depend on the transgressor's support and collaboration in setting it right. In particular, I shall ask what is at stake for a person who finds herself in this situation; when repair is not in her reach, does she still have a reason to protest her wrongful treatment?

I shall argue that she does. Drawing from ideas put forward by Bernard Boxill, I shall propose that a victim of injustice should protest her unjust treatment in order to show and know herself as self-respecting.

After introducing the problem of responding to injustice, I consider two views of the basis of a person's self-respect: on the Dignity View, a person's self-respect is based upon her inherent nature, whereas on the Honour View, it is based upon the social standing conferred on her by others. I argue that neither view can explain the epistemic role of protest, since they make the connection between a person's worth and her knowledge of it either too tight or too tenuous. Instead, in the chapter's second half, I develop and defend an alternative proposal.

Starting from the idea that a person's worth can be 'up to her', I propose that one can *take responsibility* for one's self-respect – for the person one is, or is becoming. Doing so, however, requires an act of public commitment; a person can stand up for herself by standing up before others. In this manner, protest can be a form of self-creation. I then consider whether an act of commitment can also be genuinely self-revelatory. Its apparently self-fulfilling character might seem to preclude that it can, but I suggest this conclusion can be avoided. If so, protest can be a form of self-discovery as well.

Introduction

Victims of injustice confront an often unwelcome choice: should you stand up for yourself and protest your mistreatment? It may seem obvious that the answer is yes; but why? Here's a first stab: injustices, like all moral wrongs, must be recognised as such, and will not be recognised if we let them go unopposed. We could flesh this view out in two different ways: first, we could argue that protesting injustice is normatively important because of its beneficial effects; perhaps it will compel the wrongdoer to understand our plight, transform herself, or offer remediation; perhaps it will change the attitudes of bystanders or third parties, or recruit them to our just cause; perhaps it will prevent injustice from happening again. Alternatively, we could maintain that protesting injustice is normatively important even if we have no realistic hope of bringing about such outcomes through our intervention. What makes protest important, we could argue, is not its positive consequences, but the opportunity it represents to stand up for and speak the moral truth to those who would deny or ignore it.

But while recognising these laudatory goals, it is also worth recognising that what makes the choice to protest injustice an unwelcome one is not just that it is born of undesirable circumstances (one's own mistreatment); it also that open protest carries costs and risks for the one doing it. In the worst cases, protesting mistreatment can expose one to the risks of escalation, retaliation, and violence; and even in the best cases, it can deplete one's energies and resources, directing valuable time and attention away from the pursuit of one's freely chosen projects and commitments.

This brief sketch of the possible values and risks of protest has a notable consequence. If either of the views proposed is correct, from a victim's position, the best circumstance of all would seem to be one in which *someone else* takes up the mantle of opposing her unjust treatment on her behalf. Enabled to remain in the background where things are safer and easier, a victim of injustice could then support the efforts of others, or at least endorse them quietly or privately, whilst herself putting less at stake.

My objective in this chapter is not to argue that such a strategy is never justified; for I am happy to accept that it often is[95]. Nevertheless, I believe it has a cost to refrain from standing up for oneself and protest on one's own behalf. This is not because doing so is to engage in morally problematic kind of free-riding on another's efforts; bystanders and third parties too can have a legitimate interest in promoting good outcomes and engaging in moral truth-telling. Nor is the problem that there will not always be another person to fight the good fight in our place, though this is plausibly often the case. Rather, the idea that will interest me here is that refraining from standing up for oneself can be damaging or undermining to a person's self-respect[96].

[95] Indeed, it is plausible that those who are most frequently subjected to injustice also often confront the greatest risks in opposing it: the structurally marginalised and oppressed; those without adequate material resources or political power. The relationship between these factors is undoubtedly complex, but it may be that increased exposure to injustice stems in part from the implicit assumption by wrongdoers that they can act with relative impunity.

[96] I emphasise that I am not proposing this is the *only* reason for preferring that victims of injustice protest on their own behalf. Perhaps victims themselves posses distinctive epistemic insight or authority that protests against injustice ought to encompass and reflect; perhaps it is condescending or patronising for third parties to speak on victim's behalf. The story I set out in this paper aims to identify grounds for standing up for oneself that are independent of the presence and weightiness of alternative considerations like these.

This idea is not original to me. Several philosopher have noted the connection between self-respect and protest against mistreatment, and in what follows I will draw from several of them. In particular, I will take an argument advanced by philosopher Bernard Boxill (Boxill, 2010, 1992, 1976) as my starting point.

What makes Boxill's account of the connection between protest and self-respect so intriguing, and, as I hope to show, instructive, is his sensitivity to a complex epistemological problem of self-understanding a victim of injustice confronts, and which underlies and complicates her first personal problem of deciding what to do in the face of injustice. This relation between the practical and the epistemic, I shall propose, leads to an explanatory challenge. It is to this challenge I turn in the next section.

Boxill's Puzzle

Why must a victim of injustice take a stand on her own behalf? Why can she not turn the other cheek, take 'the moral high ground', and pretend to be unaffected by the unjust treatment she suffers? In particular, does she not have adequate reason to do these things in cases where she can reasonably expect her protest to go unheard or unheeded? Finally, why should any of these questions have a bearing on her self-respect?

Boxill is aware of this sceptical line. He writes:

> "Sometimes to fail to protest injustice, to keep silent, is simply to say nothing. It is enough in such cases that the victim be sure in his own mind that he has been wronged. Why should he tell a cynical, uncaring and incorrigible world that he has been wronged? And, if he should for some reason, why should it be out of self-respect?" (Boxill, 1984, 190)

Adding to this line of thought, we could note that failing to stand up for oneself through protest need not amount to failing to resist injustice entirely. There will sometimes (and perhaps always) be the possibility of affirming one's attitude or renouncing injustice through acts of 'quiet' or 'internal' resistance[97]: muttering disavowals under one's breath, thinking them privately, writing them out in one's secret diary, or surreptitiously working to oppose the transgressor's evil cause. Why would such responses not suffice to guard one's self-respect?

"[O]nly consummate artistry", Boxill writes:

> "can permit a person continuously and elaborately to pretend servility and still know that he is self-respecting. Unless it is executed by a master, the evidence of servility will seem overwhelming and the evidence of self-respect too ambiguous. But … the self-respecting person wants to know he is self-respecting. … If only occasionally, he must shed his mask." (Boxill, 1976, 68-69)

Boxill's direct concern here is with the opposition, or rather the lack of the same, to racial oppression, as this is manifest in the behaviour of such cultural or literary archetypes as the 'Sambo' or the 'Uncle Tom', viz. the apparently "docile" and "good humoured" Black characters who seem to willingly submit themselves to racist degradation to curry trivial favours or keep themselves out of trouble. A Black person who feigns deference and servility, Boxill argues, does not just risk disadvantaging others. He also exposes himself to a grave risk:

> "such pretense has its dangers; it shakes [a person's] confidence in his self-respect. … the self-respecting person in such straits must, in some way, protest to assure himself that he has self-respect." (1976, 65-66)

In making this argument, Boxill is explicitly aligning himself with the position of W.E.B. Du Bois, whose influential writings in defence of protest put him at odds with Booker T. Washington over progress towards Black advancement in early twentieth century America. While Washington too would have been critical of servility, he argued for a focus on promoting economic equality, and recommended resorting to protest only when it was likely to be efficacious for securing political goals. By contrast, Du Bois maintained that "even when bending to the inevitable", a person

[97] Notably, these possibilities apply whether or not anyone else is also protesting openly and loudly on one's behalf. For interesting proposals along this line, see works by (Hay, 2011) or Tamara Fakhoury (Fakhoury, 2021)

must "bend with unabated protest" (Du Bois 1973, p. 43); "silent submission to civic inferiority" (Du Bois 1966, p. 514), he argued, is not consistent with maintaining self-respect[98].

While aligning himself with Du Bois, Boxill's argument has more a pronounced self-directed epistemic bent: it is for the sake of *one's own self-assurance* that a victim of injustice must not sit idly by or pretend to be unaffected by mistreatment. As Tommie Shelby rephrases the problem:

> "The worry is not so much about how others might interpret one's silence. The worry is about what such silence may reveal about one's character; it is about the threat of moral degradation. When the prospects for ending, reducing, or escaping one's oppression are dim, one can easily come to accommodate oneself to unjust conditions, effectively surrendering. The moral dissonance can even tempt one to rationalize one's condition, perhaps regarding it as not all that bad." (Shelby, 2010, p. 352)

Insisting upon the value of standing up for oneself does not mean one must *always* pursue a confrontation, no matter how trivial the slight or significant the likely costs; sometimes is it justified 'to acquiesce to injustice to avoid serious physical harm, to protect loved ones, to live to fight another day, or to die a more meaningful death at a later time' (Shelby, 2012, p. 517). However:

> "even if one knows that one's rights have been violated, how could one be confident that one's silence in the face of this injustice is not rooted in rationalization or unworthy motives such as cowardice? When one is convinced that all other modes of resistance are closed off, one can at least voice one's resentment about one's maltreatment. To be secure in one's belief that

[98] Many philosophers have taken a similar position. For instance, Agnes Callard (Callard, 2020) writes: "When people commit injustice against us, we feel it: our blood boils. At that point, we have to decide how much we want to fight to quell our anger, how much effort we are going to put into repressing and suppressing that upswell of rage. The answer is rarely none. While we do not want to let our anger get away from us and drive us to its logical, eternally vengeful conclusion, if we quash it with too heavy a hand, we lose self-respect and, more generally, our moral footing."

> one values one's rights, one must break one's silence, and suffer the consequences.' (Shelby, 2010, p. 352)

The insistence on protest thus seems to stem from its importance to a person's secure self-conception. Breaking the silence, voicing one's resentment, and exposing oneself to the consequences can prevent doubts, self-questioning, and even self-degradation from taking roots. In sum: "persons have reason to protest their wrongs not only to stop injustice but also to show self-respect and *to know themselves as self-respecting*."(Boxill, 1976, p. 59 emphasis added).

The explanatory challenge, then, is this: How could protest enable a victim of injustice to show and know that she is self-respecting? I shall call this challenge Boxill's Puzzle.

Self-respect and worth: The Dignity View

What, then, does it take to convince or assure oneself in the face of assaults on one's self-respect? In this section and the next, I want to consider two common conceptions of the basis of a person's self-respect, and evaluate them as answers to Boxill's Puzzle: can they explain how standing up for oneself and protesting one's mistreatment is also a way to maintain and protect one's self-respect in the face of challenges to it? My response, to give the game away in advance, will be that they cannot.

Consider first the position that I shall label the Dignity View. On the Dignity View, the basis of a person's self-respect lies in her distinctive rational or human nature – what, as the label reflects, we commonly call her *dignity*. The idea of dignity is most commonly associated with a Kantian moral tradition, where it denotes a kind of value that is absolute, categorical, and incomparable to any other.

A couple of notes of clarification are in order: first, though I shall explicitly borrow the label 'dignity' from a Kantian vocabulary, my project here is not exegetical, and I will not therefore be concerned with whether dignity as I will describe it is the very notion Kant himself had in mind (if, indeed, there is one such notion, rather than several more or less overlapping ones, for example). It will suffice for my purposes if my use of the term manages to map onto the idea that has become common currency in the contemporary literature as it is influenced by a broadly Kantian legacy of thinking about the nature of human worth[99].

Second, and relatedly, I will not be at pains to specify precisely which aspects or capacities of a human or rational nature *secure* a person's dignity, i.e. whether it be autonomy, rationality, agency, membership of the human species, or something else. (Properly speaking, then, my label will arguably attach to a family of views rather than a single one.) I will rely only on the following assumptions which, I shall claim, are almost universally associated with it: that a person's dignity is inalienable and unconditional.

Even this rather thin and cursory characterisation of the Dignity view should suffice to illustrate the problems it will confront in explaining or addressing Boxill's Puzzle. First, the Dignity View will struggle to explain how it can be rational for a victim of injustice of worry that her self-respect is at stake at all. Second, even if this challenge can be met in a satisfactory way, the Dignity View will be unable to account for the epistemic role the Puzzle as stated assigns to protest. I shall go over these problems in turn.

The first problem goes as follows: If the basis for a person's self-respect is her nature as the distinctive kind of rational or human creature she is, it is difficult to understand what it could possibly involve for her to come to question whether she will be able to retain it in the future, and so to make sense of the predicament that raises Boxill's Puzzle in the first place.

Let me clarify the claim I am making here. It *is* possible, of course, to be unaware of, or mistaken with respect to, one's distinctive nature: a human being can (tragically, we might add) be convinced that her value is not equal to that of her peers, and so appraise herself as unworthy of the self-respect she judges as appropriate to them; alternatively, and less troublingly, a person may simply never come to entertain the question of her own worth, and the judgement of herself it merits, at all. But note that

[99] Amongst the many proponents of views along these lines, some of the most notable examples are found in (Darwall, 2015, 2006, 1977; Dillon, 2020, 1997, 1992; Hill, 1982, 1973).

neither of these circumstances accurately describe the situation Boxill's Puzzle involves: the person who confronts the puzzle is clearly considering, or coming to consider, the question of her worth and what it will take to retain it, rather than remaining happily oblivious to the matter altogether. Nor does she simply hold a mistakenly low estimation of herself; if she did, and accordingly considered her subjection to mistreatment an accurate reflection of her inherent inferiority, she would not be motivated (not even ambivalently, in recognition of its attendant costs) to protest it.

Instead, the person described in Boxill's Puzzle is supposed to be affected by, rather than indifferent to, her mistreatment, and fears that without protesting it, she will lose confidence in her own worth. If the Dignity View gives the right account of the basis for her self-respect, we are left with the following interpretation: she must fear losing confidence in her self-respect because she fears becoming mislead or mistaken; she will fear coming to adopt an attitude towards herself which must by her own lights be irrational. In a phrase, if a person's self-respect is based on her unconditional and inalienable dignity, loss of it should be impossible; and loss of confidence in one's self-respect therefore always irrational or mistaken.

We could accept this conclusion, but it should be acknowledged that it is a significant bullet to bite. It is descriptively implausible to place the confidence and accuracy of one's self-respect on a par with confidence and accuracy of belief in any other fact or feature pertaining to oneself – concerning one's blood-type, say, or place of birth. Becoming mislead or mistaken with respect to facts about oneself need not be a trivial matter – it can be extremely inconvenient or even dangerous to be so misled or mistaken. However, the anxiety described by both Boxill and Shelby in setting out the case is not simply the fear of being mistaken in adopting a low view of one's worth; it is the fear of adopting a low view of one's worth and *being correct*.[100]

We could modify the view to specify e.g. that the basis of a person's worth lies not in her nature as such, but in a specific exercise or development of it, say. This would

[100] Cynthia Stark (Stark, 1997) raises a different, but related objection to the Dignity View, claiming that it couching doubts in one's worth as necessarily irrational is politically pernicious.

enable us to say that she could fear loss or lack of her self-respect without thereby fearing irrationality, but doing so would come at the cost of the fundamental commitments of the view I started from – viz that dignity is inalienable and unconditional, and therefore universally available to all.

Let us turn now to the second problem. Even if the first problem can be avoided, we will not get the rational justification of *protest* we had set out to provide. For the question now becomes: why should protesting *in particular* be the thing that guards a person against irrationality or mistake? If one is fearful of epistemic mistakes, shouldn't one simply issue oneself with plenty of reminders? It seems there would be no reason to stand up for oneself in the way involved in protest, rather than muttering the truth under one's own breath, writing it in one's diary, or affirming it in thought – in short, with resisting internally and quietly, instead.

Let me be clear that the problem with the Dignity View is *not* that it cannot explain how a mistreated person could have *a* reason to protest; a proponent of such a view could certainly maintain that protesting is the right thing to do, or even that it is an absolute requirement. What she cannot explain is how or why protesting would play the epistemic role of enabling one to *know* oneself as self-respecting. Protesting could not be a way of improving one's epistemic position, or form of self-discovery and self-revelation. You may have to protest to speak the moral truth, but you cannot *discover* the truth in so speaking.

Self-respect and Worth: The Honour View

Let me know turn to the second view under consideration. On this view, the basis of a person's self-respect does not reside in her human nature, but rather in the position she is afforded in a social structure or network of others. I shall call this view (or family of views) the Honour View.

The Honour View paints an entirely different picture of human worth to the one I have just been considering. Most notably, on the Honour View, human worth is not guaranteed to be equally and universally distributed to all people – in fact, it is rather likely not to be, since the basis of one's worth is here *conditional* on how one is viewed and judged by others. Honour, as characterised by this view, is not inherent, but socially conferred.

It is worth noting that a proponent of the Honour View need not reject the Dignity View; the two views can co-exist as complementary facets of a person's worth[101]. A person has both dignity and honour, we could argue, and when we say she acts out of self-respect or for the sake of self-respect, we sometimes refer to the one basis, and sometimes the other. If dignity cannot be the basis of the self-respect that is at stake in Boxill's Puzzle, as I have argued in the preceding section, perhaps honour could be instead?

Indeed, the Honour View initially appears to give us a straightforward story that explains the point of protest: By protesting, a person can attempt to occasion and earn a more favourable honour-conferral, whereas remaining silent will relegate her to remaining in the low position imposed on her by the initial mistreatment.

But this view confronts problems of its own. First, we should wonder both how and by whom honour is conferred. Discussions of honour-respect typically illustrate the idea with reference to honourable offices, positions, or hereditary connections; a Justice of the courts, or a titled Nobleman might occupy positions that render them honourable, and we can explain precisely why this is so. But no such story is available in the case of the honour of an ordinary citizen or person (or even, we might add, the honour of the Justice or Nobleman *qua* citizens or persons, when their mistreatment occurs outside the context of their formal roles). The notion of honour at stake here must be a different one.

We can see the relevance of questioning the source of a person's honour by asking ourselves whether a person subjected to injustice *thereby* loses (some of) her honour? If the answer is no (if say, the transgressor is not someone whose attitudes count for anything in establishing honour-worth), then there should be nothing for her to worry about, and Boxill's Puzzle will fail to get off the ground at all. If, however, the answer

[101] See e.g.(Darwall, 2015) for such a case. Other authors whose views can be placed in this category include (Waldron, 2012), or (Moody-Adams 1993).

is yes, we shall have to say that she has not only been treated dishonourably, but has been *dishonoured*, i.e. has been lowered in or deprived of honour she formerly enjoyed. But now it must follow that attempting to use her protest to secure greater worth must either be a plea or a sham. She must either be pleading that the authoritative conferral be revised (albeit *not* because it is inaccurate, for this is ruled out by its being authoritative), *or* duplicitously *pretending* that it is inaccurate in the hopes of manipulating her audience into a change of view.

So far, I have described the Honour View's account of the worth that forms the basis of self-respect (viz. honour conferred by others), the ways that worth might be affected by protesting (viz. through seeking to affect such conferrals with pleas or pretence), and noted some explanatory gaps and costs. Now we can ask: if the Honour View is correct, how could protest enable a person to know that she is self-respecting?

The answer, I submit, is simply that it could not. Again, as with the Dignity View, the problem is not that a proponent of the Honour View cannot explain how a mistreated person would have *a* reason to protest her mistreatment; she could maintain that protest has strategic value as a way of trying to influence the attitudes of other people, persuading them to afford one more honour. But this is not the epistemic role for protest set out by Boxill's puzzle, viz, that protest is what enables a person to know that she is self-respecting.

A response at this point might go as follows: Admittedly, protesting is not on its own sufficient for being in a position to know, but it is necessary, for a person has to protest to attract the honour upon which her self-respect will eventually rest. But note that there is no reason to accept that this is the case; we could equally well imagine that whosoever authoritatively confers honour simply alter their conferral independently, or that the agent's protesting prostrations are never successful. If the Honour View is right, we have no reason to think protesting is either sufficient or necessary to know oneself as self-respecting.

An impasse?

At this point, we seem to come to an impasse: If my argument so far is successful, neither the Dignity View nor the Honour View can vindicate the rationality of protesting for the sake of showing and knowing one's own self-respect – not, at least, without resorting to irrationality or self-deception. The Dignity view gives a person's self-respect a sound and stable basis in her inalienable nature, but in doing so makes the possibility of doubt difficult to explain, and deprives protest of the decisive epistemic role Boxill's puzzle sets out for it. By contrast, the Honour view makes self-respect conditional on the attitudes of others; this gives a purpose to one's protest, but this time one that is far too tenuous to provide successful self-reassurance. In sum, connection between the *worth* on which self-respect rests and one's *knowledge* of that worth becomes either *too tight* or *too fragile* for protest to play the role envisioned for it in Boxill's Puzzle.

The idea that protesting can be a way of showing and knowing one's self-respect looks caught between two irreconcilable ideas: one the one hand, that protest is a type of truthful self-revelation, a faithful representation of a nature that is already there; and on the other, that it is an act of self-creation, which gains its importance from the value it produces by impacting others.

In what follows, I hope to show that these two strands – protest as self-revelation and protest as self-creation– can in fact be joined together, and that a vindication *can* therefore be given to Boxill's characterisation of protest as a way to show and know one's self-respect. The reconciliation will require attending to possibility that has so far gone unexplored: that a person's worth is *up to her*: not just a question of what she irrevocably is or what others make of her, but also what she *makes of herself*. That a person is in this manner both the *maker* and the *made* in turn affects what it takes for her to know herself, and gives rise to distinctive possibilities for self-discovery and -realisation that are not available to anyone else.

To anticipate: My proposal will be laid out in two stages. In the first stage, I aim to illustrate what it involves for a feature of myself to be *up to me*, a product of my own making. This part of the argument will rest on the idea that a person can *take* *responsibility* for herself – for the person is she is, or is becoming. Doing this, I shall argue requires a distinctively *public* commitment – precisely the sort of commitment one makes in standing up for oneself in protest.

In the second stage, I shall argue that the features of myself that can be established through such acts of public commitment have a distinctive first-personal epistemology – one which makes possible precisely the connection between showing and knowing that solving Boxill's puzzle requires.

Self-creation: Taking Responsibility for Oneself

A wide range of philosophers have taken an interest in the idea of *taking responsibility*, and employed it to diverse purposes[102]. Since it is both well-developed and touches directly on a number of themes relevant to my present purposes, I will focus my attention on the account set out by David Enoch (Enoch, 2011).

Enoch proposes that it is sometimes possible to *make* oneself responsible for something through an undertaking of commitment. In setting out this idea, let us follow Enoch in first distinguishing between those things that are clearly and indisputably one's responsibility (let us say actions which are freely undertaken and fully understood), and those things that cannot possibly be so; things as far removed from one's agency as occurrences in human pre-history, or, in Enoch's own example, the movements of the planets will serve as illustration. Without worrying precisely where and how to draw the lines that delineate each of these categories, it seems plausible that there is a fairly wide gap in between them. Enoch terms this gap (or perhaps rather a privileged and sufficiently proximate proper subsection of it) the sphere of my *penumbral* agency.

Although I am not (we are stipulating) straightforwardly responsible for the things that fall into this sphere, I am able to *take* responsibility for them, and thereby *become* responsible. The relevant kind of taking, Enoch argues, can be illustrated on analogy with the undertaking of an obligation or commitment, such as the making of a promise (Enoch, 2011, p. 104).

[102] Examples include (Adams, 1985; Calhoun, 2019; Mason, 2019; Raz, 2011; Walker, 1997; Young, 2011).

Enoch's particular interest is the manner in which such takings of responsibility can give rise to obligations that would otherwise not be there, but which one could nevertheless have a duty to impose on oneself. Since this is not my concern here, I will not say anything about the plausibility Enoch's account of such duties or about how they could or should be discharged. I will instead restrict my attention to the point Enoch makes about why we should accept the proposal that we have the power to take responsibility for things in the penumbra of our agency. Enoch proposes that we ask ourselves whether a good moral engineer would have reason to give us such a power. He proposes that she would, since 'the power to take responsibility for things that lie outside the scope of one's core agency is a constitutive part of valuable relationships and ways of thinking of ourselves' (Enoch, 2011, 124), and notes points of similarity in the work of Joseph Raz. Raz writes:

> "Failure to control conduct within our domain of secure competence threatens to undermine our self-esteem and sense of who we are, what we are capable of, etc. We must react to it. We may conclude that we are no longer able securely to perform that kind of action. We have grown frail, our competence is diminishing. We come to recognize our limitations. Commonly that is not the case, and we do not allow it to be. We assert our competence by holding ourselves [responsible] for it" (Raz, 2011, p. 245).

To this list we could add also Elinor Mason (Mason, 2019), who argues that it is precisely the favourable role taking responsibility plays in our relationships that justifies the practice. Notably, however, Mason criticises Raz's claim "that holding ourselves responsible for failures within the domain of secure competence is required by self-respect" on the grounds that this renders an agent's reasons for taking responsibility "inward looking in an inappropriate way"(Mason, 2019, p. 185).

This criticism of Mason's should be resisted, however, at least when it comes to taking responsibility in the manner that interests me here, viz. by standing up for oneself in the face of challenges to one's worth. Even when such takings of responsibility are not owed to anyone else – when persuading the transgressor is out of the question, and there are no by-stander or third-party interests to consider, say – we can, and should, do so for our own sake: for the sake of our sense of who we are and what we are capable of.

To conclude this part of my proposal, I need to make one more modification to (or clarification of) Enoch's account of taking responsibility. As stated, Enoch's position is that a person can take responsibility through 'an act of will' or 'undertaking of commitment', which, as we saw, he illustrates with analogy to making a promise. However, Enoch also tentatively accepts that such undertakings may be 'implicit', carried out simply by taking action to discharge the responsibility in question, or even through an entirely "uncommunicated decision" (Enoch, 2011, p. fn 15). At some points, his discussion appears to admit of even less; for instance, he discusses the case of a person who takes pride in his nation's positive achievements but declines taking responsibility for its failures. What is criticisable about such a person, Enoch writes, is that his pride shows he "has already (implicitly) taken responsibility for his country's actions more generally" (Ibid. 112); on this point, I disagree. It is implausible to me that pride is the sort of response that can arise through an 'undertaking of commitment', 'act of will' or 'uncommunicated decision'. If this is right, a prideful response is simply not the right sort of thing to suffice for having taken responsibility, even on Enoch's own picture[103]. Moreover, this is all for the good: think again of the moral engineer and her objectives. Surely what is valuable in taking responsibility is its manifestation in action, and not simply in private thought or feeling.

I belabour this point because, recall, my project is to explain why a person should choose protest over silence in the face of an injustice she alone cannot redress. If I am right that she can protect or affect her worth through taking responsibility for herself, we need to know why the former, but not the latter, could be a way of doing so. My answer is as follows: taking responsibility is the undertaking of a commitment, and commitments like promises get their distinctive binding force, the force that enables

[103] What Enoch *should say*, I propose, is this: the proud person's pride, like the lorry driver's regret, reminds or compels him to *take* responsibility, but does not itself constitute or amount to the taking. If an affective response like pride is itself sufficient for already having taken responsibility, what would the wilful decision or commitment add? Things are different, of course, if it is not the pride (or regret) itself, but its manifestation or expression in action that constitutes taking responsibility. This would indeed give us an act of will, but would also bring us rather close to publicity after all.

them to change the normative facts, only when they are made public. They require action before others.

To be clear, my proposal is not that taking responsibility for oneself requires making a commitment *to* another. Indeed, the protesting person's commitment to herself and her worth is self-directed. Nevertheless, its normative significance requires that it has an audience. She stands up for herself by standing up to and before others[104].

We could call this an uptake-condition on taking responsibility, but on its own this kind of vocabulary risks obscuring as much as it illuminates - what does taking-up amount to? How does it occur, or fail to do so? I will restrict myself to a relatively minimal requirement of audience receptivity akin to the characterisations offered in Chapter Three: first, the protest must actually have an audience in the sense that there must be people within earshot; it will not succeed if it is made to an empty room, or in a voice too low for anyone to hear. Second, it must be designed with the intention that its audience can recognise it for what it is, viz. recognise it *as a protest*, rather than, say, a joke or a request. In other words, protest is a passive social act. This leaves in place possibilities for various kinds of partial failure: though a protest is audible, it may 'fall on dead ears' if its audience are too preoccupied or indifferent to listen; and though one might have taken pains to make oneself understood, it can never be guaranteed in advance that one's meaning will be grasped[105].

I am aware that the idea of standing up for oneself by standing up to and before others would benefit from, and ultimately deserves, a much fuller defence than I can give it here. However, I will at least suggest that it should have some intuitive plausibility. Many ordinary human practices involve making one's commitments public (swearing-in rituals; wedding vows; the Hippocratic oath), even in circumstances where they carry no formalised significance (pledges of political support; new years resolutions; abstinence vows; commitments to sobriety or addiction recovery).

[104] On this idea, see also (Calhoun, 1995)

[105] There are many more possibilities to consider here than I can do justice to in this chapter. What about a sincerely intended protest that is delivered through apparent self-deprecation or irony? What about a protester who does not believe her intention to protest will be understood; can she still do it, or will she be prevented by the expected hostility of her environment?

In extending this story, one could suggest that this is demonstrates the value we attach to interpersonal answerability; putting one's commitments out in the open makes one accountable to others, and exposes one to their criticism and judgement, even when they have no claim on one's fulfilment. More can undoubtedly be said to explain this phenomenon; perhaps a critic could maintain that the attachment to the publicity of commitment I am describing, or the idea that public commitments can increase a person's sense of her own worth, is itself irrational or mysterious; if the commitment is really made for one's own sake, and not to engage or influence the other, what is its point?

However, I have also argued that the need for publicity recurs in other, similar contexts: in needing to bear witness, in desiring to put one's experiences on the public cognitive scene and have them recognised by others; in somehow compelling a response, even if only a heavy, loaded silence. At the very least, then, I hope to have made plausible the claim that it is a kind of mysteriousness that is more familiar and less troubling than those posited by either the Dignity or the Honour View as responses to Boxill's Puzzle.

Self-discovery: Knowing oneself as self-respecting

The idea that a person can *make* her own worth by *taking* responsibility for it through protesting only gets us part of what we need to discharge the explanatory challenge of Boxill's Puzzle, however. We also need a story about how protesting enables a person to know herself as self-respecting.

I shall develop my view through addressing the following challenge: Even if we accept the preceding argument that a person can shape her own worth by publicly taking responsibility for it, and thereby *make* herself self-respecting, there is something epistemically suspect about the claim that she can also thereby come to *discover* her worth and know herself as self-respecting. The act of commitment, the challenge holds, is *self-fulfilling* in a way that either rules discovery out, or else makes it trivial and empty.

To illustrate with an example, I can *discover* that I can stand on one leg and touch my nose with my eyes closed, for example, by attempting to do it. Genuine discovery is available here precisely because it is possible that I will fail. By contrast, it seems strange to say I can *discover* or *come to know* that I am quitting drinking by hearing myself pledge that I am, or that my declaration of love *reveals to me* that I love you. I suggest that what makes it sound strange is the following: Provided an audience is place, and provided I don't stumble on my words or have a coughing fit, it seems my attempt to commit myself is *guaranteed to succeed*. This self-fulfilling character sets commitment apart from other kinds of practical achievement, and threatens to undermine the idea that there is space for a genuine epistemic enterprise here, a possibility of gaining any substantive insight in the process.

How, then, can commitment both create *and* reveal? While I believe this to be a deep and difficult problem which I cannot address exhaustively in this paper, I want to make some proposals to take us in the right direction. In the preceding section, I suggested that the (or at least some of the) special normative powers of commitment depend on its publicity, viz. on its being made before an audience, and designed to be intelligible to them. In addition to these audience-directed success conditions, I now want to highlight some conditions pertaining to the committer and her state of mind. For there are also, I will suggest, audience-*in*dependent ways for attempts at self-commitment to fail, and so room for real self-discovery after all.

The key point is that commitments can be made half-heartedly or ambivalently. What I mean by this is not that one can enter into them reluctantly or conflictedly; this is entirely true, but it need not make the commitment itself half-hearted. I can have significant misgivings about taking on some task you are entrusting to me (thinking perhaps that it is unwise to do such a thing, or that it is unlikely I shall be able to carry it out to your satisfaction), but nonetheless accept it with perfect determination to do my best to deliver. Equally, I can have willingly accepted some responsibility, but now come to regret that I did, and wish I could be free of it. In either of these cases, though I am conflicted about carrying out the task, I also fully and unwaveringly committed to considering it my task.[106] The strength of the commitment to the task is recalcitrantly immune to misgivings about having the task.

The type of ambivalence I am after is a different one. It is the sense in which a purported or attempted commitment can be undermined from the first-person perspective when one finds, as we colloquially put it, that one is merely *going through the motions*, or that *one's heart is not really in it.*

In this respect, taking responsibility for oneself in the sense I have been considering here differs from taking responsibility for some action or consequence – at least insofar as the latter typically involves other people one thereby becomes responsible to, whereas the former typically does not. The contrast is this: I can make myself responsible for dealing with my child's mess even if only because I think that is what a parent *ought* to do. That is, I can *perform the part* of a responsible parent and therein make myself responsible all the while internally rejecting that very same norm of responsible parenting, or I think myself a fraud for outwardly portraying it. Being a responsible parent, then, is the sort of thing one does *with respect to* or *in relation to* someone else – to one's child, or to those affected by the mess that child has made- and to a concrete task at hand – such as cleaning the mess. Doing it is not undermined by internal ambivalence, half-heartedness, or the sense that one is merely going through the motions.

When it comes to taking responsibility for the person one is and is becoming, however, such internal ambivalence, conflict, or doubt, if sufficiently forceful and persistent, are deleterious to the attempt. But equally, one can find that one successfully evades or suppresses them. As I stand up and express my protest, I may find that my words ring true. I may find in myself what I hoped, but did not know, was there. The possibility of this discovery, I suggest, is the reason one sometimes has to stick one's neck out, voice one's resentment, and risk suffering the consequences.

[106] Indeed, it is precisely the fact that something is, from my perspective, unambiguously *my task* that leaves me conceptual room to regret that I have it, rue taking it up, and so on.

Conclusion

I started this chapter with the idea that there is a connection between self-respect and protest. This turned into the explanatory challenge I labelled Boxil's Puzzle: in the face of apparent challenges to one's worth, how could protest be a way to show and know oneself as self-respecting?

I then considered two accounts of the basis of human worth, The Dignity View and the Honour View, but argued that neither can address the puzzle, or even make much sense of the situation in which it arises. They render the connection between one's worth and one's knowledge of it either too tight or too tenuous to fit the Puzzle's structure.

Taking up the idea that a person's worth can be 'up to her', I developed an alternative account, drawing resources from the proposal that a person can *make* herself responsible for something by *taking* responsibility for it. In addition to actions and their consequences, I suggested, a person can also take responsibility for herself – for the person she is or is becoming – and thereby make her own worth. She can do this through an act of public commitment – standing up for herself by standing up before others.

Like the Honour view, this account rejects the characterisation of worth as static, unconditional, and independent of action; but like the Dignity view, it avoids making the claim to worth mere bravado or efficacious performance; on my view, it is the character of my action itself, rather than the impact it has upon its target audience, that makes the claim to self-respect true.

I then considered the worry that the apparently self-fulfilling character of a commitment prevents it from being a source of real discovery, and suggested that we can address the worry by recognising that commitments to being or becoming some way can fail to take hold if they are half-hearted, or, in the contrary case, turn out to resonate and ring true in unexpected ways. If correct, this account allows that standing up for oneself in protest can be both self-creation and self-discovery.

Conclusion

At the outset of this thesis, I proposed that we should not regret our manifest commitment to apology, or wish that we never had recourse to using it. Apologising, I have argued, and the practices of repair of which it forms a central part, are of a piece with our most fundamental modes of interpersonal interaction. Repair is not a break from the ordinary, well-adjusted flow of social and moral life, but itself an expression of it. To seek to detach or withdraw oneself from these practices would be to withdraw from the rich and meaningful fabric of a life lived with others.

Of course, we can perfectly well have occasions to regret the transgressions we commit or suffer under; and we can sometimes celebrate the ability to simply set these aside unilaterally, and forgive without condition or delays; we can also take heart from the fact that we sometimes find others with whom we have a spontaneous and entirely effortless understanding, one that arises and persists entirely without the use of words. As such, it can sometimes be both true and good that love means never having to say you are sorry. But it is a serious mistake to try to draw from these cases a general moral standard to aspire to. Though it is sometimes both demanding and perilous, and though it can go wrong or be abused, the pursuit of repair is a vital and valuable moral resource. Repair, I have argued, helps us to construct and reconstruct the terms of interpersonal morality, and to find ourselves and each other in the times of greatest need.

Repair, as I have characterised it, is an extended and collective undertaking, a project pursued by communicative means, and without a determinate end-point or conclusion. Its terms and scope cannot be set in advance, but must themselves be negotiated and discovered along the way. This generative or productive character is only surprising if one assumes that our patterns of moral sense-making are themselves static and unchanging, or that the development of good moral agency is fundamentally autonomous and does not, or ought not, require scaffolding and learning from others[107].

Feminist moral philosophers have long opposed these assumptions of agential self-sufficiency and independence, and the idea of moral norms as timeless and placeless[108]. They have also critiqued in myriad ways what I have described as a syndrome of methodological and substantive commitments that prioritise notions like individual autonomy, rights to property, and freedom from interference, and which occludes or renders suspicious alternative ways of thinking and being.

I mean for the present project to join with these efforts in trying to tread a different path through the moral and social terrain. The proposals I have made concerning our responses to wrongdoing, our epistemic and communicative agency, and our need and norms for redress have twin aspirations: first, to provide accounts that are more descriptively adequate than those they replace; second, to open new possibilities for moral thought and action.

It is worth briefly surveying some of the results I have arrived at and some of the further possibilities I hope to pursue in future work flowing from this project.

In chapters One and Two, I laid out my conception of the problem of repair and evaluated a prominent approach to addressing it. Though there are many ways of responding to wrongdoing, whether committed by us or against us, we often experience the reality of past transgressions as setting a kind of practical task for wrongdoers in particular. When someone has done wrong, we sometimes find it both possible and desirable that they do something about it. Moreover, the execution of this task seems to involve the familiar communicative practice of apologising. An account capable of filling this gap, I argued, must offer us a conception of apology and repair that gives us a unified mechanism, and is demonstrably both explanatorily and ethically adequate.

[107]For accounts that highlight the dynamic and interactive character of moral understanding and moral agency, see especially (McGeer, 2018, 2015a, 2015b; Springer, 2013; Walker, 2007)

[108] For surveys of this work, see (Walker, 2007)

The bookkeeping model promises to do this. Drawing from the conceptual framework and commitments of legal and political normativity, it characterises wrongdoing as an unjust loss or deprivation for the victim (and/or a corresponding unjust gain for the wrongdoer), which disturbs a pre-existing balance or equilibrium and must therefore be offset by subsequent restitution through compensation or repayment. I surveyed in detail three versions of a bookkeeping approach to repair – the penance view, the self-denigration view, and the reassurance view – finding each to be wanting. They either left significant explanatory gaps, or imposed unpalatable moral costs, or both of these at once.

This lead to an unattractive conclusion: if repair is what the bookkeeping model tells us it is, we have good reason to abandon or fundamentally revise it. The conception of repair the bookkeeping model seems to offer paints it as troublingly fastidious, rationally incoherent

While developing these critiques, I also drew into question the bookkeeping model's way of borrowing normative priorities, commitments, and standards from legal and political normativity to apply them in the informal interpersonal setting. In extending this work, it would be interesting to consider how the criticisms I have made in this context would scale up when we turn our attention away from the relatively minor interpersonal wrongs that are my primary focus and onto large-scale, social-institutional responses to wrongdoing I otherwise mostly set aside: responses to genocide, political persecution and oppression, systematic and structural injustice, and assaults on human dignity.

My stance in this thesis has been that cases like these pose profound difficulty for theorising, and should for that reason not be our starting point. All the same however, they do, tragically, occur, and make the need for repair only more urgent. What lessons can be drawn for these contexts from my critique of the bookkeeping model's blind spots and shortcomings?

I would also be able to give more consideration to the contest between alternative normative approaches that have been pursued in legal and political philosophy. My investigation in Chapter Two restricts itself to substantive conceptions of distributive and retributive justice; but legal and political philosophy also has rich traditions for procedural and relational justice. Though I have found that these ideas have exerted less influence on moral philosophy than their alternatives, their explicit inclusion in the inquiry would undoubtedly enrich our explanatory and normative options, and animate new directions for research.

In chapters Three and Four, I turned my attention to the epistemic and agential character of communication. In chapter Three, I take as my starting point Richard Moran's account of testimony as a social act. While sympathetic to Moran's overall project, I argue that he misconstrues the testimony's social character. Social acts, I propose, further divide into two sub-categories: Active and Passive. Whereas the performance of the former requires the active and voluntary reciprocation from another party, the latter require only their understanding and awareness – something one cannot ordinarily withhold at will. Drawing on a phenomenon I label the impossibility of testimonial resistance, I then argue that telling belongs to the latter category, and therefore that Moran is mistaken in characterising it as a phenomenon in which the speaker and hearer are collaborative partners.

In chapter Four, however, I resuscitate the idea of speech as collaborative in a different guise. The key insight, I suggest, comes from recognising the difference between *simple* and *complex* testimony: much of our speech – and much of our testimony – does not take the form of single self-standing utterances, but is rather delivered in extended ordered discourses, sometimes proceeding over several conversational turns. When it comes to extended sequences of speech in interaction, hearers gain a number of distinctive collaborative roles that they do not have in a speaker's delivery of simple testimony. To identify and explain these, I turned to findings from the empirical conversation sciences, including in socio- and psycho-linguistics and in conversation analysis. Hearers' active responsiveness in communication, I showed, both facilitate, enrich, and direct speakers' communicative contributions in ways philosophers have generally failed to recognise and consider.

The ideas presented in these chapters set at least three directions for further research. First, Chapter Four makes the case for including the findings of empirical conversation research as a resource for philosophical theorising and in particular as a way of 'drawing the coverts of the microglot' in J. L. Austin's phrase – making the minutiae of actual ordinary language use apparent and available for theorising. Insofar as philosophers wish to base their accounts and ideas on how language is actually used in practice, we need to note the systematic divergences between what we imagine and what actually occurs, and between language use in writing and language use in speech. Though I here focus my attention on the consequences of these differences to testimony in particular, it is natural to assume that they could also have ramifications in other areas, such as in the justification and agreement procedures upon which conceptions of Public Reason traditions rely[109].

Second, the account of testimony I propose itself sets the stage for the development of what might be termed a *procedural* social epistemology: a programme of research that departs from the typical focus on individual utterances and instead investigates the distinctive epistemic character of sequences of jointly produced talk. Such a programme has the potential to call for radical revisions to received ideas about the interpretation of testimony, its capacity to convey knowledge warrants, the determination of a speaker's credibility, and the responsibility conversation partners have for what gets said between them. It also poses profound challenges to notions of speaker autonomy and the substitutability of interlocutors that call for further inquiry.

Finally, the account of the social character of communicative agency itself raises further questions about political obligation and the promotion of deliberative democracy. If, as I argue, communicative agency is socially articulated, do states have the responsibility to secure the deliberative enfranchisement of their citizens? Which environmental conditions (whether institutional or technological) best promote high quality and democratically legitimate engagement? How far can these conditions be artificially generated or sustained by policymakers? And if communicative agents are not substitutable for one another, but bring distinctive abilities and capacities to the table, do individual citizens owe their deliberative participation to their fellows?

[109] See e.g. (Quong, 2022)

In Chapter Five, I developed my alternative conception of repair as the pursuit of mutual understanding. Arguing that we can take the notion of communicative repair introduced in an earlier chapter as an apt and illuminating model, I also sought to demonstrate that the position that results nicely explains familiar aspects of the underlying phenomena, including our occasional choice *not* to pursue repair, and the insurmountable difficulties we sometimes have encounter in the pursuit, even when the parties earnestly wish to reconcile.

Finally, in Chapter Six, I turned my attention to the case of a victim of unjust wrongdoing, who has no expectation of repair, and no power to redress the injustice by her own efforts alone. I proposed that a person in this situation nevertheless has a reason to protest her mistreatment, namely that she can thereby come to show and know herself as self-respecting. Arguing that neither of two common positions on the basis of self-respect can explain how this would be so, I instead develop an alternative account. I propose that a victim of injustice can create the basis for her own self-respect by undertaking a public commitment. She stands up for herself by standing up before others.

The ideas developed in this final part of the thesis suggest the possibility of an interconnected approach to personal and social normative development. In both interpersonal repair and protest against unjust treatment, I propose, we find an intriguing generative potential for self-understanding and -creation, and resources for social normative contestation.

Bibliography

Adams, R.M., 1985. Involuntary Sins. Philosophical Review 94, 3–31. https://doi.org/10.2307/2184713

AFI's 100 YEARS…100 MOVIE QUOTES [WWW Document], n.d. . American Film Institute. URL https://www.afi.com/afis-100-years-100-movie-quotes/ (accessed 11.20.24).

Anderson, E., 1999. What is the Point of Equality. Ethics 109, 287–337. https://doi.org/10.1086/233897

Anderson, E., Pildes, R., 2000. Expressive Theories of Law: A General Restatement. University of Pennsylvania Law Review 148, 1503.

Arendt, H., 1958. The human condition. [Chicago] : University of Chicago Press.

Austin, J.L., 1975. How to Do Things with Words: Second Edition. Harvard University Press.

Austin, J.L., 1957. I.–A Plea for Excuses: The Presidential Address. Proceedings of the Aristotelian Society 57, 1–30. https://doi.org/10.1093/aristotelian/57.1.1

Baier, A., 1997. The Commons of the Mind. Open Court Publishing.

Baier, A., 1986. Trust and Antitrust. Ethics 96, 231–260. https://doi.org/10.1086/292745

Bavelas, J.B., 2022a. Appreciating Face-to-face Dialogue, in: Bavelas, J.B. (Ed.), Face-to-Face Dialogue: Theory, Research, and Applications. Oxford University Press, p. 0. https://doi.org/10.1093/oso/9780190913366.003.0001

Bavelas, J.B., 2022b. Meaning and Understanding as an Interactional Process, in: Bavelas, J.B. (Ed.), Face-to-Face Dialogue: Theory, Research, and Applications. Oxford University Press, p. 0. https://doi.org/10.1093/oso/9780190913366.003.0009

Bavelas, J.B., Coates, L., Johnson, T., 2002. Listener Responses as a Collaborative Process: The Role of Gaze. Journal of Communication 52, 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

Bavelas, J.B., Coates, L., Johnson, T., 2000. Listeners as co-narrators. Journal of Personality and Social Psychology 79, 941–952. https://doi.org/10.1037/0022-3514.79.6.941

Bavelas, J.B., Gerwing, J., 2011. The Listener as Addressee in Face-to-Face Dialogue. International Journal of Listening 25, 178–198. https://doi.org/10.1080/10904018.2010.508675

Bennett, C., 2022. What Goes on When We Apologize? Journal of Ethics and Social Philosophy 23. https://doi.org/10.26556/jesp.v23i1.1294

Bennett, C., 2012. Précis of" the Apology Ritual". Teorema: International Journal of Philosophy 31, 73–94.

Bennett, C., 2008. The Apology Ritual: A Philosophical Theory of Punishment. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511487477

Bertrand, R., Ferré, G., Blache, P., Espesser, R., Rauzy, S., 2007. Backchannels revisited from a multimodal perspective, in: Auditory-Visual Speech Processing. Hilvarenbeek, Netherlands, pp. 1–5.

Bittner, R., 1992. Is It Reasonable to Regret Things One Did? The Journal of Philosophy 89, 262. https://doi.org/10.2307/2027168

Bovens, L., 2008. XII-Apologies. Proceedings of the Aristotelian Society (Hardback) 108, 219–239. https://doi.org/10.1111/j.1467-9264.2008.00244.x

Boxill, B.R., 2010. The Responsibility of the Oppressed to Resist Their Own Oppression. Journal of Social Philosophy 41, 1–12. https://doi.org/10.1111/j.1467-9833.2009.01474.x

Boxill, B.R., 1992. Blacks and Social Justice. Rowman & Littlefield.

Boxill, B.R., 1976. Self-Respect and Protest. Philosophy and Public Affairs 6, 58–69.

Brandom, R., 1980. Asserting. Journal of Philosophy 77, 766–767. https://doi.org/10.5840/jphil1980771117

Bruner, J.S., 2003. Making Stories: Law, Literature, Life. Harvard University Press.

Calhoun, C., 2019. XI—Responsibilities and Taking on Responsibility. Proceedings of the Aristotelian Society 119, 231–251. https://doi.org/10.1093/arisoc/aoz017

Calhoun, C., 1995. Standing for Something. Journal of Philosophy 92, 235–260.

Calhoun, C., 1992. Changing One's Heart. Ethics 103, 76–96. https://doi.org/10.1086/293471

Callard, A., 2020. On Anger. Boston Review 8–28.

Caponetto, L., 2021. A Comprehensive Definition of Illocutionary Silencing. Topoi 40, 191–202. https://doi.org/10.1007/s11245-020-09705-2

Cavell, S., 1999. The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy. Oxford University Press, USA.

Clark, H.H., 1996. Using Language, "Using" Linguistic Books. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511620539

Coady, C.A.J., 1992. Testimony: A Philosophical Study. Oxford University Press, New York.

Code, L., 2011. An Ecology of Epistemic Authority. Episteme 8, 24–37. https://doi.org/10.3366/epi.2011.0004

Code, L., 1995. Rhetorical spaces: essays on gendered locations. Routledge, New York.

Code, L., 1991. Second Persons, in: What Can She Know?, Feminist Theory and the Construction of Knowledge. Cornell University Press, pp. 71–109.

Craig, E., 1990. Knowledge and the State of Nature: An Essay in Conceptual Synthesis. Clarendon Press, Oxford, GB.

Darwall, S., 2015. Respect as Honor and as Accountability, in: Reason, Value, and Respect. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199699575.003.0004

Darwall, S., 2006. The Second-Person Standpoint: Morality, Respect, and Accountability. Harvard University Press.

Darwall, S.L., 1977. Two Kinds of Respect. Ethics 88, 36–49.

Davidson, D., 2001. Essays on Actions and Events (2nd edition). Clarendon Press, Oxford.

Dillon, R.S., 2020. Humility and Self-respect: Kantian and feminist perspectives 1, in: The Routledge Handbook of Philosophy of Humility. Routledge.

Dillon, R.S., 1997. Self-Respect: Moral, Emotional, Political. Ethics 107, 226–249.

Dillon, R.S., 1992. How to Lose Your Self-Respect. American Philosophical Quarterly 29, 125–139.

Dotson, K., 2011. Tracking Epistemic Violence, Tracking Practices of Silencing. Hypatia 26, 236–257. https://doi.org/10.1111/j.1527-2001.2011.01177.x

Dover, D., 2023a. Identity and Influence. Synthese 202, 1–24. https://doi.org/10.1007/s11229-023-04279-z

Dover, D., 2023b. Two Kinds of Curiosity. Philosophy and Phenomenological Research 108, 811–832. https://doi.org/10.1111/phpr.12976

Dover, D., 2022. The Conversational Self. Mind 131, 193–230. https://doi.org/10.1093/mind/fzab069

Dover, D., 2019. The Walk and the Talk. Philosophical Review 128, 387–422. https://doi.org/10.1215/00318108-7697850

Duff, R.A., 2006. Iv-Answering for Crime. Proceedings of the Aristotelian Society 106, 87–113. https://doi.org/10.1111/j.1467-9264.2006.00140.x

Ebels-Duggan, K., 2013. Dealing with the Past: Responsibility and Personal History. Philosophical Studies 164, 141–161. https://doi.org/10.1007/s11098-013-0090-1

Ekins, R., 2012. Equal Protection and Social Meaning. American Journal of Jurisprudence 57, 21–48. https://doi.org/10.1093/ajj/57.1.21

Enfield, N.J., 2017. How We Talk: The Inner Workings of Conversation. Basic Books.

Enoch, D., 2011. Being Responsible, Taking Responsibility, and Penumbral Agency, in: Heuer, Lang (Eds.), Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams. Oxford University Press, Usa.

Fakhoury, T., 2021. Quiet Resistance: The Value of Personal Defiance. J Ethics 25, 403–422. https://doi.org/10.1007/s10892-020-09356-w

Feinberg, J., 1970. Doing & Deserving; Essays in the Theory of Responsibility. Princeton University Press, Princeton, N.J.,.

Fraser, R., 2021. Narrative Testimony. Philosophical Studies 178, 4025–4052. https://doi.org/10.1007/s11098-021-01635-y

Fricker, M., 2018. Ambivalence About Forgiveness. Roy. Inst. Philos. Suppl. 84, 161–185. https://doi.org/10.1017/S1358246118000590

Fricker, M., 2007. Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press, New York.

Gaita, R., 2004. Good and Evil: An Absolute Conception. Routledge.

Gardner, J., 2013. What is Tort Law for? Part 2. The Place of Distributive Justice (SSRN Scholarly Paper No. 2269615). Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.2269615

Gardner, J., 2011. What is Tort Law For? Part 1. The Place of Corrective Justice. Law and Philosophy 30, 1–50. https://doi.org/10.1007/s10982-010-9086-6

Garrard, E., McNaughton, D., 2003. In Defence of Unconditional Forgiveness. Proceedings of the Aristotelian Society 103, 39–60. https://doi.org/10.1111/1467-9264.00127

Gert, H.J., Radzik, L., Hand, and M., 2004. Hampton on the Expressive Power of Punishment. Journal of Social Philosophy 35, 79–90. https://doi.org/10.1111/j.1467-9833.2004.00217.x

Gilligan, C., 2016. In a Different Voice: Psychological Theory and Women's Development. Harvard University Press.

Goffman, E., 2009. Relations in Public. Transaction Publishers.

Govier, T., 1999. What is acknowledgement and why is it important? OSSA Conference Archive.

Grice, H.P., 1957. Meaning. The Philosophical Review 66, 377. https://doi.org/10.2307/2182440

Griswold, C., 2006. Forgiveness: A Philosophical Exploration. Cambridge University Press, New York.

H, Y.V., 1970. On getting a word in edgewise. Papers from the sixth regional meeting Chicago Linguistic Society, April 16-18, 1970, Chicago Linguistic Society, Chicago 567–578.

Hampton, J., 1991. Correction Harms versus Righting Wrongs: The Goal of Retribution. UCLA L. Rev. 39, 1659–1702.

Hart, H.L.A., 1968. Punishment and Responsibility: Essays in the Philosophy of Law. Oxford University Press.

Hay, C., 2011. The Obligation to Resist Oppression. Journal of Social Philosophy 42, 21–45. https://doi.org/10.1111/j.1467-9833.2010.01518.x

Helmreich, J.S., 2015. The Apologetic Stance: The Apologetic Stance. Philos Public Aff 43, 75–108. https://doi.org/10.1111/papa.12053

Hieronymi, P., 2001. Articulating an Uncompromising Forgiveness. Philosophy and Phenomenological Research 62, 529–555. https://doi.org/10.1111/j.1933-1592.2001.tb00073.x

Hill, T.E., 1982. Self-Respect Reconsidered. Tulane Studies in Philosophy 31, 129–137. https://doi.org/10.5840/tulane1982319

Hill, T.E., 1973. SERVILITY AND SELF-RESPECT. The Monist 57, 87–104.

Hughes, P.M., Warmke, B., 2024. Forgiveness, in: Zalta, E.N., Nodelman, U. (Eds.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Jankélévitch, V., Hobart, A., 1996. Should We Pardon Them? Critical Inquiry 22, 552–572.

Kolnai, A., 1974. Forgiveness. Proceedings of the Aristotelian Society 74, 91–106. https://doi.org/10.1093/aristotelian/74.1.91

Korta, K., Perry, J., 2024. Pragmatics, in: Zalta, E.N., Nodelman, U. (Eds.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Kukla, R., Lance, M., 2009. ?Yo!? and ?Lo!?: The Pragmatic Topography of the Space of Reasons. Harvard University Press.

Lance, M., Kukla, R., 2013. Leave the Gun; Take the Cannoli! The Pragmatic Topography of Second-Person Calls. Ethics 123, 456–478. https://doi.org/10.1086/669565

Lerner, H.G., 2017. Why won't you apologize? : healing big betrayals and everyday hurts. London : Duckworth Overlook.

Levy, K., 2014. Why Retributivism Needs Consequentialism: The Rightful Place of Revenge in the Criminal Justice System. Rutgers Law Review 66, 629–684.

Lewis, D., 1969. Convention: A Philosophical Study. Synthese 26, 153–157.

Lugones, M., 1987. Playfulness, ?World?-Travelling, and Loving Perception. Hypatia 2, 3–19. https://doi.org/10.1111/j.1527-2001.1987.tb01062.x

Maitra, I., 2009. Silencing Speech. Canadian Journal of Philosophy 39, 309–338. https://doi.org/10.1353/cjp.0.0050

Maitra, I., 2004. Silence and Responsibility. Philosophical Perspectives 18, 189–208. https://doi.org/10.1111/j.1520-8583.2004.00025.x

Martin, A.M., 2010. Owning Up and Lowering Down: The Power of Apology. Journal of Philosophy 107, 534–553. https://doi.org/10.5840/jphil20101071037

Marusić, B., 2022. On the Temporality of Emotions: An Essay on Grief, Anger, and Love. Oxford University Press, Oxford.

Mason, E., 2019. Ways to be Blameworthy: Rightness, Wrongness, and Responsibility. Oxford University Press. https://doi.org/10.1093/oso/9780198833604.001.0001

McDonald, L., 2021. Cat-Calls, Compliments and Coercion. Pacific Philosophical Quarterly 103, 208–230. https://doi.org/10.1111/papq.12385

McGeer, V., 2022. Moral travel and the narrative work of forgiveness: A Tribute to Ronald de Sousa. A Tribute to Ronald de Sousa.

McGeer, V., 2018. Scaffolding Agency: A Proleptic Account of the Reactive Attitudes. European Journal of Philosophy 27, 301–323. https://doi.org/10.1111/ejop.12408

McGeer, V., 2015a. Mind-Making Practices: The Social Infrastructure of Self-Knowing Agency and Responsibility. Philosophical Explorations 18, 259–281. https://doi.org/10.1080/13869795.2015.1032331

McGeer, V., 2015b. Building a better theory of responsibility. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 172, 2635–2649.

McGowan, M.K., 2019. Just Words: On Speech and Hidden Harm. Oxford University Press.

McKenna, M., 2011. Conversation and Responsibility. Oxford University Press USA, , US.

Medina, J., 2013. Active Ignorance, Epistemic Others, and Epistemic Friction, in: Medina, J. (Ed.), The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination. Oxford University Press, p. 0. https://doi.org/10.1093/acprof:oso/9780199929023.003.0001

Moodyadams, M.M., 1993. Race, Class, and the Social Construction of Self-Respect. Philosophical Forum 24, 251–266.

Moore, M.S., 1988. The Moral Worth of Retribution, in: Schoeman, F. (Ed.), Responsibility, Character, and the Emotions: New Essays in Moral Psychology. Cambridge University Press, Cambridge, pp. 179–219. https://doi.org/10.1017/CBO9780511625411.008

Moran, R., 2019. The Exchange of Words: Replies to critics. Eur J Philos 27, 786–795. https://doi.org/10.1111/ejop.12491

Moran, R., 2018. The Exchange of Words: Speech, Testimony, and Intersubjectivity. Oup Usa, New York City.

Moran, R., 2005. Getting Told and Being Believed. Philosophers' Imprint 5, 1–29.

Morris, H., 1971. Guilt and Suffering. Philosophy East and West 21, 419–434. https://doi.org/10.2307/1398170

Morris, H., 1968. Persons and Punishment. The Monist 52, 475–501. https://doi.org/10.5840/monist196852436

Murphy, J.G., 2007. Legal moralism and retribution revisited. Criminal Law, Philosophy 1, 5–20. https://doi.org/10.1007/s11572-006-9000-3

Murphy, J.G., Hampton, J., 1988. Forgiveness and Mercy, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511625121

Norlock, K., 2008. Forgiveness From a Feminist Perspective. Lexington Books.

Nozick, R., 1974. Anarchy, State, and Utopia. Basic Books, New York.

Nussbaum, M.C., 2016. Anger and Forgiveness: Resentment, Generosity, Justice. Oxford University Press, New York.

Pereboom, D., 2012. Free Will Skepticism, Blame, and Obligation, in: Coates, D.J., Tognazzini, N.A. (Eds.), Blame: Its Nature and Norms. Oxford University Press, p. 0. https://doi.org/10.1093/acprof:oso/9780199860821.003.0010

Pettigrove, G., 2007. Understanding, Excusing, Forgiving. Philosophy and Phenomenological Research 74, 156–175. https://doi.org/10.1111/j.1933-1592.2007.00007.x

Quong, J., 2022. Public Reason, in: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

Radzik, L., 2009a. Making Amends: Atonement in Morality, Law, and Politics, Making Amends. Oxford University Press.

Radzik, L., 2009b. Repaying Moral Debts: Self-Punishment and Restitution, in: Radzik, L. (Ed.), Making Amends: Atonement in Morality, Law, and Politics. Oxford University Press, p. 0. https://doi.org/10.1093/acprof:oso/9780195373660.003.0002

Radzik, L., 2009c. Chapter Four Reforming Relationships: The Reconciliation Theory of Atonement, in: Radzik, L. (Ed.), Making Amends: Atonement in Morality, Law, and Politics. Oxford University Press, p. 0. https://doi.org/10.1093/acprof:oso/9780195373660.003.0004

Raz, J., 2011. Being in the World, in: Raz, J. (Ed.), From Normativity to Responsibility. Oxford University Press, p. 0. https://doi.org/10.1093/acprof:oso/9780199693818.003.0012

Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., Glazer, T., Knox, D., Liebscher, A., Marin, S., 2023. The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation. Science Advances 9, eadf3197. https://doi.org/10.1126/sciadv.adf3197

Ross, A., 1986. Why Do We Believe What We Are Told? Ratio 69–88.

Rushdy, A.H.A., 2015. A Guilted Age: Apologies for the Past. Temple University Press, Philadelphia.

Scanlon, T., 1990. Promises and Practices. Philosophy and Public Affairs 19, 199–226.

Schegloff, E.A., 1992. Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. American Journal of Sociology 97, 1295–1345.

Schegloff, E.A., Jefferson, G., Sacks, H., 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. Language 53, 361–382. https://doi.org/10.2307/413107

Schelling, T.C., 1980. The Strategy of Conflict: With a New Preface by the Author. Harvard University Press.

Segal, E., 1970. Love Story. Buccaneer Books.

Shelby, T., 2012. The Ethics of Uncle Tom's Children. Critical Inquiry 38, 513–532. https://doi.org/10.1086/664549

Shelby, T., 2010. Reflections on Boxill's Blacks and Social Justice. Journal of Social Philosophy 41, 343–353. https://doi.org/10.1111/j.1467-9833.2010.01494.x

Sliwa, P., 2019. The Power of Excuses. Philosophy & Public Affairs 47, 37–71. https://doi.org/10.1111/papa.12139

Sliwa, P., 2010. Excuse Without Exculpation: The Case of Moral Ignorance, in: Shafer-Landau, R. (Ed.), Oxford Studies in Metaethics. Oxford University Press, pp. 72–95.

Sommers, T., 2007. The Objective Attitude. Philosophical Quarterly 57, 321–341. https://doi.org/10.1111/j.1467-9213.2007.487.x

Springer, E., 2013. Communicating Moral Concern: An Ethics of Critical Responsiveness. MIT Press.

Stalnaker, R., 2014. Context. Oxford University Press, Oxford.

Stark, C.A., 1997. The Rationality of Valuing Oneself: A Critique of Kant on Self-Respect. Journal of the History of Philosophy 35, 65–82. https://doi.org/10.1353/hph.1997.0006

Strawson, P., 1962. Freedom and Resentment. Proceedings of the British Academy 48, 187–211.

Swinburne, R., 1989. Guilt, Atonement, and Forgiveness, in: Responsibility and Atonement. Oxford University Press, Oxford. https://doi.org/10.1093/0198248490.003.0006

Tavuchis, N., 1991. Mea culpa: A sociology of apology and reconciliation, Mea culpa: A sociology of apology and reconciliation. Stanford University Press.

Tolins, J., Fox Tree, J.E., 2014. Addressee backchannels steer narrative development. Journal of Pragmatics 70, 152–164. https://doi.org/10.1016/j.pragma.2014.06.006

van Roojen, M., 2020. Promising and Assertion, in: Goldberg, S. (Ed.), The Oxford Handbook of Assertion. Oxford University Press, pp. 178–200. https://doi.org/10.1093/oxfordhb/9780190675233.013.39

Waldron, J., 2012. Dignity, Rank, and Rights. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199915439.001.0001

Walker, M.U., 2014. Moral Vulnerability and the Task of Reparations.

Walker, M.U., 2013. Third Parties and the Social Scaffolding of Forgiveness. Journal of Religious Ethics 41, 495–512. https://doi.org/10.1111/jore.12026

Walker, M.U., 2007. Moral understandings: a feminist study in ethics, 2nd ed. ed, Studies in feminist philosophy. Oxford University Press, New York.

Walker, M.U., 2006a. Moral Repair: Reconstructing Moral Relations after Wrongdoing 264.

Walker, M.U., 2006b. Restorative Justice and Reparations. J Social Philosophy 37, 377–395. https://doi.org/10.1111/j.1467-9833.2006.00343.x

Walker, M.U., 1997. Moral Understandings: A Feminist Study in Ethics. Routledge, New York, US.

Wanderer, J., 2010. Inhabiting the Space of Reasoning. Analysis 70, 367–378. https://doi.org/10.1093/analys/anp147

Warmke, B., 2016. The Economic Model of Forgiveness: The Economic Model of Forgiveness. Pacific Philosophical Quarterly 97, 570–589. https://doi.org/10.1111/papq.12055

Watson, G., 2004. Asserting and Promising. Philosophical Studies 117, 57–77. https://doi.org/10.1023/B:PHIL.0000014525.93335.9e

Watson, G., 1993. 4. Responsibility and the Limits of Evil: Variations on a Strawsonian Theme, in: Fischer, J.M., Ravizza, M. (Eds.), Perspectives on Moral Responsibility. Cornell University Press, pp. 119–148.

W.E.B. Du Bois, 1973. Thoughts and Ideas at the Turn of the Twentieth Century, in: William M. Tuttle (Ed.), W. E. B. Du Bois. Prentice Hall.

W.E.B. Du Bois, 1966. Of Mr. Booker T. Washington and Others, in: Howard Brotz (Ed.), Negro Social and Political Thought,1850-1920 : Representative Texts. Basic Books, New York.

Williams, B., 2002. Truth and Truthfulness: An Essay in Genealogy. Princeton University Press, Princeton.

Williams, B., 1981. Moral Luck: Philosophical Papers 1973?1980. Cambridge University Press, New York.

Williams, B.A.O., Nagel, T., 1976. Moral Luck. Proceedings of the Aristotelian Society, Supplementary Volumes 50, 115–151.

Wolf, S., 1990. Freedom Within Reason. Oup Usa, New York.

Young, I.M., 2011. Responsibility for Justice. Oxford University Press USA, , US.

Young, I.M., 1990. Justice and the Politics of Difference. Princeton University Press.