Title
More than Words: Stances as an Alternative Model for Apology, Forgiveness and Similar Speech Acts

Author
Helmreich, Jeffrey Stuart

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

More than Words:

Stances as an Alternative Model for Apology, Forgiveness and Similar Speech Acts

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Philosophy

by

Jeffrey S. Helmreich

2013

ABSTRACT OF THE DISSERTATION

More than Words:

Stances as an Alternative Model for Apology, Forgiveness and Similar Speech Acts

By

Jeffrey S. Helmreich

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2013

Professor Seana Shiffrin, Chair

We have the power to make dramatic moral differences with words. In particular, certain speech acts – apologizing, forgiving, taking responsibility – change the moral dynamics between people, thereby restoring relationships, relieving moral debts and grounding historic reconciliation.

Few dispute this power, even as it continues to amaze us in practice. Yet, despite an illuminating burst of scholarly attention to many aspects of apology, forgiveness and moral repair in recent years, their power remains elusive. That is, we still have an incomplete grasp of what, exactly, about saying "I'm sorry," or "I forgive you" effects such significant moral change.

The dissertation seeks to understand and account for that power. At the same time, it also seeks a new account of these speech acts and their sincerity conditions, not only for individual

people but for corporate and institutional bodies, as well. These twin projects are joined by the same underlying conviction: to capture how these speech acts accomplish so much, we need a new understanding of what they are.

Chapters I and II begin this explanatory project by focusing on the classic case of apologies, taking issue with traditional accounts of apologetic expression as representing or revealing something – such as how the speaker feels or what she believes. These views cannot make sense of the way an apology responds remedially to a past wrongdoing, as illustrated most dramatically by cases where the victim knows everything the apology could reveal.

Instead, I argue that apologies, and similar speech acts, should be understood less as expressions than as ways of treating someone that counteract the mistreatment begun by the actions for which one apologizes. This model requires a new, relational understanding of both apologies and of the actions that give rise to them.

Chapter III focuses on the formal speech act of forgiveness, by which a victim can alter the moral status of her offender – rendering apologies and other acts of moral repair unnecessary, and their absence no longer blameworthy. I argue that forgiveness has this impact because, and to the extent that, it takes place in contexts in which the moral power of other remedial steps like apology are already at work. As with apologies, then, forgiveness emerges as less a unilateral expression than an interactive approach to another person, which can help restore their relationship.

Together the accounts present apologies, forgiveness and similar speech acts as active ways for people to relate to each other, whose sincerity depends more on commitments and dispositions to act than on emotions and psychological states. With this framework in place, it

becomes clear why even institutions – countries, courts, companies – can sincerely apologize, forgive and engage each other in similar speech acts, as I argue in Chapter IV.

The resulting picture of utterances like apology and forgiveness departs from the speech-action dichotomy that Austin and Searle began challenging half a century ago. On the account developed here, certain speech can function as action over and above what it communicates, while some actions have meaning beyond their material impact on the world. The area of moral repair, then, sheds new light on what action can mean and speech can do.

The dissertation of Jeffrey S. Helmreich is approved

Mark Greenberg

Barbara Herman

Pamela Hieronymi

David Kaplan

Herbert Morris

Seana Shiffrin, Committee Chair

University of California, Los Angeles

2013

To Alan Michael Helmreich, unforgettable inspiration

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Not long after I arrived at UCLA, I asked Seana Shiffrin to be my advisor. It has proved to be one of the best decisions I've ever made, not least in its impact on every aspect of the work that follows. Her contribution and influence is simply too great to try to capture here, except to say that it includes the stunning example she set as philosopher, teacher, and person. I cannot imagine a better and more dedicated advisor, or a more perfect model of how to synthesize intellect and principle.

Given my luck in choice of advisor, it should have been too much to expect to have, in the same department, an additional faculty supporter such as Barbara Herman. To my great fortune, she gave enormously to this project and to my journey here, from the transformative comments on my drafts to the thrilling talks in her office, always leaving me at once inspired and unsettled. My dissertation – and the way I try to do philosophy – owe immeasurably to Barbara.

Herbert Morris came later to this project, but I still feel guilty about having grabbed so much of his striking and insightful attention, which reshaped all my thinking about regret, remorse, moral repair and so much else. I'm equally grateful to Herb for sharing his paradigm-shifting thoughts with me in person, in a wonderful series of sessions across West Los Angeles that taught me so much and felt way too short.

All three of these teachers – my dissertation's principal readers – also impressed upon me the value of keeping moral philosophy sensitive to moral life and experience, rather than treating it as the mere application of conceptual analysis to ethical questions. I hope a trace of this orientation will be found in the dissertation, at least in its choice of method and emphasis.

My work also benefited greatly from the invaluable feedback and, at times, much-needed criticism of Mark Greenberg, who kept me rooted in the ideas that first attracted me to philosophy, and Pamela Hieronymi, whose inspiring work in moral repair and moral psychology proved as influential as her input at key junctures. Finally, David Kaplan challenged me to think beyond the categories typically applied to speech acts and their ethical contexts, and helped me fine-tune the notion of stance-taking that dominates the dissertation.

In developing the ideas that led up to this project, I also gained immeasurably from exchanging work with peers, especially Thi Nguyen, Mandel Cabrera, Jorah Dannenberg, Brent Kious and one or two others who I hope will forgive my memory lapse. At the same time, my project was enriched by regular meetings with David Goldman and Stephen White, in what Dave aptly dubbed the "The Moral Aftermath Group."

In addition, my thinking in general was stimulated and improved through the dissertation years by multiple talks with Howie Wettstein, Joseph Almog, John Carriero, Marc Cohen, Ari Lev, Julie Wulfemeyer, Alex Radulescu, Tiffany Teeman Cvrkel, Arudra Burra, David Plunkett, Jesse Summers, Elliot Michaelson, Louis-Philippe Hodgson and Sari Kisilevsky, among others.

I remain particularly grateful to my parents, Helaine and William Helmreich, along with my brother Joe and sister Deb, for their love and steady support of my decision to do what most fulfilled me, rather than what seemed practical.

Most of all I thank my wife, Esther, who proved that one could have both a best friend and most perceptive interlocutor right here at home. I depended more than I should admit on her boundless love, brilliance and support.

VITA

EDUCATION

Columbia University
A.B. *cum laude* 1997

Georgetown University Law Center
J.D. *magna cum laude* 2004

University of California, Los Angeles
M. Phil 2009

EMPLOYMENT

Harvard Law School Program on Negotiation
Visiting Scholar 2012-13

University of California, Los Angeles
Teaching Associate 2009-11
Teaching Assistant 2006-09

United States District Court, Northern District of Georgia
Judicial Law Clerk 2004-05

PUBLICATIONS

Does 'Sorry' Incriminate? Evidence, Harm and the Protection of Apologies
Cornell Journal of Law and Public Policy 21:3 (2012), pp. 567-609

Putting Down: Expressive Subordination and Equal Protection
U.C.L.A. Law Review Discourse 59 (2012), pp. 112-27

PRESENTATIONS

"The Metaphysics of Reconciliation: How Forgiveness Relieves Moral Debt"
American Philosophical Association,
Pacific Division Meeting, San Francisco, CA — April, 2013

"The Problems of Agent Regret"
University of California, Irvine — February, 2012

"Regret, Remorse and Accidents:
Where the New Apology Laws Go Wrong"
Law and Humanities Junior Scholars Workshop
University of Southern California — June, 2011

"Stance-Takings as Transfers:
The Pragmatics of Apologizing and Promising"
Semantics Seminar, Philosophy and Linguistics
University of California, Los Angeles — April, 2011

"Beyond Intention: Sincerity in Promising
and Stance-Taking"
Ethics Writing Workshop, Philosophy Department
University of California, Los Angeles — February, 2010

"Thank you and I'm Sorry"
Philosophy of Language Workshop
University of California, Los Angeles — February, 2009

"Phenomenal Concepts and the Concept of Phenomenality"
Society for Philosophy and Psychology Annual Meeting
Wake Forest University — June, 2005

INTRODUCTION:

We seem to do extraordinary things just by saying the right words at the right times, and meaning them. We apologize – and somehow redress a wrong done, at least partially. We forgive someone – and a relationship is restored, and an offense absolved. We consent – and an otherwise forbidden invasion becomes morally permissible.

There is little disagreement, among ethicists and speech act theorists, about the power of these communicative acts. Yet there remains dispute, or else sheer puzzlement, over just how they can accomplish what they do, even in new and unprecedented contexts – in legal or inter-group conflicts, for example. Just what is it about saying "I'm sorry" that warrants forgiveness or at least foregoing a demand for full restitution, even if the victim already knows everything about how the offender feels? And what is it about forgiveness that could relieve the wrongdoer of some duties of repair, and make it appropriate for him to protest if the victim changes her mind and punishes him after all? Finally, how can countries and corporations, who – as institutions – feel no resentment or regret, manage to apologize and reconcile with sincerity to the point that group-level forgiveness may be warranted?

The dissertation seeks to understand and account for these transformative powers. It is motivated by a conviction that the key to the moral impact of these speech acts lies in their nature and function as actions. To capture how they accomplish so much, then, we need to understand what exactly they are.

To that end, chapters I and II explore the paradigmatic case of apology. Traditional accounts of apologies and similar speech acts treat them as either asserting some fact – such as

that the speaker appreciates that he was wrong and will not do it again – or expressing some psychological state. As such, these accounts cannot make sense of the way an apology responds remedially to a past wrongdoing, in part because they leave unclear how an offender can counteract what he did merely by asserting or expressing certain things about himself. This problem is illustrated most dramatically by cases in which the victim already knows everything relevant about what the offender feels and believes, and still justifiably demands an apology. In the case of blameless harm, the focus of Chapter II, the offender presumably has no beliefs or feelings to improve upon – as he did or felt nothing wrong in the first place. So if he makes a moral difference by saying he is sorry, which I argue he does, there must be some other function performed by the speech act besides transferring information.

Instead, I argue that apologies, and similar speech acts, should be understood less as expressions than as ways of treating people – and committing to do so – that counteract a certain mistreatment put in place by earlier actions. This model requires a new, relational understanding of both apologetic expression and of the actions that give rise to them, such that one can meaningfully counteract the other. If an apology is to change something about what the apologizer did beforehand, it must be able to counteract some effect of his earlier action. Indeed, I argue, that is what they do: wrongdoings have an adverse effect beyond the material harm, if any, that they inflict; they also put in place a way of treating the victim that has objectionable meaning. An apology counteracts this mistreatment with an equally meaningful way of treating the victim, as I explain. In apologizing, one performs the speech act I will call "stance-taking," a kind of speech act that both constitutes and commits to a way of treating someone in light of a normative position one has adopted. I introduce stance-taking in Chapter I and elaborate upon it throughout the dissertation.

The account shows that the reasons to apologize relate to the way an offender otherwise treats his victims. That picture, however, is challenged powerfully by an important feature of another stance-taking: forgiveness. Through the formal act of forgiveness, a victim can relieve her offender of the duty to apologize. Chapter III explores this feature of forgiveness and seeks to reconcile it with the account of apologies and moral repair developed so far. I first try to explain why a victim can, by forgiving her offender, alter his moral status – rendering apologies and other acts of moral repair unnecessary, and their absence no longer blameworthy. I argue that her forgiveness has this power because, and to the extent that, it helps fulfill the function of other remedial steps like apology and restitution, though in a different form. As with apologies, then, forgiveness emerges as less a unilateral expression than an interactive approach to another person, which can help restore their relationship and improve the moral landscape between them.

These speech acts, then, work less as ways of revealing the speaker's psychology than as ways for them to relate to others. That, however, does not mean they involve none of the familiar psychological states typically associated with them, such as sadness, grief and affectations of regret. When people perform these speech acts, they will likely undergo these other psychological episodes, if only because that is part of how they experience the forms of treatment and commitments required by sincere stance taking.

Still, it will prove important that such psychological episodes are not essential to acts like apology and forgiveness, in the way that commitments and dispositions to act might be. That is why even institutions – countries, courts, companies – can sincerely apologize, forgive and engage each other in similar speech acts, to dramatic effect. In Chapter IV, I try to account for how institutions can sincerely take the stances involved in moral repair, apologizing and

forgiving meaningfully, even as they lack any of the psychological states typically associated with these acts.

The resulting account of apologies, forgiveness and other speech acts, whether by individuals or institutions, supports a break from the speech-action dichotomy that Austin and Searle began to undermine half a century ago. On the account proposed here, speech – in the case of stance-taking – functions as action over and above what it communicates, and actions have meaning beyond their material impact on the world. The area of moral repair, then, uniquely illustrates the moral significance of what action can mean and speech can do.

CHAPTER I:

APOLOGIES AS STANCE-TAKINGS

Apologies wield enormous power. They can warrant forgiveness, even relieving wrongdoers of further debts of repair. They can restore relationships, sometimes even improving them. Failing to apologize, moreover, grounds continued resentment and worse, no matter what else is done to make things right. And yet, saying "I'm sorry" or "I apologize" hardly seems to repair the harm done by a wrongful act, or to compensate for it. At most, it seems to convey an attitude or acknowledgment, and it is not clear exactly what difference such conveyance alone makes, especially when the victim already knows the information relayed. How and why, then, does it matter so much whether a wrongdoer apologizes?

Philosophers have sought to explain the remedial power of apologies in terms of what they reveal to their audience – that an offender commits never to do it again, for example,[1] or that he recognizes his wrongdoing,[2] or even that he wants to repent.[3] For this family of views, an apology works by dissuading the victim of reasons that the initial wrongful act had given her to resent, fear or distrust the offender or his behavior.[4] Others argue that the communicative aspect of apologies is unnecessary, and that in fact they work by opposing the wrongful action with a

[1] See, for example, Martin 2010, arguing that an apology works by reaffirming one's commitment to the victim to treat her differently.
[2] See, for example, Griswold 2007, 136-37.
[3] David Velleman discusses a version of this view in Velleman 2003, 241.
[4] Some have also insisted that apologies aim not only to remove the victim's *reasons* for resentment but to actually assuage the victim and cause her resentment to recede, even if her resentment is unjustified. See, for the seminal case, Maimonides 1987.

single counteractive sort of action, such as subordinating oneself where the initial act presumed superiority, or honoring the victim where the initial act disrespected her.[5]

In contrast, I argue here that apologies work by doing more (and sometimes less) than conveying information. Nonetheless, their communicative element is essential and central to their remedial power. In particular, I argue that they work by putting in place a new way of treating the victim, which itself reverses a prior way of treating the victim that began with the initial wrongdoing.

My account depends on characterizing apologies as a unique kind of speech act, of a piece with thanking, absolving, taking responsibility and others, which I call "stance-taking." To take a stance is, in the typical case, to perform a speech act that both acts on and commits to act on a normative claim one adopts. For example, to endorse a candidate or a cause ("I'm for A!") is both a way of acting on a political principle (A is to be supported), whereby the speech is itself a form of support, and to verbally commit oneself to so acting. When an offender apologizes to his victim, I will argue, he treats the victim as someone he owes not to have wronged as he did, and verbally commits himself to such treatment. In this way, the apologizer begins a relationship as the victim's moral debtor, so to speak. This way of treating the victim, moreover, reverses a prior way of treating her that began with the wrongful behavior: namely, treating her as one who could be wronged that way. I will try to show how wronging someone involves this kind of mistreatment, and how apologies can be ways of reversing and redressing it.

In the first four parts of the chapter, I will consider and respond to alternative accounts of how apologies work, some of them familiar, others just recently proposed. I will start, in Part I,

[5] For an account along those lines, see Bovens 1993.

with the question of why mere assertions, on the part of an offender to her victim, cannot do the work of apologizing, or at least not most of it. I will then argue that expressing psychological states cannot do this work either, notwithstanding the standard accounts of apologies as expressions of regret or remorse.[6]

So, if not by asserting propositions of some content ("I was wrong," for example, or "I am sorry"), or expressing emotions like sorrow or regret, how do apologies make moral differences? I will explore two possible answers to the question and show why I find them incomplete. I will then introduce and argue for an alternative account of how an apology could in fact help act remedially in response to wrongdoing. That will be the account based on stance-taking, previewed above. It will be laid out in greater detail, below, and defended against likely objections.

I. 'I'm sorry' isn't enough: apologies as assertives.

My principal objective is to investigate how apologies act remedially in response to wrongdoing. This includes the question of precisely what aspect, if any, of a wrongdoing do they affect, and how do they affect it remedially. Whatever the answer turns out to be, the questions arise in part because it is undisputed that apologies have remedial power of *some* sort. In particular, apologies seem to have the following features in need of explanation:

1. They are owed, or due, to the victim of one's wrongful behavior, as something presented directly to her.

[6] Searle 1979, 12.

2. They have staying power. Once the apology has been made, it is "in place," so to speak; the wrongful act thereby apologized-for. It is henceforth less sensible to ask the offender, "Do you still apologize?" than it is to ask, about an expressed opinion, "Do you still believe that?"
3. They improve the wrongdoer's moral standing with regard to what she's done. Once an offender has sincerely apologized, she has given the victim a reason to resent her less, and at times to absolve her of the wrong done, even relieving her of duties of further repair.

My chief concern is to answer the question of how speech acts like apologies can do all that. To start, though, it may be useful to consider what *can't* achieve such effects, and determine what an apology adds to these insufficient counterparts. As a non-controversial starting point, I want to propose the following thesis: one cannot redress a wrong, in the way apologies can, simply by believing something or feeling something, like regret, guilt or remorse. Even if Jack privately feels guilty about wrongfully harming Jill, or any other state an apology might be thought to express, he nevertheless has not yet done the remedial work apologies do. This assumption, I expect, is easily accepted. To make the moral difference apologies make, it might be agreed, one has to do more than think or feel something.

To this I want to propose adding one, only slightly more ambitious, premise as a further starting point: an apology's work is not done even once the victim *learns* how her offender feels or thinks about what he did. While it may cause Jill some comfort to discover Jack's remorseful state of mind, it is perfectly apt for her to complain: "Glad he feels that way, as far as it goes, but he still needs to step up and apologize."

This point – that the function of apologies cannot be completely fulfilled by the victim's receiving information about her offender's state of mind – may be thought sufficient to establish that apologies cannot be effectively performed by assertion alone. But that would be too quick. A fundamental feature of assertion has not yet been raised: in making an assertion, the speaker *intends to communicate certain content* to the listener, and to be understood by the listener as so intending.[7] When Jack tells Jill how badly he feels about what he did, he does more than transfer a pre-existing fact; his very telling reveals a new fact – his intention to tell her – which is itself relevant to the moral evaluation of his status as a repentant wrongdoer. In apologizing, then, one could help make up for a past wrong *both* by having the requisite attitudes, feelings or beliefs about one's wrongdoing and, in addition, intentionally communicating them to one's victim in a way that also makes this intention clear to her.

i. The interview apology

These two steps, however, can be shown still insufficient to achieve the moral remedial effects of apologies. For it is possible to assert, while clearly intending to assert and to be understood as such, all those facts about oneself in an obviously unapologetic context. For example, one can assert them in response to a series of questions. Consider the following scenario:

Jack betrays Jill by revealing a scandalous secret she confided in him. He seems to avoid her for days afterwards, but Jill confronts him, and questions him as follows:

Jill: Did you tell?

[7] Grice, 1957, 383: "For A to mean something by X…A must intend to induce by X a belief in an audience, and he must also intend his utterance to be recognized as so intended."

Jack: Yes, it was me.

Jill: Do you think that was right?

Jack: No.

Jill: Did you mean to do it?

Jack: Yes.

Jill: Wasn't that wrong?

Jack: Yes.

Jill: Do you regret it?

Jack: Yes.

Jill: Do you regret it because it was wrong?

Jack: Yeah, of course I do.

Jill: Will you do it again?

Jack: No, obviously.

Jack's responses assert all the propositions thought to be essential to what a paradigmatic, or ideal, apology conveys.[8] He acknowledges his wrongdoing, admits that he is culpable for it, states that he feels badly about its wrongness, and that he does not intend to do it again. Moreover, he spoke with the intention of asserting these propositions to Jill, and with her taking

[8] Smith 2008, 28. Smith does not offer this analysis of the paradigmatic apology as an explanation of how apologies work or achieve moral repair; he simply aims to characterize the essential content that proper or complete apologies convey. My critique of the Interview Apology, then, is compatible with his account of apologies.

him to be doing so. Yet it seems equally intuitive that such answers, to a victim's questions, do not do the moral work of apologizing.

One may want to challenge this intuitive response by proposing the following exchange: 'Do you apologize?' 'Yes.' If, however, that counts as a sincere apology, it prompts the question of what the phrase 'I apologize' adds to the assertion of some truth about an offender. If much of the moral work of apologies can be done by asserting content, we ought to be able to translate that content into statements of fact about the offender: "I did it"; "I'm guilty"; "I feel such and such." If, however, the phrase "I apologize" is irreducibly distinct from any such assertions of fact about the offender, then, by hypothesis, uttering the phrase does more than intentionally assert something to the victim. The question remains: what else does 'I apologize' do, which can account for its unique effects?

II. Expressives

It has been argued so far that apologies cannot do their moral remedial work merely through the assertion of facts about the offender. There is, however, an alternative way that apologies have been thought to undermine or work against one's wrongdoings – one that requires communication, but is not reducible to assertion. In fact, the classical accounts of apologies as speech acts treat them not as mere assertions of propositions but as expressions of psychological states. On John Searle's view, for example, apologies are expressions of regret.[9] To apologize, then, an agent expresses her regret to the victim of her wrongdoing.

An apology, then, is a kind of speech act different from mere assertion, namely an *expressive*. An expressive is a speech act that expresses a state of the speaker and, if sincerely or

[9] Searle 1979, 12

felicitously uttered, expresses one that the speaker correctly believes actually obtains at the time of utterance. The phrase "ouch" is an expression of pain, for example, and is appropriately uttered when the speaker correctly believes he is in pain.[10] Notice that expressives, on this view, do not merely report that the speaker is in the state expressed (as in "I am really impressed by your performance."). They give voice to the state itself ("Bravo!").

By way of illustration, one can compare a psychological state to a room lit up inside a house. An expressive, then, is akin to opening the shades in that room – the light shines out, reaching the observer on the lawn. A mere assertion, on the other hand, more closely approximates passing a note under the door informing the outsider which room is lit. It lets the observer know about the light, but doesn't directly expose him to it. Notice that expressives, on this understanding, involve two components: first, they vocally display a state of the speaker (rather than report it), and second, they are intentionally communicated by the speaker to the listener for that purpose. To Searle, apologies are expressives that, when performed appropriately, give voice to the offender's regret as it obtains.

Suppose then, with Searle, that regret captures the psychological state an apology is meant to express, where by "regret" is meant *moral* regret: a negative attitude towards what one did because of its having been wrong, rather than merely because of adverse consequences of having done it.[11] If that is so, then the expressive element of apologies may supply what's missing in the Interview case, and account for the key remedial role played by the speech act. This view is not implausible. Consider comforting a mourner. It is presumably more effective to

[10] David Kaplan, "The Meaning of Ouch and Oops," Paper delivered as UC Berkeley Graduate Council Lecture, accessed via web at http://www.youtube.com/watch?v=iaGRLlgPl6w (April 24, 2004).

[11] I use "moral regret" instead of "guilt" or "remorse," because guilt is sometimes understood as a judgment (i.e. that one did wrong), rather than a psychological state or attitude, while remorse may be inappropriate in cases where no wrongful loss or harm was imposed. Thanks to Herbert Morris for clarifying this distinction.

cry with him than simply to assert that you feel bad about what happened. And it would be unduly reductive to chalk this difference up to the epistemic advantages of crying – i.e. it's better evidence of true feeling. There is, rather, something intimate and compromising in baring one's emotions in front of another. It is a way of making oneself vulnerable, exposed, before someone else. And maybe this emotional sharing, alone, accounts for the additional remedial effect of apologies, beyond that of asserting some truth or other.

Without discounting the value of such expressions, I dispute that they can account for the remedial effect of the apologetic speech act. Notice that if expressives do much more than assertives, that advantage should lie, at least partly, in what they express. They share or reveal something, beyond the communicator's mere avowal of some fact or other. But it is not clear whether regret, moral or otherwise, is the kind of state that can be shared to such effect.

Two possibilities seem available: on one, such regret is *affective*, involving an arresting emotional state that is only sincerely expressed when the speaker is actively feeling or undergoing it. On the other, it is an intentional state that need not actively arrest the regretful – no "pangs" of regret, in other words[12] – but is equated with an agent's taking the position that she did something wrong and wishes she did otherwise.

On the first possibility, apologies – if they worked by expressing regret – would involve the display of a vivid, felt state just as it overtook the apologizer. This, however, would run up against a familiar feature of apologies: they should be timely. As time passes after the wrongdoing, an apology becomes wrongfully late.[13] Yet the instant proposal – that apologies work to the extent that they express an active affective state – would render the timeliness

[12] The phrase comes from Gilbert 2001.
[13] That granted, apologies may be otherwise defective if uttered too early, before one appreciates the reasons to do so. See Shiffrin 2002.

requirement difficult if not impossible to meet. Imagine, for example, that I have wronged someone, and I realize it was wrong, and appreciate that I have a duty to apologize to him. So I contrive to run into him as he leaves his office and apologize on the spot. But when I spot him as I lurk, poised to apologize, still ever appreciating how wrong I was, I realize something is missing: I don't exactly feel the kind of affective moral regret that can be vividly displayed. Or at least I'm not sure I'm feeling it, actively, at that moment. True, I am utterly persuaded that I owe the apology, shouldn't have wronged him, and perhaps even that I *should* feel an affective state of regret. But, sensitive introspecter that I am, I know I'm not experiencing it at that moment. On the proposed view – that apologies paradigmatically express a vivid, affective state – I should not apologize in the state just described, for that would rob the apology of its essential function. Indeed, it may also be insincere, like saying "Bravo" when I feel no positive reaction to anything. Either way, I couldn't be required to apologize unless I'm required, as well, to be in the throes of the right feeling just at the moment, those few seconds, when my victim passes me. This seems too much to ask for the fulfillment of such a commonplace duty, particularly one that is owed in a timely fashion and at a moment when in dialogue with the victim.

But rejecting this possibility – that apologies centrally express an active, affective state – seems to leave only the alternative account of what it is to express moral regret, to wit: expressing one's affirmation of the essential positions implied by an apology. For example, an apology might express one's position that one did something wrong and wishes one did otherwise, and perhaps also that one intends not to do it again. The problem with this possibility is that it robs apologies – as expressive speech acts – of any power over and above that of assertion. That is because expressing one's position about something – that one did wrong and

wishes one did otherwise, say – is akin to expressing a belief. And assertions already do that. Yet assertion, as we saw, cannot do the moral work of apologizing. That leaves the question of what, beyond asserting content or – which amounts to the same – expressing beliefs, apologies actually do that constitutes a remedial response to past wrongs.

III. Apologies as agent-initiated action

So far, I have argued that it is insufficient, for doing the moral work of apologizing, to simply assert facts about oneself, such as that one feels badly about what one did or intends to change. The Interview Apology fails. It has now been argued, as well, that expressing psychological states is not the essential remedial feature of apologies missing from the Interview Apology.

Why, then, does the Interview Apology fail? It seems that part of where the interviewee falls short, even as he affirms his guilt, regret and responsibility, is his passivity. The offender, in such a case, may have affirmed or expressed important information about what he did. But he took no action or initiative towards apologizing to the victim. He did not initiate the communications that were extracted from him ("Do you feel sorry?").

What, then, *about* initiating an apology lessens or counteracts the impact of a past wrong? Initiating the apology might be thought to reveal something of moral significance about the offender, such as the strength of his regret or his sincerity, as expressed by such judgments as "If he really felt bad, he'd go the extra mile and apologize." But that option is not available, because if it were, apologies would work as assertives, sharing with the victim whatever psychological state is reflected in taking such action. Yet it has been argued so far that whatever work

apologies do, it must extend beyond the accurate report of some emotional state. So if initiating the apology makes a difference, it cannot be by way of merely *revealing* that the offender feels or believes something. It also cannot be by way of expressing some psychological state, for the reasons already presented above. That is, unless it is a psychological state whose activity is mainly constituted by a tendency to take initiative action, which would render it quite different from the sort of emotional state whose expression would add to an offender's vulnerability. In either case, we are left with the question of what about initiating action is so important to the remedial work of apologies?

The instant proposal, again, is that agent-initiated action may be necessary to do the moral work of apologizing. Moreover, the work done by such action is not, or not principally, that of revealing to the victim some fact about the offender or making oneself vulnerable by baring some state to the victim. What, then, is accomplished by this irreducibly active component of apologizing? Two possibilities bear closer consideration.

A. Pledges to reform

Suppose that in apologizing, the offender makes a pledge, a commitment: she resolves not to repeat the act. Indeed, this proposal is not implausible: an apology is, after all, considered insincere if the victim intends to repeat the violation. So in apologizing, the offender takes the performative step of pledging to change and not to repeat the offense. This feature, perhaps, accounts for why a victim needs more than mere factual information about the offender. For merely hearing that the offender is currently repentant is not enough, perhaps, to truly assure her

that the threat is gone.[14] But a pledge on the offender's part is different; once the wrongdoer commits to reforming his ways, foreswearing future wrongs, the victim has that commitment as an additional reason to feel less threatened by him and his future behavior. True, he is morally committed to such reform anyway, inasmuch as he is obligated not to repeat. But it was his apparent willingness to shirk this duty that motivated the victim's initial perception of a threat; the pledge both counteracts this impression and adds a further, verbal commitment to the offender's moral commitment not to repeat. If nothing else, this is useful evidence in support of fearing the offender a bit less.

Unlike previous proposals, the pledge model reflects the irreducibly performative feature of apologies, the sense that, even once the victim knows everything she could possibly learn about the offender, she can still justifiably find the *act* of apology absent. On this account, the missing element is the performance of the pledge – the self-obligation to avoid repeating the offense.

This particular proposal fails, however, because it cannot account for cases of sincere and effective apologies by offenders who will definitely never encounter the victim again. Consider the evocative example of the deathbed apology: one apologizes to a friend, whose death is imminent, for having waited too long to do what she asked him to do. Such apologies say nothing about the offender's future behavior, nor does it seem to matter if they do. Their future relationship is beside the point: she is apologizing for the *past*, so as to make things right *for now*. A similar point follows from cases in which the offender is at death's door: he has no future behavior to pledge, yet he seems capable of sincere and meaningful apology that justifies

[14] Pamela Hieronymi characterizes wrongful actions as posing threats by revealing (and in effect expressing) the agent's evaluation of the victim as someone he can wrong – an evaluation exposed by what he did. Hieronymi 2001, 546.

forgiveness and reconciliation. Or consider apologies between strangers, momentarily passing through each other's lives. If one wrongs the other during their once-in-a-lifetime encounter, it seems an apology is due – and can make a difference. But by hypothesis, it will have no impact on their future interactions. As a result, the difference apologizing makes could not come down to its reassurance about the future.

B. Gestures of subordination

Luc Bovens proposes a completely different explanation of how apologies effect moral repair through action. Rather than focus on the commitment, if any, that the apology makes, his account emphasizes the status relationship it restores.[15] On Bovens's view, a wrongdoing is a way of disrespecting the victim, specifically by treating her as less than a moral equal, entitled to the same rights and restraint as oneself. The failure to regard the victim as a moral equal amounts to the offender placing herself above the victim, looking down on her as inferior. As a result, there is now a "respect deficit" between them.

How can the offender restore the equilibrium in respect that ought to obtain between her and her victim? By reversing the respect dynamic, argues Bovens. Specifically, "The offender pays excess respect to the victim to restore this deficit and transfers power to the victim as a form of respect."[16] On this picture, the apology is a way of subordinating and humbling oneself before one's victim. And in this way, one cancels out, as it were, the presumption of superiority with which one had treated the victim just before.

[15] Bovens 2008, 220-239.
[16] Bovens 2008, 233.

This account, in my view, helpfully shifts the focus of the apology's remedial power from the information conveyed to the way the offender treats the victim in apologizing, a feature it shares with the account I will propose. Furthermore, it should be stressed that Bovens's notion of disrespect is not literal. On his view, wronging someone simply *is* a form of disrespect, even where the wrong does not, in itself, put someone down. For example, suppose a teacher asks her most admiring student to review a manuscript, demanding that he commit to "be critical, forget who I am – I'm depending on your commitment to do that." Suppose further that the student does commit to review the manuscript undeferentially, just as she would like, but he succumbs to his deep admiration for the teacher and, when confronted with what looks like a glaring error, he assumes the teacher must have meant it deliberately and that he misread it. In other words, the student wrongfully violates his expressed commitment to the teacher – to be tough, ditching deference – on which the teacher said she depended. Yet the violation reflects no disrespecting attitude on the part of the offender – (seemingly) quite the opposite. Nevertheless, on Bovens's view, it *is* in fact a form of disrespect more broadly understood: the student takes an undue liberty with his interaction with his teacher, failing to do what he committed to her to do. The wrong is *itself a form of disrespect* – a liberty inappropriately taken with the rights of another.

The drawback of Bovens's account, however, is that his own proposed form of redress takes respect more literally than his account of wronging did, and implausibly so. He argues that an apology's remedial power lies in its presenting the offender to the victim as subordinate, humbled. The apologizer takes a humiliated or at least inferior posture before the victim. But if wrongdoing is not literally a form of subordination, and so had nothing to do with acting as though one is better than one's victim, how would acting subordinately redress the wrong? The question is one of mechanism: what *about* bowing one's head, say, or shaming oneself, could

help mitigate a past transgression – like failing to fulfill one's commitment to one's teacher – or some consequence of it? If we grant Bovens that wronging someone manifests a kind of disrespect, it cannot be a conscious form of taking the other to be inferior or unworthy of respect – by hypothesis in the case above, the wrongdoer, if anything, *excessively* regards the victim as superior. If he then treats her like a superior, or presents himself as an ashamed inferior, he is merely continuing the sort of treatment he already carried out, and one he embraced too whole-heartedly at that.

In short, the problem with Bovens's account of how apologies work is that he proposes a *literal* form of extra respect as a remedy for an entirely *non-literal* form of disrespect, one so abstract as to be practically synonymous with wrongdoing; thus his remedy misses the mistreatment it targets. There is no reason that putting someone above oneself literally, such as rendering her superior in some way, has any counteractive effect on, say, failing to warn her about a risk she stood to incur. This is particularly clear in cases where the offender is already the victim's subordinate in some straightforward way, such as being her employee.

Importantly, I am not disputing Bovens's suggestion that wronging someone involves some disrespect, theoretically abstractable from the rest of the wrongdoing (such as the harm it may inflict). Nor am I disputing his characterization of what apologies convey or present about the offender. My worry is, rather, that the wrongdoing may have nothing to do with actual, literal subordination of the victim. As a result, Bovens owes an account of how the offender's own literal, actual self-subordination could help remedy it. In the absence of such an account, it remains premature to accept Bovens's interpretation of the remedial work apologies do.

IV. Taking a moral stance: how apologies succeed

A. The need for a different account

It has been argued so far that assertive accounts of apologies fail to explain their remedial power, because it is not clear what good it does to inform the victim of the facts apologies purportedly convey. It has been further argued that apologies as expressives add too little to their power as assertives – in which case expressing psychological states cannot do the missing remedial work – or too much to be required of wrongdoers as soon as apologies are due. Yet neither pledging to reform nor subordinating oneself to the victim seems to do the essential remedial work of apologies, either.

The goal that remains, then, is to explain how the mere utterance of an apology could work against or respond remedially to a prior wrong, if not by the means already discussed. Of course, in aiming for such an explanation, it bears mention that one should not try to portray apologies as correcting all or even most of the damage wrought by a wrongdoing. If apologies perform moral repair, this may be merely by mitigating some of the harm done, or some further morally undesirable outcome that accompanies the primary transgression. But what about the wrong could apologies change? They clearly do not mitigate the physical or property damage of many wrongs. Words won't repair a negligently broken arm, or replace the book I borrowed and lost, or – if uttered one-on-one – restore a person's good name. What remedial work, then, do they perform?

The answer turns on the fact that wrongdoing involves a certain mistreatment of the victim, over and above what else it inflicts upon her, namely: treating the victim as one whom it is acceptable to violate in that way. In the case of intentional wrongs, this mistreatment is constituted *both* by first wronging someone intentionally, *and* by subsequently leaving it in place, so to speak, without attempts at redress. Intentionally committing a wrongful action treats the victim as one whom it is acceptable to violate that way, or, put differently, as one whom the offender is free to violate that way. Moving on, or continuing to act as before, after wronging someone – taking no deliberately remedial steps, for example – *also* treats the victim as one whom it is acceptable to wrong as one did. Indeed, the mistreatment begins with the initial wronging and continues for as long as the offender allows it to stand without redress. As long as the wrong is left unredressed, then, the offender is treating the victim as one whom it is acceptable to violate as he did.

In the case of unintentional wrongs, by contrast, the mistreatment that apologies target begins only *after* the wrong is committed. Failing to provide for one's employees or one's children, for example, even if the failure is inadvertent and non-negligent, may violate some duty to them.[17] And having violated a duty – and leaving it to stand without redress – the violator now begins to treat the victim as someone violable, as one to whom it is acceptable to do what one did. So with intentional wrongs, the mistreatment begins with the wrongful act, while with unintentional wrongs, it begins afterwards, with the failure to take any steps at redress. Importantly, though, committing *both* intentional and unintentional wrongdoings and leaving it in place, without redress, constitutes treating the victim as one whom it is acceptable to violate.

[17] See, for example, Waldron 1993, 203-24.

How, then, do we end that treatment? One way, it might be thought, is to simply reassure the victim that she's not violable, in whatever way our actions – and subsequent inactions – seemed to treat her as violable. Jean Hampton and Jeffrie Murphy argue that an apology serves as a kind of statement of position, as though the offender is an author editing her work. Her initial action put her view one way, which meant something offensive to the victim, and now she is correcting the prior misstatement.[18] Hampton and Murphy do not necessarily mean here that the wrong is a form of communication, whereby the offender intentionally expresses his point of view about the victim and her violability. Rather, as Pamela Hieronymi puts it, "an action carries meaning by revealing the evaluations of its author."[19] The action reveals his view, even if it does not intentionally express it. So if wronging someone and acting as though nothing happened treats her as violable, moral repair may consist in letting her know she is not violable, after all.

The problem with such a communicative form of repair, however, is – as in Bovens's proposal – that it is misdirected. Note that the target of moral repair is the *treatment* of the victim as violable, not the expression of the view that she is violable. This distinction – between treating someone a certain way and expressing one's opinion about her – is crucial to my account, and how it differs from the message-revising accounts of Hampton and Murphy. I propose, again, that wronging someone and continuing to act as though nothing has happened *treats* or takes the victim as violable in a certain way, particularly if it stands unredressed. It does not *represent* the victim as violable, or express the offender's *attitude* to that effect. Rather, I'm claiming, *the action, and the subsequent inaction, treats her this way.*

[18] Murphy and Hampton 1988, 44 (quote attributed to Murphy).

[19] Hieronymi 2001, 546.

This difference can be illustrated by a dispute between an earlybird and a latecomer. Earlybird arrives punctually; Latecomer shows up half an hour after that, several times, each with an unimpressive excuse, such as, most recently, that he got caught up in a political argument and lost track of time. Earlybird complains: "You act as though my time is worth less than yours." To that Latecomer, with visible sincerity, responds: "You have no idea. I do not think less of your time. In fact, I think your time is much more important than mine!" Latecomer has, however, missed the point, as Earlybird responds: "I never said you actually *agree that* my time is less valuable. I meant you're *treating* me as though it is." Latecomer's behavior, in other words, is a way of acting as though Earlybird's time is less valuable. The behavior itself *acts on* the insignificance of Earlybird's time, even if it doesn't reveal Latecomer's conscious view to that effect.[20] Again, it's the difference between action that *treats* someone as though P, and action that *reveals that one takes the view* or position that P.

If, as I'm arguing, wronging someone without redress constitutes a mistreatment that stands apart from the offender's apparent or actual attitude about the victim, then nothing like an expressed "change of heart," or repudiation, could remedy it. Since the mistreatment is not necessarily an expression of the offender's attitude – apparent or actual – a professed change of attitude need not have any effect on it. What is needed, instead, is a different, less purely expressive account of how one can remedy the mistreatment involved in both wronging someone and failing to do anything to redress it.

[20] This type of behavior – that treats someone in an objectively insulting way, regardless of what subjective view it expresses – constitutes what Elizabeth Anderson and Richard Pildes call "expressive harm." See Anderson and Pildes 2000, esp. 1503, 1527.

B. The fundamental asymmetry of moral repair

I have argued, again, that wronging someone and not attempting to redress it treats the victim as though one is free to violate her in that way. And this treatment (begun by the initial wrongdoing), or mistreatment, is in need of redress. That raises the question of what might redress it. One seemingly obvious answer may lie in a familiar form of moral remedy: repair the concrete harm done. For example, if property was wrongfully damaged, compensation presents itself as a natural answer: repair the *moral* damage, if there is any, by repairing all the other damage. The problem for this proposal is that wronging someone and compensating her (as any comparable physical repair) does not, in itself, amount to treating someone as though she is not violable. It could, alternatively, be seen as treating her as though she can be wronged "for a price," so to speak; wronged and then compensated.

We see this kind of behavior most often in the way the powerful treat their subjects. Kings and governments routinely seize property or inflict harms and then, in a spell of benevolence or under the threat of legal action, pay their victims back. Yet even if they cover the physical cost, this behavior does not treat the subjects as people whom their rulers are un-free to violate or relieve of property. Instead, it treats them as people from whom one can steal and then pay at will – which is different.

Similarly, suppose I took your coffeemaker and recklessly broke it. As argued above, if I simply go on acting as though nothing has happened between us, then I am treating you as someone whose possessions I am free to take and break. Now suppose that, instead, I show up at your door with an envelope full of money and simply hand it to you without a word (save the polite greeting). This behavior would certainly be compatible with my feeling bad about what I

did. But it would also be compatible – equally compatible – with treating your possessions as rental items, or objects I can take, break and then pay up. In other words, harming and paying treats you as *both* someone I'm unfree to wrong *and*, alternatively, as someone I'm free to wrong for a price. Importantly, to remind the reader, I am not claiming that the psychology compatible with behavior is what gives it its objective expressive content. I am simply using the psychological attitude reflected by certain behaviors as a heuristic for *uncovering* its objective meaning. And using that method here, we find that wronging someone and subsequently compensating her is ambiguous: it could treat the victim as one who can acceptably be harmed and compensated. That would amount to what could be called an *objective insult* – a way of being treated to which someone has reason to object, or to take offense.

Compensation, then, would not rise to the level of reversing the mistreatment that follows wronging someone and leaving it unredressed. But the problem this presents for moral remedy is that nearly anything one does after wrongdoing can be seen as compensation, or at least as a price one is paying for the misdeed, like a fine. And as we have seen, wronging and then paying, so to speak, does not treat someone as though she was inviolable in the first place. But what else is there to do after a wrongdoing, if any subsequent action treats the victim as though it is the price of the earlier transgression? How, in other words, can one take remedial action while treating the victim as though such action isn't enough?

To put the problem more relevantly: what form of moral remedy can be offered without taking itself to be complete or adequate (as in "Here, I've paid X and said Y, so we're square")? Anything he might do, by way of compensation or even self-punishment, could be a form of paying a price or fine for the wrongdoing – which amounts to treating the victim as wrongable for a price, an objective insult. How, then, can the offender treat her otherwise?

One answer is already familiar from the more directly monetary cases of owing someone more than one can repay, as when one damages a priceless heirloom and tries to get its owner to accept a sum of cash. That way involves forgiveness: the one who owes the debt asks his debtor to release him of it, to forgive the debt. This way he takes his offered sum to be insufficient, acting as though only the owner can – through his generous act of forgiveness – make it suffice.[21] The money is, in other words, offered as an insufficient amount that the recipient is nevertheless in a position to accept or reject.

This is the model I am proposing for moral remedy more generally: to avoid the mistreatment described above, it must be offered for the victim as something to accept or reject, rather than something that is objectively sufficient. This offering treats the victim as one who is owed *more* than any compensatory act can give her; it treats her as one who is owed not to have been wronged in the first place.

Notice that this dynamic is reflected in the analogous case of prior consent. If I am going to assault someone, no amount of bribery or lavishing of benefits, or self-flagellation, can "buy" my right to the violent act. Indeed, if I attempt any such act and then assault my victim anyway, I will have treated him as violable for a price or token gift. On the other hand, if I seek his *permission* to engage in the (minor) assault, then my behavior no longer expresses his violability. Seeking forgiveness is, on this analogy, similar to a post-hoc form of consent, as if to ask: "will you accept what I am doing as sufficient to make up for my prior trespass, even though it isn't?"

But it is important to note that the treatment I am describing – as someone's debtor – does not need to rise to the level of seeking forgiveness, although that certainly suffices. One can, instead, give something less than what is owed in such a manner that it is not taken to be sufficient; it is offered as insufficient. That is the model I am suggesting for apology.

[21] This feature of forgiveness will figure prominently in the arguments of Chapter III, infra.

When a wrongdoer apologizes, then, she offers both the apologetic gesture itself, along with anything else she does by way of moral repair, as an attempt to make up for a wrong *but which is insufficient*. It amounts to presenting oneself as in no position to make up for the wrong done, but offering something in the hope that the victim might accept less than what she is entitled to. On this model, then, apology is not really a unilateral expression, revealing one's opinion or position. It is, rather, an interpersonal performative act, offering something to the victim to be accepted or rejected. And it need not take the form of saying one is sorry. One can just as easily present oneself as having wronged the victim, and having no way to make up for it, but seeking the victim's acceptance of what is done anyway, or seeking her forgiveness of the outstanding moral debt. That explains why the expression, "Please forgive me for the horrible thing I've done, though I don't deserve it" can be a workable substitute for apology. In this way, one treats the wrongful act as unacceptable, and as not fungible – as not doable for a price.

Importantly, though, it is not enough simply to say all these facts: that one cannot make up for what one did, for example, or that one believes it is up to the victim to accept any moral repair. That is because, as already discussed, the wrongful act left unredressed is problematic because of how it *treats* the victim, not because of what attitude or position it reveals. As a result, it is not enough merely to say that one does not believe the victim is wrongable, so to speak, or to express that belief by asserting these facts. Rather, one must treat the victim in the remedial way, relating to her as one who cannot be wronged, and who is owed more than compensation, and who alone has the power to relieve that excess debt. Apologies, then, cannot redress the prior mistreatment – put in place whenever we wrong someone without redress – by expressing beliefs alone. They would, rather, have to put in place a new kind of *treatment*, a new

way of relating to the victim as one who is owed better and who alone can forgive or relieve that debt. And they would have to *be* part of that treatment, as well.

C. Moral stances

How can apologies – as communicative acts – put such treatment in place? By taking a certain kind of *stance* toward the victim. A stance is, roughly, a way an agent is disposed to act or to treat someone or something in light of a normative position she adopts. Forgiveness, for example, arguably involves being inclined to act as though an offender should no longer be blamed, resented, punished or called upon to make amends.[22] Maintaining a grievance, political or personal, is a stance: it involves being disposed to raise the complaint, resist reconciliation, perhaps even feel less than content about it. Even being a friend or fan of a baseball team is a stance in this sense: the latter involves being disposed to support, champion, root for and be loyal to one's team, for example, or to act by the position that the team is the best and most worthy of enthusiastic support. Behaviors, attitudes, even emotions will end up figuring into a stance over time. But a stance is not reducible to any of these, as all may figure into acting in line with the normative position embodied in the stance.

Taking a stance is a communicative action by which the communicator both acts on and commits to the stance. For example, to take a political stance – "I'm for Warren" – involves both an act that supports a certain candidate, and a speech act that commits me to continue to act in support of her candidacy. Apologies, on the instant proposal, are cases of stance-takings: through

[22] See Chapter III, *infra*.

the verbal commitment to the victim, the offender both acts on and becomes committed to the *apologetic stance*. That stance involves treating or relating to the victim as one whom the offender owes more than he can repay – specifically, he owes her not to do what he did – and who therefore alone has the discretion to accept or reject his moral repair (since it is insufficient). In this way, the apologizer verbally puts in place a treatment that, by definition, involves *not* treating the victim as violable for a price – as it explicitly involves treating her as someone for whom no "price," no compensatory act, is sufficient. That is just the very treatment committed to; it defines the relationship.

Second, an apology is itself offered as part of the relationship of moral delinquent to moral debtor; it is offered as a necessarily insufficient attempt to make up for the wrong one owes the victim not to have done in the first place. That feature is reflected even in the form the commissive gesture takes – apologies are *offered*, presented as insufficient. One cannot apologize by simply declaring the apologetic content unilaterally with the invitation, "take it or leave it." In apologizing, then, we both *treat* the victim as someone we owe more than we can repay way *and* commit to doing so thereafter.

In this way, the apologetic stance-taking both establishes a relationship that redresses the earlier mistreatment of the victim, and also constitutes an instance of that new way of relating – acting towards the victim as one whom the offender owes more than he could repay. Notice, then, that the apologetic stance-taking involves three elements: a way of treating the victim, a way of committing to such a treatment, and a verbal commissive speech act [23] expressed to the victim. That might raise the question of why have the latter two elements. If the problem to redress was a mistreatment of the victim that began with the initial wrongdoing, why not simply

[23] Austin 1959, 156-160.

treat her in a counteractive way and leave it at that? In essence, that suggestion is reflected in Bovens's proposal, as well – a one-time gesture in opposition to an earlier action (though in his account the gesture is one of self-subordination). Why not follow that structure, if not exactly that substantive proposal? The offender, then, would need only to do some action that treats the victim as one whom she owed not wrong as she did, and present the action as insufficient in the right ways just proposed.

The problem with this pure action-based substitute for apologies, even with the right kind of presentation, is that the state of affairs that the remedial treatment is supposed to reflect – owing the victim more than can be repaid – persists over time. It is not a state one can visit for the duration of a single gesture and move on. The wrong requires *relating* to the victim a certain way, which takes time. So a commitment to actually be in the relationship toward the victim expressed in the stance-taking is necessary. That, however, still leaves the question of why voice the commissive stance-taking to the victim directly. Why not simply commit to a third party, or declare over the radio something on the order of, "I apologize to Smith"? One answer is that the step that acts on the stance, like the treatment it redresses, is directed: it is a way of treating the victim; it is meant to embody a form of relating to the victim. It therefore has to be an act done *to* the victim. Merely committing to the stance in some other form is insufficient, both because it leaves out the action that itself constitutes treating the victim differently, and also because it would no longer be an action done to or toward the victim.

A second reason for the directed apology is that the mistreatment is, as explained earlier, an objective insult – it is a mistreatment to which the victim has reason to take offense (whether or not she does). She suffers the expressive harm of having been mistreated that way. Actually relating to her in the counteractive way discussed here takes time; it is constituted by a series of

acts of which the apologetic action is only the first step. She thus does not have reason to be reassured in a way that mitigates the objective insult until well after the initial act. Yet there is reason to reassure her: one reason to counteract an objective insult is that it is, so to speak, *insulting* – it gives the victim reason to take offense, which would be a wrongful harm inflicted by the insulting action. As a result, the offender should want to prevent that harm by reassuring the victim directly that the insult is repudiated in action. The offender cannot give this reason for reassurance by committing to a third party; it is of little value to a third party, and so neither the third party nor the victim can hold him to such a misdirected commitment. In contrast, a verbal commitment to the victim binds the offender to future behavior, which the victim may claim of him.[24] She then has reason, upon receiving or "uptaking" the commitment, to be reassured that the objective insult of the prior mistreatment is about to end, even before the new relationship that redresses it can be fully manifest. And, separately, she has reason to believe the new way of relating to her has already begun, inasmuch as the apology is itself a way of acting on it, for reasons already discussed.

Why, then, is this account different from other "commitment"-based accounts, like pledging to treat someone better in the future, as proposed earlier? The key difference is that the current account is not about reassuring the victim of some *future* change. It involves committing to an approach that is already wrongfully absent – a certain way of treating the victim that reverses a different way of doing so, already in place, as soon as one wrongs a victim and fails to redress it. True, the object of the expressed commitment is future behavior. But the object of the new behavior – the new treatment – is the past wrongdoing, and the indebtedness it puts in place.

[24] See, for example, Shiffrin 2008, 518: "If one invites trust in a particular way – e.g. by offering a gift or by declaring one is the sort of person to whom another could safely confide in – it is wrong to retract the offer (unless it was clear from the start that a timely acceptance was necessary) or to act in ways that undermine or counteract its value." I submit that the "wrong" here is not limited to solicitation of trust. It applies to any verbal "offering" that has value for the recipient if she chooses to use it (whether or not she actually values it as such).

C. Stance-Takings vs. assertives and expressives

It will be recalled that I rejected the expressive view of apologies, standard among philosophers of language, in part because it required unrealistically that an apologizer be in the throes of an affective psychological state just at the moment she apologizes, on pain of insincerity. I argued, instead, that the act of taking the apologetic stance – which I now claimed is the required form of apology – does *not* need to be accompanied by any such psychological state. At the same time, the stance can be taken insincerely. What, then, is required of sincere apologizers, such that it escapes the excessive demands of expressives while meeting our pre-theoretical demands for sincerity? The apologizer, I want to propose, should – on pain of insincerity – be in and committed to the apologetic stance.

Recall that I defined a stance as a way of being inclined to act in light of a normative position one has adopted and internalized. To return to the example of a political stance, say being pro-choice, one who truly takes this stance would be disposed to direct her actions in light of the position she has taken. In speaking to people who ask about abortion, she will advocate the pro-choice position. In choosing among candidates for office, their position on the legal status of abortion will figure in her evaluation. It may also figure in the advice she gives friends and the causes to which she donates. In much the same way, someone in the *apologetic* stance will be inclined to act by the normative position that she owes the victim not to have done what she did. Unlike being pro-choice, though, the apologetic stance – as argued in the previous section – is bilateral. It involves being inclined to act *toward* someone a certain way; it is a *directed* stance,

one taken toward the victim. The apologizer will therefore be inclined to treat the victim in ways consistent with the stance; to not repeat the offense, to seek to alleviate and sympathize with her suffering, to seek and try to earn her forgiveness, to characterize herself in discussion with the victim in ways consistent with the normative position of moral debtor that she has taken, and much else.

Notice, though, that I described stance-takings as verbal commitments. As noted in the first section, apologies have staying power. Once I've apologized to you, it is in some sense no longer "my" state to change. If I revert to treating you as though my prior wrongful act was appropriate, or that I owe you nothing, or that I should be proud of what I did to you, you can always counter with something like: "Hey, I thought you apologized." In contrast, if I merely privately enter the apologetic stance, I may leave it without anyone in particular having a claim to my commitment. Of course, you may have independent grounds to object – including whatever the reasons I *should* have apologized to you in the first place. But my internal resolution, alone, doesn't give you grounds to charge betrayal later on when I recant. My resolution wasn't yours to hold me to, so to speak.

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. [25] One upshot, then, is that sincerely taking

[25] This way of putting it owes a lot to Seana Shiffrin's account of the "rights transfer" view of promising, which, among speech acts, is the paragon of Austinian "commissives."

stances requires two elements. First, one must enter or at least begin to enter the stance in question; a sincere apologizer should be in or about to enter the apologetic stance. Second, one must also be sincere about committing, to the listener, to *stay* in the stance, at least for some time beyond the utterance.

Notice, then, that I have described stances in a way that, contra the expressive account, seems to leave emotions out of the picture. Internalizing a reason and acting accordingly can be done without any particular feeling or affectation. And yet, it may be worried that many instances of what I've called stance-taking, including apologies, are strongly associated with emotions, often powerful ones. Apologies can be given tearfully and with visible pain. More importantly, as noted earlier, sometimes the expression of these emotions seems sufficient to do the work of these speech acts. If I erupt in visible agony over what I did, and tearfully bare my tortured soul to my victim, she is liable to forgive me on grounds that I not only apologized but did so much more. If a vivid display of genuine emotion can constitute or even substitute for apology, how likely is it that, in fact, apologeticness is a stance, rather than an emotional state after all?

On the analysis of stances, however, this result is not surprising. True, the essential component of stances is something seemingly dry: a disposition to act in light of having internalized a reason or normative position. But it needs to be recalled that human beings *experience* the internalization of normative positions quite emotionally. Consider what the apologetic stance involves: the apologizer internalizes and acts on the view that he owes his victim not to do what he did and can never make up for it; a debt that is greater the more severe the violation. How is this position, if genuinely internalized, experienced? It is, among other things, a state of tension, because it involves taking seriously that it's too late to do what one

should have, too late to undo a violation one had no right to commit. Imagine the way one's body contorts after dropping something, or bowling a gutterball. Constitutive of the state is the yearning, the anguish, to somehow undo what's too late, the regret, the guilt.

And the connection with emotions runs the other way, too: a person who is truly morally regretful, to the point of tears, is likely in that state *because* of appreciating the reason to be in it; she has, in other words, internalized the reason to apologize and remain apologetic. An emotional outburst, then, will likely be the symptom of a stance, rather than a more vivid alternative to it.[26] Suppose, instead, someone was in an emotional state that looked like moral regret – crying, beating one's chest, say, bowing one's head, expressing how horrible the act was – but without any appreciation of the reasons an apology is owed. We might regard such a display, and even the intense state it reveals, as ultimately insufficient, a momentary fit rather than a decisive step toward moral repair.

The same point applies to a similar worry about contradictory emotions: there are emotional states that are plainly incompatible with apologies. For example, I can't be sincerely apologetic while celebrating or happily reminiscing or boasting about the wrong I did. It follows, seemingly, that emotional states are required to be genuinely apologetic. Stances, consisting as they do of internalizing reasons and relating to others in certain ways, wouldn't be enough. But that would be too quick. True, being inclined by a reason to treat someone apologetically, so to speak, is still different from actually feeling badly. But as a matter of natural fact, it seems almost impossible to undertake or adopt such a reason while continuing to feel proud or happy about the wrongdoing. As a natural human fact, those states will almost never coexist.

[26] Thanks to Seana Shiffrin for pointing out this possibility.

Finally, recall also that the apologetic stance is not all there is to sincere apologizing; as a commissive speech act, a stance-taking like "I apologize" also requires *committing* to the stance. The apologizer must not only be inclined to treat the victim in accordance with the stance; he must also continually try to maintain that stance and its attendant dispositions and inclinations. That requires resisting psychological states and embracing others; shunning states like pride or enjoyment of the wrongdoing one did; embracing states like regret, sorrow and sympathy. These are the states that help maintain the stance, at least as actual human beings live them. So while the essential individuating conditions of apologies, as stance-takings, do not require emotional states, it is difficult to imagine a human being meeting them unemotionally.[27]

If I am truly moved to act by my having internalized a position in favor of animal rights, for example, there is no reason that motivation will restrict itself to purely external acts. I would, rather, be inclined to embrace some psychological states – love and sympathy for animals I encounter – and resist, or become alienated, from others, such as enjoying a display of animal exploitation like a cockfight or horse raise, if they should somehow overtake me. Similarly, one who has truly internalized the normative position involved in apologies, and remains committed to it, will be inclined not only to act apologetically toward that person, but to resist and become alienated from bouts of hostility, malice and perverse pride or nostalgia about the wrongdoing.

On this hypothesis, then, a typical person who takes seriously that she stands in disrespect of another person in an irreparable way, and commits to that stance, will likely be in certain emotional states, at various times thereafter, and not in others. The "I'm sorry" expressed by

[27] But the case of non-human institutions, like countries and corporations, will prove different in this respect, as argued in Chapter IV, infra.

apologies, then, would capture a natural way human beings typically experience the apologetic stance or the internalizing of the reasons to take it.

V. Limits of theories of apologies

It may be objected that the present account of apologies, as taking and thereby entering and committing to stances, does not fully capture the phenomenon of apologizing as we know it. The contemporary practice of apologies involves many recognizable features. For example, a typical apology involves the phrase "I'm sorry," and we sometimes apologize for others. These features, and many more, are hardly necessitated or explained by the account proposed here, characterizing apologies as a type of stance-taking.

But that is inevitable. Apologies as a practice must, by now, be underdetermined by the moral reasons for performing them. That is because once a societal convention meets the moral demand for some kind of apology-like act – some way of respecting a victim's right not to have been wronged – it becomes established practice. As a result, it becomes an expected behavior of repentant wrongdoers. That, however, adds to the moral reasons to apologize, and to do it in the specific way established. After all, if that's what people tend to do when they recognize their wrongful behavior, then failure to do so is to single out a particular victim as not entitled to the same respect. It is to act towards her as an exception to the established practice for redressing wrongs done to victims. And that makes it additionally wrong not to apologize to her, or, put differently, that fact puts in place a *new* duty to apologize and to do so in whatever way is conventional. Similarly, the specific conventions associated with apologies may be necessary to

communicate to the victim that one is apologizing. And to the degree that the possibility of uptake is necessary for verbal commitment, those further conventions may be necessary to take the apologetic stance. As a result, further features of conventional apologies may enter into the formal, basic obligation to take the apologetic stance advocated here or the methods for doing so. Nothing in the present account rules out this possibility.

Conclusion

There is much more to be said about stance-taking and apologies. But I have hopefully shown here some of the ways apologies as stance-taking differ from merely asserting or expressing something about the offender, and why those ways capture the remedial power of apologies. In particular, I have argued that wrongdoing – without redress – is also a way of treating or relating to someone, and one that persists well beyond the initial wrongful action. The right stance, then, as a way of treating someone in its own right, can counteract or at least end that aspect of wrongful behavior. In fact, I have argued, it is primarily through stance-taking that utterances like apologies help remedy past wrongs. They take a stance of treating the victim in just the way the initial offense – and the inaction that followed – wrongfully failed to treat her.

There is, of course, more to the actual human practice of apology, especially across different cultures, than stance-taking, just as there is more to stance-taking than apologies. But the merging of these two notions, in the way described here, has hopefully shed some light on an otherwise mysterious fact: simply uttering a word like "sorry," if meant and understood properly, can by itself repair relationships, relieve obligations, warrant forgiveness and mitigate an offender's duty to repair and compensate her victim.

CHAPTER II:

THE PROBLEMS OF AGENT REGRET

The previous chapter argued that apologies work by ending a kind of mistreatment that would otherwise persist from the moment we wrong someone. The account thereby links the need to apologize to the fact that a wrong was done. There is, however, a large class of cases in which apologies, or similar speech acts ("I'm sorry," "Forgive me"), are expected and commonly offered where no wrong was done. I refer here to blameless action, by which I mean actions that faultlessly inflict harm on other people. By "harm," I mean either an injury that is morally undesirable (no decent person would want it to take place),[28] or an insult (even when it offends no one).[29] Examples run the gamut from grave physical injury, as when a doctor administers a proven safe drug that unpredictably harms her patient, to minor snubs, as when a speaker mistakenly excludes someone in listing whom to credit for an idea. Although these actions are done without fault, they inflict harm or at least insult.

When we perform these types of blameless action, which I will refer to as blameless harms or blameless injury, it is considered appropriate not only to say something apologetic,

[28] I qualify harm in this way so to exclude harms one might be justified in inflicting deliberately, such as humiliating an overzealous prosecutor and thereby shaming him to be more careful, or performing surgery, or grounding a child.

[29] "Harm" then is used loosely here, inasmuch as some blameless actions for which we commonly apologize do not cause any kind of loss or suffering: think of speaking disrespectfully to someone who fails to take offense, for example, or mistakenly taking someone's belonging (it looks to be one's own) that she does not value or need anyway. Nevertheless, I will for simplicity refer to all such actions as "blameless harms," "blameless harming," "blameless injury," and cognate phrases, stretching the notion of harm to include what Elizabeth Anderson and Richard Pildes call "expressive harm." Anderson and Pildes, 2000.

along the lines of "Sorry," but to *feel* sorry, as well. Moreover, these responses are treated as reasonable, even on the part of those who fully appreciate that their actions are blameless and above criticism from anyone else. When a driver unpreventably collides with a vehicle that had suddenly swerved into his path, he would be reasonable in feeling bad about what he did – especially if he caused damage – and expressing as much to the victim. As Bernard Williams observed, it is fitting for blameless injurers to feel badly about what they did, even as it is wrong for anyone else to criticize them and they know it. [30] The injurer will view his action negatively, despite his blamelessness, experiencing what Williams calls "agent regret."[31]

The phenomenon of agent regret, so described, raises a problem not widely discussed in the literature on it.[32] The facts viewed objectively warrant no criticism of the blameless injurer – indeed, they refute it – and yet we find nothing unreasonable for him to persist in agent regret, which seems to involve a critical attitude towards himself. Agent regret, in other words, seems to involve an attitude or state of mind that clashes with a reasonable assessment of the facts to which it responds. Specifically, it involves a self-critical state where the facts seem to warrant no criticism at all. Call this the *internal* problem of agent regret.

In addition, there is also what could be called the *external* problem of agent regret: besides *feeling* badly when we blamelessly harm others, we think it appropriate to *express* regret or something like it. In particular, we think it appropriate to apologize, or at least say we are

[30] Williams 1982, 20-40, 28

[31] Williams 1982, 28. Earlier descriptions of the phenomenon may be found, in an obviously very different context, in Maimonides 1987.

[32] The landmark article on agent regret is Williams 1982, but this point is not meant to criticize Williams (or anyone else who wrote on the topic). Williams cites agent regret only as a counterexample to traditional moral theorizing, which he took to require that moral assessments attach only to intentional actions, not their consequences. Any further challenges posed by agent regret to other settled moral intuitions need not concern Williams, except inasmuch as they bolster his case against standard moral theorizing.

sorry or something similar, for blamelessly harming someone else. In fact, this response may be more than simply appropriate. As Adam Smith put it:

> To make no apology, to offer no atonement, is regarded as the highest brutality. Yet why should [the blameless injurer] make an apology more than any other person? Why should he, since he was equally innocent with any bystander, be thus singled out from among all mankind…?[33]

From a third personal point of view, his act is blameless and therefore unworthy of holding him accountable for it. Why, nevertheless, would it be reasonable for the agent to apologize or at least express regret, if only a muttered "sorry," to his victim – and to feel as though he *should* do so? What sense is there in *his* having to expressing regret for something that wasn't his moral responsibility? And why is it improper – "the highest brutality" – for him to do nothing, to walk off and leave his victim to her suffering?[34]

I call this the "external" problem of agent regret because, unlike the feeling that blameless injurers might privately experience, the behavior at issue is visible and interpersonal – a way someone acts *towards* (and in full view of) another. Specifically, he acts apologetically towards the victim of his blameless injuring. My plan is to propose solutions to both the internal and external problems of agent regret, as evoked here. In other words, I hope to account for why it is reasonable for blameless injurers to respond self-critically to what they did, and to apologize or express regret to their victims. The two accounts, for the two distinct questions, build on one-another. The first draws on the investments that moral agents should have. As I will try to show,

[33] Smith, 2002: 123. His example, involving a horse rider who loses control of the suddenly volatile animal, which goes on to injure a pedestrian, caused injury only due to a blameless lack of "excessive care" that would have been inappropriately "timid," as never riding horses ("so far from being regarded as blamable, that the contrary quality is rather considered as such").

[34] It may be noticed that these are very different questions, as my answers in Part IV will reflect.

a moral agent must be deeply invested in not inflicting harm, and this investment gives rise to a self-critical view of the harms she does inflict. The second draws on the treatment of others, indeed the stance, that follows from the investment I have described. That treatment, among other things, explains the insult or disrespect suffered by victims of injury when the injurer fails to take steps to redress what he did. As I will try to show, an injurer – even a blameless injurer – who harms an undeserving victim and then proceeds to go about his business treats the victim insultingly, with behavior that objectively expresses an insult (whatever it reveals, or even ostensibly reveals, about the agent himself).[35] Specifically, he treats the victim as though it is acceptable to have harmed him. Saying he is sorry, or something like it, avoids or ends this mistreatment.

I. The challenge: agent regret looks unreasonable from outside

As noted already, Bernard Williams most prominently pointed out that blameless injurers naturally feel bad about the harm they inflicted.[36] He illustrates the phenomenon with the evocative example of a lorry driver, who – through no fault of his own, including no negligence – runs over a child who had quickly crawled into the street, hidden from view.[37] Although everyone on the scene, including the handful of spectators gathered at the roadside, properly regards the fatal accident as tragic and horrible, the driver alone feels what Williams calls agent regret. He feels a special sort of negative reaction to the fact that he inflicted the damage, even if he did so blamelessly. And, Williams suggests, it is appropriate for him to react that way.

[35] Anderson and Pildes 2000, 1503, 1527.

[36] Thomas Nagel discusses the same overall phenomenon that prompted Williams to raise the possibility of agent regret, namely the phenomenon of becoming morally liable for consequences that are partly due to luck, or "moral luck" as Williams coined it. Unlike Williams, however, Nagel denies that moral luck extends to cases of truly blameless injurers, whom Nagel says do not have to respond self-critically. See Nagel 1979.

[37] Williams 1982, 28.

The example is brought to challenge the claim, sometimes attributed to Kant, that no action should be the object of guilt or moral criticism merely in virtue of its consequences.[38] It is, notwithstanding such thinking, *precisely* the consequences of his blameless action that prompt the lorry driver's agent regret; had the same behavior resulted in no harm, he would experience nothing of the sort. And, Williams implies, this reaction is appropriate, despite the tendency of some ethical theorists to dismiss consequences.

The driver will plainly – and appropriately – feel remorseful about the harm he inflicted. He might experience his remorse as feeling "guilty." But by this he does not take himself to actually be guilty, in the sense of "culpable." Instead, he will feel about himself and his action *as though* he did something with which he disapproves – except without any actual disapproval, and without believing that he did wrong. Agent regret of this sort, then, belongs to the category that Herbert Morris calls "nonmoral guilt"[39] – where guilt feelings are experienced without the judgment that one is guilty of something or culpable in some way; indeed, in many such cases, the judgment would be wrong. Agent regret, in particular, is the sort of state evoked by claims such as "I can't help feeling guilty about it even though it wasn't my fault." A key feature of the remorseful feeling involved in agent regret, then, is that it is *self-critical*: it involves a negative attitude or orientation towards oneself over having done what one did, captured by such phrases

[38] Some claim that moral reactions are appropriately applied only to the will and its selection of maxims, reasons, or motives. Thomas Nagel calls this constraint Kantian (Nagel 1979, 24). Not everyone interprets Kant this way, however. See, for example, Barbara Herman's view in her 1993. Further, Julie Tannenbaum proposes an original account of morally grounded agent regret for harming others. Tannenbaum's account draws on another part of Kant's moral theory: moral agents should value and work to preserve rational agency, on her account, in light of the Kantian principle of treating humanity as an end in itself and the (Kantian) value of rational will. For that reason, they should disvalue actions that fail to realize the end of promoting and protecting rational agency. Tannenbaum 2007, 53-54. My own account, in contrast, derives the self-critical stance blameless harms from the moral practices that must be undertaken on pain of negligence. I also deny that the self-critical stance involves an evaluative judgment of one's actions; indeed, one must judge one's actions above reproach if they violate no moral wrongs, or so I want to allow.

[39] See Morris 1988, 221-22.

as "I feel terrible about it," or "I'm really down on myself over it." Yet it lacks a negative *judgment* – a belief that one actually is worthy of negative regard.

In contrast, examples of regret that is not self-critical include self-pity over having become involved in a tragic incident, as captured by the thought, 'Poor me, what a horrible thing to be caught up in.' In addition, we might contrast mere regret – wising one hadn't done something – with self-critical regret: feeling negatively toward, or "down on," oneself for having done something. Suppose I wish I hadn't signed the worthy petition for gay rights that ultimately cost me a job at a religious institution. But I might at the same time be proud of having had the courage to do it – thinking it praiseworthy, rather than feeling negatively toward myself over it. In that case, I would have regret, but not self-critical agent regret. Blameless injurers like the lorry driver, as Williams observed, *do* have self-critical agent regret, which I will simply call "agent regret" from here on. They feel remorseful about what they did. Moreover, as Julie Tannenbaum observes, this self-critical state is distinctly *moral*: they take a *morally* negative view of what they did, inasmuch as there are moral reasons to disvalue the harm they caused (death, injury, damage to rational agency or dignity).[40]

My task is to explain why it is reasonable for blameless injurers to react that way, in light of an important reason to think otherwise. In calling an internal response or state like agent regret "reasonable," I mean two things, in particular, one positive, the other negative. On the positive side, I mean that the response can be explained by what it responds to, by its object, without appealing to any brute relationship between the two (such as: Y is always the response to X, or X tends to elicit Y). Fear, for example, would be a reasonable response to predators, inasmuch as predators are dangerous (taking as a premise that fear is a reasonable response to danger). In contrast, fear would not – on this view alone – be a reasonable response to a harmless insect, no

[40] Tannenbaum 2007, 53-54.

matter how grotesque, because the sight of the insect cannot explain the fear without appealing to a brute relationship such as the rule that people fear insects. On this view, agent regret would be reasonable if the fact that one harmed someone else, even blamelessly, explains one's having agent regret.

Second, and more importantly, on the negative side: a response is reasonable if there is no compelling reason to overcome or resist it. On the previous example, fear of harmless insects faces a powerful challenge from their harmlessness – the fact that they pose no danger constitutes a good reason to overcome the fear. So on this negative test, such fear is unreasonable. For agent regret to be reasonable, then, there would have to be no fact or reason that, when fully appreciated, would persuade someone to abandon or at least think she should overcome her agent regret.

Notice that on this last test for "reasonableness," it will not be enough to show that agent regret is natural for blameless injurers, or even so utterly natural that we would judge someone "odd" for lacking it. There are, after all, responses we consider quintessentially natural or human that nevertheless face compelling challenges from a full appreciation of the relevant facts. People tend to be proud, for example, of physical traits that distinguish them, such as perfect pitch or an exquisite natural hair color, or – in some cultures – masculinity or femininity. And they are ashamed of others, such as snoring. But it may be appropriately pointed out that they did nothing praiseworthy or admirable to acquire or even maintain their hair color or ability to recognize musical notes, much less some aspects of their conformity to gender stereotypes, just as they did nothing shameful to snore. They have good reason to discontinue being proud or ashamed of these sorts of traits, though we would never expect them to do so.

I mention these last examples because it has been observed (by Harry Frankfurt, for example)[41] that agent regret, too, is natural, as an instance of a more general type of natural reaction to being the cause of something untoward. We do not ordinarily want to be the cause of anything we disvalue. This "natural" desire might be thought to show agent regret a reasonable response after all. But it is, in fact, just the sort of reaction that is open to rational scrutiny to determine whether, despite its naturalness, it is *also* reasonable. If there is a good reason to overcome the feeling, then it is not clear what about its "naturalness" comes to its defense.

The point of the present section is to raise the worry that there is, in fact, such a good reason to overcome agent regret, however natural it might be. In particular, it faces a challenge from the objective point of view. That challenge has been stated already, but bears repeating: we have nothing to criticize about someone who blamelessly harms someone else, even if we know all the relevant facts, including everything the injurer herself knows. Therefore, she has no basis to be critical of herself either, and so should not react self-critically to what she did. Her being above criticism, in other words, constitutes a strong reason for her, too, not to be in the least bit displeased with herself over it.

Consider cases of ethical intervention: someone volunteers, for example, to donate blood upon reading about a stranger who needs it to survive, even though the blood drive conflicts with her long-planned birthday party. She goes anyway, reasoning that the time of the drive – Superbowl Sunday – will draw too few volunteers, and she fears if people like her don't go, the victim may not get the blood she needs. Yet, despite the most rigorous and overcautious pre-testing, her blood, unlike that of the dozen or so others who show up, turns out to be oddly *too* healthy: it is flush with antibodies that well-serve most people but, in this particular unhealthy recipient, overreact and cause a rare, fatal allergic reaction. Had she not intervened and added her

[41] Frankfurt 2008, 10.

blood to the donor pool, he would have survived. Her agency is thus conspicuous in the affair.[42] It is, therefore, just the sort of case in which agent regret might be expected – as the agent caused grave harm, however blamelessly.

Importantly, though, from an objective standpoint no criticism of her is warranted. Her activities reflected nothing but morally ideal behavior, coupled perhaps with extraordinarily bad luck. Indeed, her infliction of harm is itself a result of her saintliness, of traits that should only be celebrated. True, it might be said, she caused a morally undesirable event. And that alone may count as a morally critical assessment. But it comes up short as a ground for her to react self-critically or feel guilty, in light of the fact that it serves as no basis for anyone else to do respond critically to her. Even from an objective point of view, after all, she is the cause of a morally disvalued event – that is merely a straightforward description of what happened. Yet she is still judged above all criticism.

I am suggesting, then, that there should be some consistency between a person's evaluative response to her own actions and that of a third party, at least if both are reasonable. But one might protest that almost no reaction to causing an event can be mirrored by that of spectators, unless they experience it vicariously. For example, pride at hitting a home run cannot be shared by the cheering fans. Conceding that much, I want to suggest that on another level, both agents and spectators almost always *do* share versions of the same evaluative response to some action. Take the case of the home run hitter. True, the agent reacts with pride, while the spectators react with admiration, which is different. But they can also plausibly be redescribed as distinct forms of the same thicker response: admiration, with spectator admiration experienced as positive regard for someone else, and self-admiration experienced as pride. In that sense, the

[42] For an account of why harms from well-meaning intervention ground more agent regret than those resulting from actions not intended to effect the victim, see Driver 1997.

reaction of the agent is not deeply perspectival; it is a form of the same reaction the spectators have. With that framework in mind, we can now see why agent regret is *uniquely* problematic: In cases of feeling guilty about one's own blameless action, the proper response of neutral spectators shares *nothing at all* with the agent's self-critical response. If they have no connection to the players in the affair, their appropriate response will be either praise or pity, with perhaps some revulsion about the tragedy. Yet agent regret is not experienced as self-pity, nor is it a self-regarding version of pity, as might be captured by the thought, 'it's too bad for me that I happened to cause this.' Rather, agent regret is a self-critical state. And that is not at all reflected in the reactions of third parties; criticism, rather, is to be thought unwarranted and unreasonable from their point of view.

The proper (and natural) objective response of blameless harm – that the injurer is above criticism -- challenges her self-critical agent regret. It presents itself arguably as a reason to overcome that unpleasant state. The blameless injurer should, instead, feel pleased or at least content with herself, or something equally far from guilt, such as self-pity over having become involved in such a horrible tragedy, along with sympathy for the victim (just as third parties might feel). Or, equally appropriate, she might feel resentment at the fates for throwing a hapless victim into the path of her blameless behavior. But the fact that criticism is objectively unwarranted seems to render any *self*-critical state unwarranted, too, and therefore constitutes a good reason for her to overcome any agent regret. And there being such a reason, on the definition of "reasonableness" introduced earlier, renders the state unreasonable.

In order to refute this claim, then, and show agent regret reasonable after all, it is necessary to show why, despite appearances, the uncritical assessment from an objective standpoint does not clash with the agent's self-critical reaction. One way of doing this would be

to show that the agent's self-critical state does not constitute or involve a judgment that she is worthy of criticism (which contradicts the objective judgment that she is not). That is what I now set out to do.

II. Proposal: why agent regret is reasonable

Taking stock, I have set the goal of trying to solve, among other things, the internal problem of agent regret – why it is reasonable to feel self-critical, or regard ourselves negatively, for harming others even when the harm is blameless, we appreciate why it is blameless, and nobody else should feel this way about us (and we know it). I now turn to an account that, I believe, does explain these features. First, I will argue that moral agents are engaged – at times unconsciously – in a constant project of not causing harm to others, and are deeply invested in its success. Second, I will argue that this ongoing project, and the extent to which moral agents are invested in it, grounds a negative view of inflicting harm after all, including blameless harms. This view of harmful behavior, though it does not involve self-blame, amounts to a morally self-critical state – one that has all the features already identified here with agent regret.

(i) Moral agents are constantly engaged in a project of not harming others

The first step begins with an observation: moral agents are at all times seeking to avoid causing unjustified harm to others. All but the negligent or malicious engage in an elaborate set of behaviors designed to ward off the possibility of causing injury, which become more intense and urgent as that possibility grows likelier, or the injury more severe. They drive safely, watch where they walk, take precautions when operating hazardous machinery and sanitize the objects they share with others. And, perhaps more importantly, as soon as an activity they thought was

safe appears likely to cause harm after all — driving just above the speed limit on an abandoned road when suddenly a pedestrian crosses — they immediately change course to stave off the danger, trying very hard to prevent it.

These steps reflect an ongoing project of seeking not to harm other people, which intensifies with either the degree or likelihood of harm. Indeed, as soon as they seem about to cause injury, moral agents try very hard – on pain of negligence – to avoid inflicting it. That brings out an important feature of the project of avoiding the infliction of harm: moral agents are deeply invested in its success. They do not merely take steps to avoid causing injury; as soon as they appear on the verge of causing injury anyway, as when their cars suddenly seem to veer towards a stranded pedestrian, they do *all* in their power to avoid it, with extreme effort. The project of avoiding injuring others is, then, one that demands ongoing pursuit and deep investment.

That said, it needs to be distinguished from still more extreme projects with which it is easily confused. Imagine, for example, someone so averse to injuring others that she never drives cars or takes a single step without canvassing the environment for danger. That would amount to more than just being invested in a constant project of ensuring that one's activities remain safe for others; it would frame one's entire life around that project. Instead, the claim here is, more modestly, that moral agents are engaged in this project – that of avoiding injury to others – among many others, though they do not drop the project, and they remain highly invested in its success.

Second, more importantly, it is easy to confuse the project just now described – of avoiding *causing* injury – with that of trying to *prevent* harm from being visited upon another

altogether. While I think the latter is laudable, and often a goal of moral agents, I do not think it fair to characterize moral agents as generally or intensely engaged in it as a rule. They will arguably try to prevent harms to others that they believe they could possibly prevent, especially when intervening involves no comparable cost to themselves (ideally they won't even consider this question). But that endeavor is distinct from the project described here, and it would be implausibly demanding to say that all but the negligent are constantly engaged and deeply invested in it. On the other hand, I claim it is a necessary condition of ordinary moral agency – of being not malicious or negligent – that one seeks not to inflict harm on others, outright, and that one be deeply invested in this project. The project I'm describing, then, is concerned with what results from *one's own agency*. So it includes my aim to tread carefully so as not to accidentally push my companion into the river, but it does *not* include my effort to save him if he was pushed by someone else. Nor is it implicated in my bad feelings about being swept into someone by a gust of wind, however much I wish it not to happen.

Finally, I described the project of not harming others as constant, ongoing. So it is crucial to clarify that this does not suggest it is always on someone's mind. To the contrary: a person taking reasonable care not to hurt someone may, at times, stop even considering whether she poses a danger to others. That is because once they are reassured of the safety of an activity like sitting on a park bench, for example, moral agents may take the liberty of discounting altogether the risks they might impose. They may recline on the bench, perhaps losing themselves in a book or a breeze. That is, perhaps, until they notice, say, that the bench leg is perched on someone's foot; at that point, they will feel compelled to change their behavior immediately. In other words, their investment in avoiding harm tracks what they see as the likelihood of a particular activity causing it; if harm is unlikely, even the non-negligent may ignore the possibility. They need not

act as though preventing injury is worth the all-consuming cost of remaining vigilant even during reasonably safe, everyday activities. What makes the project, nevertheless, constant is that they remain disposed to change their behavior unless they continue to be reasonably assured of its safety to others.

There are, of course, those who argue that taking reasonable care to avoid injuring others requires more, specifically an active psychological state. John Gardner, for example, points out that moral negligence, like its legal counterpart, is identified as a failure to take "due care" in order not to harm others. An essential feature of such "due care," he argues, is *actually caring*, in the sense of forming and maintaining an occurrent intention not to cause harm.[43] Moral agents, on this view, are — on pain of negligence — actively intending not to harm others (as connoted by the phrase "taking care"). My claim here is weaker: moral agents may desist from the active, occurrent state of working to avoid injuring others — indeed, it may play no role in their occurrent psychology for long stretches — just as long as they have no reason to regard their activities as dangerous. But the project of avoiding harmful behavior, and the investment in it, remains in place, as the moral agent is poised to act on it as soon as the likelihood of harm presents itself.[44] The intensity and urgency with which they will strive to act as danger presents itself, proportional to the extent and likeliness of harm, shows their high investment in the

[43] Gardner 2001, 13: "Taking care is an essentially intentional action. One cannot take care not to Φ without trying not to Φ."

[44] I say "poised" rather than "disposed" to act as a way of excluding moral agents who become indifferent to the potential harmfulness of their behavior whenever their behavior appears safe. For them, it is only when their actions suddenly threaten someone's safety that they become jolted into taking due care. Such people might be called reckless, inasmuch as they have no interest in making sure they do not harm someone for long stretches of time. This intentional lack of concern or precaution, in my view, violates the requirement of taking due care even on my less conscious, occurrently psychological version of the required state (as opposed to Gardner's version, which it obviously violates as well).

project of avoiding harm. On either view, then, moral agents are always engaged and highly invested in a project of preventing themselves from injuring others.

(ii) The project grounds a self-critical reaction to failure

The previous subsection argued that a moral agent who is non-negligent is always either actively taking steps to avoid harming others, or poised to take harm-avoiding steps in case – and as soon as – her behavior should cease to be safe for others. In short, moral agents are always engaged in a project of avoiding harming others, in which they are highly invested, whether it's on their minds or not.

The actions to avoid, then, acquire a normative character for the agent. Striving constantly and at times intensely not to perform some action involves (trivially) treating that action as *not to be done*. It is as though one is actually telling oneself, 'Don't do that.' This is a feature that runs across moral and nonmoral projects. Consider, for example, the effort to stay healthy. Suppose someone with an injured shoulder sets out to do his best to recover fully, which requires not raising his arm above his head. So he takes on the project of not raising his arm – or doing any other physical behavior – in a way that will damage his shoulder. But in a sudden, instinctual moment, he notices a vase in danger of falling off a high shelf, and quickly reaches up to block it. The spontaneous action causes his shoulder to deteriorate and prevents full recovery. The patient could reasonably be self-critical of his destructive move, quite displeased with himself over having done the very thing he was deeply invested in not doing. He might even say he feels guilty about ruining his recovery. It is a similar sense of guilt that people may report

when they break their diets or exercise regimens, even accidentally and through no lapse in skill, effort or discipline.

Notice, though, that guilt feelings in these instances – the sense of having messed something up, accompanied by displeasure over having done so – is compatible with there being no basis for anyone else to criticize the agent. From an objective perspective, they did absolutely nothing wrong; what happened was merely unfortunate. And they know it, too. Why, then, do the agents themselves continue to feel remorseful about what they did? As noted, the normative character of a project of avoiding some action orients us against that action. We think something along the lines of, 'that's the very thing that I must not do.' If, nevertheless, we do the action so characterized, we experience ourselves as having violated a project in which we're deeply invested. The clash between what we've done, and what we're deeply invested in not doing is experienced self-critically, or as feeling remorseful about it. So it is with actions that undermine our health and safety (if we're striving not to do them), and with actions that cause harm to others (if we're ordinary moral agents). And it probably characterizes many other projects, as well.

Yet the clash between the project's implicit admonishment (Don't injure someone!) and one's actual behavior is not merely dichotomous – either present or absent; it can be experienced with varying levels of intensity. That is because moral agents are not merely acting to avoid inflicting harm. They become more invested, and engaged more intensely, the more severe the harm in question. Thus Williams's lorry driver would be an extreme case; the project of not causing harm is much more intensely directed against killing a child than nudging a fellow commuter on the subway.

Yet it is important to appreciate that the same sort of clash, which grounds feeling remorseful or self-critical, applies to much lower levels of harm, as well. For example, suppose one calls an old acquaintance, after not seeing her for 10 years, by the name of her late and recently departed sibling. The remark sparks sadness and discomfort, and perhaps vicarious embarrassment for the speaker. But it was done blamelessly; the speaker had uncontrollably confused the two names in his memory, they looked alike, and someone in the room had yelled out the sibling's name in her general direction. It was a reasonable mistake, and would count as blameless. Still, long after the speaker has any contact or interaction with the acquaintance, he likely regrets what he did, bemoaning the misfire of his memory and speech. He did the very thing he was invested in avoiding, namely saddening and embarrassing his acquaintance. Again, his view of this misfire of speech and memory is self-critical. But he does not regard blame as appropriate, nor does he assign himself guilt in the episode. He simply regards himself as having done the very thing (insult others) that he is, even now, deeply invested in not doing. And that is experienced self-critically.

In short, when moral agents blamelessly inflict morally undesirable harm, they have done something they are deeply and actively invested in not doing. That clash is experienced self-critically. This explanation hopefully shows that agent regret can meet both criteria of reasonableness laid out for internal states in the previous section. First, it renders agent regret a predictable response to the facts as a reasonable moral agent views them: that one acted precisely as one remains actively and intensely invested in not acting grounds a self-critical state like agent regret. The response can be explained satisfyingly by what prompted it.

Second, it survives candidate reasons to abandon it. The main reason to think one should set aside agent regret, namely the challenge from the objective standpoint, does not emerge as a

good reason to do so. The challenge amounted to the observation that the blameless injurer is above criticism; there is no objective reason to criticize him for what he did, no matter how much the would-be critic knows. It seems to follow that he has no reason to criticize himself, and that self-critical states like agent regret are therefore inapt. But the challenge misses the mark. The blameless injurer agrees that he has no basis to criticize himself, and will not raise any such criticism. He knows that he should *have* the project I described here – seeking not to harm others unjustifiably – on pain of negligence, but he also knows has no duty to *succeed* in the project as long as he pursues it in earnest. Still, the fact that in harming someone he did exactly what he is striving not to do – the clash between what he did and what he is at all times, in effect, telling himself *not* to do – will be *experienced* self-critically.

This point brings out an important distinction between a self-critical state and a self-critical judgment. A self-critical state – like that of the self-injurer with the bad shoulder, or the lorry driver who harms another – involves an experienced violation of what one is striving not to do. It is, then, compatible with a completely uncritical *description* of the same action: 'A harmed B blamelessly and is unworthy of any criticism.' I can, in other words, know that I am above criticism in injuring my still-recovering shoulder, but still hate myself for doing so – or at least feel guilty about it. And the same can be true of injuring another person, on the account presented here. I know I am faultless, perhaps, but I am displeased with myself over it. A self-critical judgment, on the other hand, amounts to a critical claim or conclusion about oneself, to the effect that one did something worthy of criticism.

Importantly, it is only a self-critical judgment that is directly challenged by the objective evaluation of a blameless injurer and her action. For that evaluation responds only to what can be correctly described about the action, and nothing in its description warrants criticism (indeed, the

description *refutes* criticism). But a self-critical state, such as agent regret, does not involve an evaluation taken to follow from the facts of the case; it does not involve any self-critical judgment. And that is why it is insulated from the uncritical objective point of view, which would refute such a judgment. In other words, agent regret and the uncritical objective assessment simply go past each other, as they are different types of response. The one poses no threat to the other.

III. Objections

A. *Ought implies can* and substantive virtue theory

The account in the last section began with a fact about moral agents, to wit: non-negligent moral agents are deeply invested and engaged in a project of not harming others. This claim was used to impute to moral agents a self-critical state toward their harmful behavior. Such actions involved doing the very thing these agents remain intensely invested in not doing – and that can be experienced as a tension, which is a self-critical state, even if it is devoid of self-critical judgments.

There are, however, familiar philosophical positions that might be seen to challenge this picture of moral agents. One is the view, sometimes attributed to Kant, that acts of will, rather than their consequences, are the proper object of moral concern.[45] Therefore a moral agent need not have any investment one way or the other in whether she *actually* causes harm — only in whether she did her best to avoid it. Thus she may be very concerned to take reasonable care, in the sense of not being negligent. And in practice, this may require taking all sorts of steps meant

[45] Thomas Nagel, for example, argues that cases like Williams's, in which regret is grounded in the result of someone's actions — rather than the actions themselves — do not fit Kant's moral theory, which limits the reach of moral judgments to such results-blind factors as the will and its choice of maxims. Nagel 1979, 24. But see Tannenbaum 2007, 53-54.

to prevent foreseeable harms. But she may, at the same time, have no special investment in whether these steps succeed; her purpose is merely to have fulfilled her duty to take them. In that case, she is not trying hard to avoid injuring others; she is merely trying hard to avoid any behavior that could constitute negligence or recklessness toward the safety of others.

This legalistic moral agent may be comparable to someone who obeys all traffic laws and directives in her driver's manual, despite no personal investment in avoiding accidents. She drives cautiously, signals and makes sure she has room before any change in direction, and stays ready to stop the car suddenly if needed – but, again, only because these activities are required by a set of rules she follows. Perhaps she believes in following those rules out of some sense of fairness to the others who do so, or because she believes laws and even legally backed guidelines are sacrosanct. But as long as she can satisfy herself that she was doing everything these laws put in place to avoid accidents, she is indifferent to whether, as things turn out, she has one anyway.

While this stance may be coherent, there is reason to doubt that moral agents can actually maintain anything like it in practice. True, there may be no *duty* not to actually harm people, and arguably no duty to be invested in not harming people, either; the only duty is to avoid actions reasonably likely to cause harm.[46] But being invested in not *actually* harming people may be essential to *success* at fulfilling the narrower duty of taking reasonable care, or of not being negligent. Adapting the methodology that Seana Shiffrin calls "substantive virtue theory,"[47] one can ask what it is that agents committed to a principle against negligently harming others should do, in order to realize and abide by the principle reliably. There will be other activities, goals and dispositions that they will take up and cultivate, whose importance is implied – if not outright

[46] But see Gardner 2001 for a contrary view.
[47] Shiffrin 2010, 113-16.

mandated – by the principle itself. These are the "virtues" that people properly committed to a principle, or who are capable of reliably following it, will tend to manifest. For example, people who follow the principle of respecting the religions of others will likely do more than merely refrain from disparaging or grudgingly tolerating exotic rituals. They will need to try to appreciate these practices, even seeking to discover at least some value in them. This takes effort, but without it they are likely to become bored or impatient with bizarre practices to which they cannot at all relate. True, such self-sensitization is hardly required, explicitly, by a directive to respect the religions of others. But it is what people who accept or want to reliably follow the principle will, in all likelihood, have to do to succeed at it.

Similarly, in the case of fulfilling the duty against negligent harm, it is difficult to imagine how one might go about avoiding culpable injury to others without trying to avoid injury altogether. The reason for this can be appreciated by reflecting on what reasonable care actually requires on pain of negligence: mainly, it requires that people take due care that their behavior not injure others. This requires, among other things, taking reasonable steps to prevent one's behavior from causing harm once it appears likely to do so. The problem for any moral agent, even a legalistic one, can be illustrated by the case of close calls. These are cases where behavior stands a reasonable chance of causing harm, but almost as reasonable a chance of turning out to be safe. Agents in such cases do not know whether taking preventive measures will work. What they do know is that if a reasonable effort *might* prevent harm, and they nevertheless decline to make the effort, they will definitely be counted negligent if harm results in the end.

Consider a driver like Williams's lorry driver, whose vehicle, at 30 miles per hour, is headed for a pedestrian who fainted half a block ahead. Slamming on the brakes could be futile;

it may already be too late to stop and so one would not be culpable in failing to do so. It could also be unnecessary; a doctor appears to be trying to dash into the street and remove the fainted fellow. The problem is that if the driver – our legalist, let's suppose – declines to hit the brakes because it appears to be unnecessary, he will surely be negligent if, nevertheless, the car *does* hit the pedestrian after all. The only way to avoid negligence in such cases, then, is to act as though one can prevent harm and try to do so, leaving till afterwards the discovery of whether this turned out to be reasonably possible. In practice, the legalist must undertake the same project as the moral agent described earlier: she, too, must often enough be invested in a project of not causing harm. If harm is even reasonably likely, and she does not act to prevent it, she will be negligent if it is inflicted. And that means that even if her goal is only to avoid *negligent* injury, her means of doing so must be to try to avoid injury, period.

In other words, even if one doubts that moral agents *must* be invested in not harming others (as contrasted with merely taking all reasonable measures to avoid it), they would have extreme difficulty – if it is even possible – doing what is necessary to avoid negligent harm, or the recurrence of previously non-negligent injury, without such an investment. And with that investment, I have argued, they will experience those harmful actions self-critically. It also bears mention that the project of avoiding causing harm to others is not merely one that arises each time harm is imminent. It is, rather, an ongoing project, which takes experience as evidence to be used in the future.[48] So upon realizing that some blameless behavior turned out to be harmful, in a way he could be excused for failing to anticipate, a moral agent would be now inclined to regard that behavior as the very sort he both *should have* avoided and, from now on, *should* avoid. That figures into his now improved approach to the project of not harming others – and it,

[48] This point is made by Barbara Herman to explain how a moral agent who adopts the Kantian constraint that only culpably willed actions can be wrong might nevertheless have a duty to engage in moral repair after accidental harm. See Herman 1993, 102.

too, orients him self-critically toward that behavior, even if it does not ground a self-critical judgment.

B. Too high a moral standard?

The picture of moral agency suggested by my account of blameless self-criticism may seem implausible in several ways. First, it may seem unrecognizably saintly or severe. That is in part, perhaps, because of the seemingly constant nature of the project of avoiding harm as I have presented it, involving an ongoing project in which moral agents are deeply invested. But "ongoing project" and "deeply invested" are in fact rather open concepts. Parents, for example, could be aptly described as deeply invested in their children's dental health (besides financially). This means both that they will try not to damage their child's teeth, making sure that their day-to-day activities are not harmful to their child's enamels, for example, and that they will immediately stop themselves when they realize they are about to inflict dental damage, such as by leaving a jar of caramel in reach. One can correctly describe this disposition as an "ongoing project," inasmuch as *every* activity of the parent is intended, *inter alia*, not to inflict such damage, and *every* activity that appears likely to cause such damage will be stopped, once detected.

Notice, however, that such an approach hardly impacts most of the parent's day-to-day life. Most of his activities will bear no trace of this investment, or the efforts in support of it – even as, all the while, the parent stands ready to cease or avoid any impingement on the child's teeth, should the prospect come up. That is because very few activities need be done differently as a result of the project or her investment in it. Similarly, many human activities bear no risk of causing meaningful harm to another person. And as I mentioned, a moral agent's investment in

not harming others is but one of a perhaps infinitely many investments she may have. Thus many daily activities will barely implicate it. Still, every activity that moral agents perform is meant, among other things, not to be dangerous. It just usually takes little effort, in practice, to keep one's behavior safe. Moral agents, in other words, engage in this constant project, remaining deeply invested in its success, without being noticeably affected by doing so.

Another worry is a self-critical reaction to blameless harms implies perfectionism. It seems to implicate a standard that insists on success even at the impossible task of preventing the unforeseeable or unavoidable. That, however, overstates the nature of the self-critical state described here. It is not a finding of failure to meet some standard. Indeed – as emphasized repeatedly in the previous section – it is not a finding or a judgment at all. The self-critical take implied here amounts to nothing more than the experience that one did something that one was, in some way, deeply invested in one's not doing. It is accompanied by no moral evaluation, nor does it involve acceptance of blame, punishment or even the victim's resentment. It is compatible with rejecting, outright, all of those responses.

C. Third party consolation

Still, the reasons to view one's own behavior self-critically, even if it is blameless, run up against a powerful observation Williams raises in discussing "agent regret." As already noted, third parties should react very differently to the blameless infliction of injury than the injurers themselves. One way, not yet discussed, is that they try to console the blameless injurer, perhaps even trying to dissuade her from feeling guilty. And, as Julie Tannenbaum points out, these third party stances are appropriate.[49] It is perfectly in order, in other words, to try to console blameless

[49] Tannenbaum 2007, 56-57.

injurers, especially if they are consumed by guilt. How, then, can the appropriateness of third party consolation be reconciled with the self-critical state that, on the argument just presented, moral agents reasonably enter when they blamelessly inflict harm?

One way to reconcile the two standpoints is to show that they are compatible. The moral agent's reasons to be self-critical stem from his own investment in not harming others, and the realization that he did the thing he was invested in not doing. The third parties, in contrast, are not in the same way engaged in a project of *his* not harming others or in *his* proper use of efforts or actions, and they certainly weren't directing *his* actions so as not to result in harm. So what happened does not represent a clash of *their* actions with any of their projects or investments; the clash was his alone. More importantly, as already noted, they correctly assess that no self-criticism is warranted. While a self-critical state is reasonable, as argued in the previous section, a self-critical *judgment* is unreasonable, indeed wrong. That leaves the observers with the realization that the injurer is both above criticism and, at the same time, suffering (undeservingly, at that). The consoler's target, then, is the injurer's suffering as a result of being moved into the painful self-critical state, however reasonable on his part.

Moreover, it is properly regarded as unfortunate that the injurer must suffer pain and anguish as a result of a self-critical state he reasonably enters due, in large part, to events beyond his control. That does not mean his self-critical reaction is mistaken. It only means that it is right for the rest of us to want his *experience* of that view to be less painful. As a result, the consolers could reasonably seek to alleviate that suffering, by focusing his attention on his innocence in the affair. Their consolation, then, would not be aimed at the injurer's appropriate self-critical reaction; it is, rather, aimed at his pain, by focusing his attention on the *limits* of the criticism to which he should subject himself. In particular, that criticism should fall short of blame, or any

judgment whatsoever. They rightly remind him of this limit as a way to alleviate his suffering over what he did, even while this reminder – and the attempt to make him feel better – remain compatible with the morally self-critical state he enters.

IV. The external problem

The past two sections have hopefully shed light on the phenomenon of agent regret, inasmuch as they explain why it is reasonable to react self-critically to the harms we blamelessly inflict. But that leaves another half of the problem still unsolved: why, in addition to *re*acting self-critically, do we also *act* self-critically in a very visible, not to say "external," way – apologizing or otherwise expressing regret, or at least saying "sorry," to those we blamelessly harm? This question is not merely that of why it is reasonable to express the internal agent regret already defended in the previous sections. The practice of external agent regret goes beyond giving voice to a feeling: it is, indeed, a practice. The "sorry" uttered by the blameless injurer – even the commuter who accidentally shoves another hidden from view – is a deliberate and recognizable action, performed out of a sense that it *should* be performed.

In other words, external agent regret – expressing regret to the victims of blameless harms on the belief that we should do so – reflects a belief that it is better to voice some such expression than simply to move on. Is that belief, and the performance that acts on it, reasonable? There are grounds to think otherwise. Take a case of an obviously blameless accident: you're driving a car while the passenger to your right is sipping a mostly empty cup of coffee. Suddenly, an undetectable optical illusion at that part of the road causes you to think you see an object in your path, which reasonably spurs you to stop short, causing your passenger to spill her scalding coffee on herself. Having witnessed the whole thing, she knows full well that it

was accidental and beyond your control (you know she knows it, and she knows you do, too). She would have no reason to resent you or demand that you clarify that it was unintended; indeed, she needs no information from you at all. Nor can she hold you responsible or accountable for what happened, since it was unintentional and faultless. What purpose could be served, then, by uttering "sorry" such that doing so would be not merely polite but called for, and not doing so at all problematic? If we are truly blameless in injuring someone, then why not just go on, leaving the victim to her suffering – especially if there are others on hand better able to help and comfort her (suppose it's a three-seater, and the opposite passenger is already ably cleaning up the mess)? Why should we say anything at all?

The phenomenon I'm questioning, then, and the form of the question, differ from the case of internal agent regret. Nevertheless, external agent regret is, in fact, deeply related to the internal version. As I will now try to show, the solution just now proposed for the internal problem lays the groundwork for solving the external one. That is, it helps explain the sense that we should apologize or at least express regret to victims of blameless injury. The argument proceeds in several steps.

First, I will show that the arguments about internal agent regret – of the last two sections – also show that moral agents naturally treat each other as people it is bad to harm. I will then argue that people therefore have a well-grounded objection to being treated otherwise. I will then argue that leaving victims of one's blameless conduct to suffer, moving on without a word, breaks the treatment I have described, which people reasonably expect. It treats them as people it *is* acceptable to harm, as I will show, contrary to what they would reasonably expect. And, as mentioned, they have a well-grounded objection to such treatment. Finally, apologetic behavior

prevents this mistreatment. That is why it is reasonable both to express regret to the victims of one's blameless harm and to believe one should do so.

(a) The other-valuing stance

Recall that the preceding sections sought to account for agent regret on the basis of a project in which moral agents are always engaged and deeply invested, on pain of negligence. That project was that of avoiding harm to others. The nature of the project grounded a self-critical internal response to injuring others – an experienced clash between what one did and what one is deeply invested in not doing. But it should be clear that the project itself is decidedly *external*. Moral agents do not merely *wish* not to harm others. Their investment is manifest in action. They take steps to ensure they are safe to others – from watching where they walk to monitoring the meanings of their words, and the impact of their actions. And, as I took pains to argue, they are not merely going through the motions – they strive not to cause harm, with deep investment.

In short, moral agents find themselves in what the previous chapter described as a *stance*: a way of relating to someone or something in light of a normative position one has adopted. In this case, their stance involves treating others as not to be harmed, and they indeed do so in light of an internalized position that it would be bad to harm them, a view reflected both in their external care-taking and their internal agent regret when they fail. Moral agents, on pain of negligence or worse, generally adopt this stance – they treat others as people it would be bad to harm, and very bad to harm seriously, no matter how blamelessly.

Importantly, if obviously, the stance I have described is valuable to its beneficiaries; it is *better* to be treated as one it is bad to harm than to be treated the way the "legalist" described in

the previous section treats others – as someone it is fine to harm blamelessly, as long as he is morally in the clear. While there may be no moral prohibition in being treated this legalistic way, people would prefer not to be. They want others to disvalue harming them.

The next important step is to appreciate that the general tendency of moral agents to be in the stance I have described – treating others as people it is bad to harm, because of adopting such a view about harming them – grounds an *expectation* to be treated in this valuable way. The doctor who gives us treatments, asks how we're feeling, checks the safety of all the procedures she performs and the medications she prescribes, is reasonably thought to *actually* be invested in our health, treating us as people it would be bad to harm. We would be surprised to learn otherwise; to discover, for example, that she is merely out to preserve her moral clean hands and avoid being culpable of negligence or worse, with no actual investment in our wellbeing. We would be surprised to discover that she is, actually, a version of the legalist as described in the prior section.

From this point another closely follows: it is disappointing, or worse, to discover that someone does not adopt this other-regarding stance, instead acting with no genuine investment in not causing us harm. This disappointment may be experienced as an insult, or it may not be experienced at all, but there is definitely grounds to be disappointed and insulted by it. That explains why someone would be insulted by negligent behavior, over and above the imposition of a risk of harm. Someone who rushes through the train station, thrashing about without regard to whom he may injure or at least ruffle, is problematic in *two* separate ways: first, he endangers others, which is just a derivative wrong of negligently inflicting harm; second, equally important for present purposes, he acts with disregard for others. Although he will incur no moral

responsibility if no harm results, the sense of disappointment and insult remains: he is failing to treat us as people it is bad to harm.

Finally, and most importantly: this callous treatment – as people it is acceptable to harm – is not merely perpetrated by acting negligently towards us and risking harm. It is also implicated in moving on after *non*-negligent or blameless harm, acting as though nothing has happened. Put simply: when we harm someone and then act as though nothing happened, casually moving on with our affairs, we treat the victim as though it is fine or acceptable for us to have harmed her. Importantly, the treatment is that it is acceptable "to have harmed here," rather than that it is fine or acceptable *to harm her*. (The latter, unlike "to have harmed," would amount to a different insult, suggesting I may intentionally inflict harm, it being acceptable to do so. Here, in contrast, the insult lies not in the injurious action, but in what the injurer does *afterwards*.) The insult expressed by moving on, after blamelessly injuring someone, is that it is fine *to have harmed* someone. Recall that an objective insult is an action that treats someone as though some insulting claim were true. In this case, it is to treat the victim as though it were true that it is entirely acceptable or of no consequence that one harmed her (however blamelessly). That putative "truth" is insulting, not least because it is a disappointing departure from the standard treatment of moral agents. Doing nothing, simply leaving her to her suffering, thereby amounts to an objective insult. To be clear: I am *not* suggesting that there is anything untoward in harming someone blamelessly. The objectively insulting behavior begins when, despite having inflicted harm – albeit blamelessly – one simply moves on and does nothing about it.

From this it follows that people should not simply move on after harming another person, even blamelessly. That action, or inaction, amounts to an objective insult to the victim, particularly in light of the expectation that a different stance is in place. But it does not obviously

follow, on what has been stated so far, that any particular action could redress this insult. All that follows is that whatever is done instead of doing nothing, it would avoid the objective insult just described if it treated the victim as though harming her was *not* acceptable.

(b) Why say "sorry"?

I want to argue now that expressing one's agent regret to the victim, or something similar, can be one way of preventing or preempting the objective insult that lies in doing nothing.

This point may, at first glance, seem not to need argument. If the goal is to refute or preempt the objective insult that harming someone doesn't matter, or is acceptable, it may seem straightforward that the best means to achieve that goal is to express this refutation to the victim. If my actions would otherwise treat her as though it is acceptable to harm her, then the solution is simply to tell her it is not acceptable. The obvious remedy, in other words, would be to un-say the insult, to so speak.

But this simple answer gives rise to an explanatory problem. Consider, first, that the insult is objective, rather than something (necessarily) communicated intentionally by the agent. We want to allow, in fact, that people who harm others and move on – those who inflict the objective insult in question – may not actually believe the insult they objectively express. They may, in fact, *disbelieve* that harming the victim is acceptable. Our careless commuter, just now mentioned, may simply be caught up in the rush home, and may – if asked – reveal a sincerely negative view of harming others. In other words, believing the insult is not necessary to inflicting it. If so, then the victim's learning that her injurer does not believe the objective insult – by way of him telling her he hates causing injury, say – would not seem to make much difference. We

could imagine a victim of a blameless injury, such as the coffee pilling described above, tracking down her injurer and asking him: "Do you really believe it's okay to do that to me?" And we can imagine him answering, "No! Absolutely not," and her judging him sincere. But she would still have a legitimate grievance against him, charging that – his beliefs notwithstanding – his *actions* treated her as though it was acceptable to injure her. If this exchange does not remedy the insult, then it seems unexplained how the injurer can do so by simply informing the victim that he does not believe it is acceptable to harm her.

One thing missing, in the injurer's mere report that he disbelieves it is acceptable to harm the victim, is *action*. Specifically absent is action directed toward the victim that amounts to a way of treating her (just as the objective insult is a different way of treating her), as one whom it is unacceptable to harm. I propose, then, that the action of apologizing to the victim, along the lines of making an active point of saying "Sorry I did that," or "Pardon me for that," meets this criterion, because of at least four elements: 1) it is a break, an interruption in one's activities. 2) It is itself a form of rejecting or not accepting what one did; 3) it is done as something owed or required, or at least as something one *should* do; And 4) it is done *to* the victim, as part of the way one relates to the victim.

Actions, like apologies, that involve these elements are sufficient to constitute treating the victim as though it was not acceptable to do what one did. The first element – interruption – amounts to an avoidance of simply moving on as though nothing has happened. It is, in fact, quintessentially an act of *not* moving on after harming another, a deliberate refusal to simply leave the scene. Second, an apology is a way of acting out one's non-acceptance of what happened. Accepting or not accepting a state of affairs is not merely an attitude; it is constituted by action or (often in the case of acceptance) inaction. Consider the performative: "I do not

accept that," or acts of protest as a way of rejecting some position or state of affairs. Social space has made room for human behavior that constitutes their pro- or against- relation towards something or some event. When we declare, "I do not accept X," or "I take no part in Y," we do so performatively, acting out what we say. Indeed, it is a paradigmatic stance-taking – both committing to the non-acceptance and, at the same time, acting on it. No further action is necessary. So an apology or expression like "I'm sorry" enacts my non-acceptance of what I did, much the way a protest or similar public act might enact resistance to something.

Third, the apology enacts my non-acceptance not merely as an isolated response, but as something I *should* do. The speech act, or at least the act of doing *something*, is performed as something owed, or preferable, as though the agent does not permit himself to simply move on (quite apart from not wanting to move on). It has the ritual character of an obligatory – if still deeply sincere – gesture, like certain salutes or expressions of condolences or congratulations. It is as though the apologizer communicates, in addition to the apology, the message: apart from how I feel, I have a duty to do this or at least to not move on. That transforms the agent's speech act from one enacting *his* non-acceptance of what he did to one treating the injurious action as, in general, not *to be accepted*. Put differently, it acts as though what was done is both not accepted *and* not acceptable.

Still, acting as though one's having injured someone is unacceptable, as these last elements amount to doing, falls short of *treating the victim* as though it is unacceptable to harm her. Put crudely: acting as if X is not the same as treating S as if X. To constitute a way of treating someone, an action has to be done to her, or towards her. If I announce to my partner that I oppose some candidate for mayor, I have definitely acted as though this candidate is unworthy of support, but – until I vote, say – I have not treated the candidate this way. My action

may be about the candidate, but it is not done *to* her or in the context of relating to her. There is, in other words, no intelligible way of seeing it as a treatment of the candidate. When we harm someone, however, and then move on, the insulting treatment is done *to* the victim. We harm *her* and then we leave *her.* Therefore, the set of actions aimed at remedying this insult or blocking it need to be given this same directed quality. They, too, need to be ways of treating the victim.

How can that be done? Mostly, that work is done in element (4). The act of non-acceptance is performed as an act owed or entitled to the victim. Indeed, all the steps have elements of this quality: the injurer interrupts his affairs and heads *towards* the victim or stays *with* her; he expresses his rejection or non-acceptance of what he did *to the victim*; and he does so out of a sense of owing *to the victim* not to accept what he did. In this way, he does not merely act as though harming her was unacceptable; he *treats the victim* as though having harmed her was unacceptable. And that, straightforwardly, counteracts or prevents the objective insult of treating her as though harming her is acceptable. In that way, an apology – and similar expressions of regret, sorrow or rejection of the action – is a way of avoiding the insulting treatment of leaving the victim to her suffering.

(c) Clarification

On the account just presented, apologies serve to preempt or counteract a way of treating the victim as though it is acceptable to have harmed her. They do so by taking an active step that rejects, or acts on the non-acceptance, of one's past injurious behavior, as an action owed to (and therefore directed towards) the victim. In so doing, the agent acts towards the victim as though she owes her to not accept what he did, or (which amounts to the same) as though it is not acceptable to have harmed her.

This account has several notable features. First, it does not require the "apology" in question to be a full-on, ideal apology as is typically demanded of wrongdoers. One can satisfy criteria 1-4 above simply by expressing one's regret – "I feel terrible about what I did" – as long as it is an act done to the victim that constitutes one's non-acceptance of what one did as a way of acting on the principle that harming her is, in general, unacceptable. But it need not, and should not, involve such elements as accepting responsibility, admitting fault and pledges to repent – paradigmatic features of apologies for wrongdoing.

In fact, the difference between the two types of speech acts – "apologies" for blameless harms, as described above, and those for outright wrongdoings – can account for a familiar feature of the former. Often victims of blameless harm will rebuff the injurer's apology, insisting that it is unnecessary or even inappropriate.[50] Yet we do not interpret the victim's response to mean the injurer should have done nothing, moving on without a word. "Sorry" or "Forgive me" are simply ambiguous; the same locutions stand for apologies for wrongdoings, as well as those meant to act only on one's non-acceptance of what one did. The victim should appropriately reject the standard, paradigmatic apology for wrongdoing, even if not the expression "I'm sorry" and its cognates.

A second feature of this account is that it does not explain why apologies, or some similar speech act, would be *required* of blameless injurers. It merely shows that they constitute one sufficient means of remedying an objective insult, namely treating the victim as though it is acceptable to have harmed her. There may well be other ways of doing the same, and the arguments here do not specify the normative pull of the need to do it, except that it is stronger than etiquette (because the objective insult is not set by convention, and can be shown insulting

[50] Thanks to Barbara Herman for pointing out this feature of the "blameless apology."

independently). But it bears mention that such ways are not easy to find. For example, suppose that instead of moving on after we blamelessly harm others, we took some step aimed at remedying the harm we caused. After we bump someone blamelessly, causing her to fall and suffer a minor scrape, we would – for example – help her up and give her a band aid and antibiotic treatment. Or, more coarsely, we would drop a few dollars her way for the treatment, and cajole someone else to help her up (presumably we're in a rush, after all).

The problem with these behaviors is that, as we saw in the discussion outright wrongdoing in Chapter I, harming and repairing, or harming and compensating, does not treat the victim as someone whom it was unacceptable, or – as in cases of grave harm – horrible, to have harmed. It is, equally, a way of treating the harm as a behavior that has a price, which can be dispensed with and otherwise ignored, even tolerated. We see this kind of treatment in the behavior of tyrants or careless government agencies, who in the course of pursuing their plans or policies, end up damaging someone or her property. Then they simply send money or provide the necessary repairs to the affected subjects. This does not amount to treating the victims as people it was unacceptable to harm. It treats them as people whom harming is fine, just potentially costly. It fails to involve any non-acceptance of what one did, at least not on an objective reading of the behavior in question.

There are, however, ways to make such remedial steps more likely to perform the function of apologizing as described here. All that would be needed is that they be the sorts of behaviors that, looked upon objectively, treat the victim as though it is unacceptable to harm her, or act with non-acceptance of the harm. The philosopher and activist Sari Nusseibeh, of Al Quds University in East Jerusalem, describes a practice in Palestinian culture done in response to accidentally harming someone else (including when the harm is blameless). Nusseibeh was once

involved in a car accident that slightly injured an elderly woman, though it was understood by all that he was blameless. Although the woman recovered, Nusseibeh's father insisted he needed to do something to remedy what he did – despite everyone's awareness of its blamelessness.[51] So he followed the local custom, which requires that he meet with the woman's kin and not only apologize but offer the head of the family millions of dollars by way of compensation, even though that is far in excess of the damages (she was fine).[52] Custom also dictates that the offer be rejected. The self-punishing offer, in this context, does not suggest that harming is acceptable though it has a price; the exorbitance of the price is supposed to express precisely the opposite: that the injury was extremely unacceptable.

An apology can be an even more direct way of treating the victim as though the harm was unacceptable. It simply *is* a form of not accepting what one did, and it is directed at the victim as something required or owed to her. In this way, apology prevents or refutes the wrongful act of objectively insulting the victim. Still, it bears re-emphasizing that although an apology, or an apology-like expression, fulfills a necessary moral function, it may not be the only means of doing so, as Nusseibeh's example illustrates. What is owed, on the foregoing argument, can be put in purely negative terms: we should *not* leave the victim to her suffering, nor should we do anything instead that likewise amounts to treating her as someone it is acceptable to have harmed. From this point it follows only that blameless injurers must do *something* to stop this mistreatment, and something that – unlike leaving (or compensating) – constitutes non-acceptance of it, and which is done to the victim. But I want to allow that there could be multiple mechanisms by which one might avoid or counteract this insult. I have proposed merely one such

[51] Nusseibeh 2007, 165-66: "I explained it wasn't my fault…Father said, "this time you've really blown it…By not apologizing, you impugned the honor of their family and ours."
[52] Ibid, 167.

mechanism, which I claim is involved in the already common practice of apologizing or saying "sorry" or "pardon me" for blameless harm.

(d) Intentional blameless harms

On the account proposed here, apologies – or any speech acts performed to express internal agent regret – are among the ways to avoid an objective insult of treating the act of blamelessly harming another as acceptable. But this seems inapt for at least one class of blameless harms: those done intentionally.[53] For acts taken deliberately, it is hard to show that one does not accept the harm; after all, the agent knowingly caused it. Consider a manager of a corporation branch who is forced by the firm's executives to fire all but three of her 40 employees. She has no way of avoiding the act, but it will impose considerable harm. It seems appropriate, even required, for her to do something – apologize, say, or express her regret. But it would seem these acts cannot perform the usual function of counteracting the objective insult of treating the victims as though harming them is acceptable. Here, harming them is precisely what the manager intends to do.

Or consider an even starker example, that of morally necessary harm. In a variety of cases, we must inflict harm for moral reasons. For example, we may have to steal a neighbor's taxi to take someone to the emergency room, even if it causes the neighbor to miss a flight and a job interview, at great personal expense. In this case, I believe, it would still be an objective insult to simply leave the incident in place, never apologizing to the neighbor, even once the neighbor learns and accepts the good reason for the affront. On the other hand, the agent clearly

[53] Thanks to Herb Morris for pressing this example.

thought the harm was acceptable, as inflicting it was actually required – so how could he fulfill the requirement to treat it as *un*acceptable?

This diagnosis, however, would be too quick. These two cases only show that some blameless harm, for which apology is appropriate, can be inflicted deliberately. It does not quite show that the injurers in such cases *accept* what they are doing. In both the case of the forced firings and the taxi stealer, the harms are best seen as unwanted double effects. The intended action is to carry out one's compelled professional duties, or to get someone to the hospital. If there were ways to do the same actions without causing harm, the injurers in these cases presumably would avail themselves of them. That one is forced to inflict harm, then, is something one wishes to resist, protest, stand against. One does not want to treat this as acceptable, nor treat her victims as though harming them is acceptable. In these cases, then, apologizing in advance would still serve the same functions described above: it would constitute the injurer's active non-acceptance of the infliction of harm he stands to cause; it would involve a step one does not permit oneself to harm others without taking (as in "interruption," above), and it would direct these behaviors toward the victim. None of that changes with the temporal vantage point from which these behaviors are directed.

V. Between internal and external

It will be noticed that none of what has been said here about external agent regret requires that it express the self-critical agent regret described earlier. The disrespect, or insult, redressed

by apology happens *after* the harm was done, whereas the internal, self-critical agent regret concerns the act of inflicting harm in the first place.

That said, the two phenomena discussed in this chapter – agent regret and apology for blameless harm – are, as I have already said, deeply interrelated. One reason has already been presented: the project that grounds *internal* agent regret also puts in place the expectation, on the part of victims, that helps make doing nothing after the injury disappointing and insulting. But there is another way that internal and external agent regret relate: they reinforce each other. Recall that the self-critical stance moral agents take toward harming other people arises from their ongoing project of not causing such harm. This project becomes more intense and urgent the more seriously moral agents take the suffering of those they would harm. So even before inflicting blameless harms, moral agents are already acting towards various would-be victims as though it is unacceptable to harm them. They are manifesting that non-acceptance by trying hard, with strong moral valence, not to inflict harm in the first place. They are not accepting the prospect of causing harm in the future. Indeed, they view such hypothetical actions critically, albeit in a way they do not yet direct against themselves.

Second, the self-critical way blameless injurers react to their harmful behavior helps render their formal non-acceptance of it, through the speech act of apologizing, sincere. Being self-critical of something I did implies that I find it unacceptable, and that view is at least part of what might be expressed by a formal speech act that constitutes my non-acceptance of it. Finally, the moral requirement that we try very hard not to harm others may reflect similar values as the requirement not to treat them as though it is acceptable to have harmed them. Or, put differently, the reason ignoring the harm, after it was inflicted, constitutes an insult might relate to the reasons we try so hard not to inflict it in the first place. There is, of course, dispute about what

those reasons may be. But on the arguments offered here, injuring others is to be viewed both before and after the fact as unacceptable, regardless of the agent's moral status in causing the injury.

This negative view of harming others not only connects internal and external agent regret. Arguably, it also explains any apparent continuity between agent regret for blameless activity and guilt for outright wrongs. The reason for self-criticism when one harms another, even blamelessly, may stem from the same source as the reason to feel guilty for culpable harms. In other words, morality may require that we disvalue harms suffered by others, period, which could help explain both why we must not intentionally inflict them, and why we must intentionally avoid them. I offer this possibility only as a suggestion, for now, which might shed even more light on why we criticize ourselves when we harm others, blame ourselves when we do so deliberately or negligently, and why the two reactions are sometimes hard to distinguish in experience. Either way, in harming others we caused an event that, as moral agents, we disvalue, refuse to accept, and never stop trying to prevent.

Conclusion

I have sought to make sense of what I take to be widely practiced, but philosophically puzzling, responses to blamelessly inflicting harm. The two responses differ – one is the internal feeling of agent regret; the other is the external utterance of "sorry" or cognate speech acts. But they raise the same difficulty: they reflect the injurer refusing to excuse himself for what he did, even though objectively his action is entirely excusable, even above criticism. Rather than argue that the two perspectives – internal, first-personal and objective, third-personal – clash, and one

is mistaken, I sought to reconcile them, showing why the agent may reasonably be self-critical and apologetic even as the rest of us, taking the objective point of view, should refrain from criticizing him or holding him accountable.

The self-critical response to blamelessly harming another person, I argued, is how a reasonable moral agent experiences the tension between what she did (harm others) and what she is deeply invested in not doing. Her investment grounds an experience of what she did as something not to do, something she is against herself doing, but which she did anyway – a self-critical state. On the external side, the expression of "I'm sorry" or "Forgive me" ends or prevents a way of treating the victim that is otherwise put in place by harming her and leaving it at that: treating the victim as one whom it is acceptable to have harmed. Apologizing changes that, by acting to not accept one's injurious behavior, this action being directed towards the victim, and performed as something normatively required.

This way of framing the issues is addressed primarily to those who experience or appreciate the phenomena of internal and external agent regret, but who have not fully been able to make sense of them. Other readers, however, may resist that starting point, feeling that agent regret is foreign to them or that they need not apologize for blameless harms. To them I offer the foregoing arguments not as an explanation of a familiar reaction, but as a positive case for them: a reason to have self-critical agent regret after blamelessly harming someone else, and a reason to apologize.

CHAPTER III:

FORGIVENESS

In the past two chapters, I sought to explain how wrongdoers, and even blameless injurers, can make moral differences by apologizing and engaging in similar acts of moral repair. The arguments relied on the claim that we treat people in objectively insulting ways when we wrong or harm them and do nothing about it, and certain speech acts like apology – and making amends more generally – can stop this mistreatment. Notice, then, that at least some of the reasons to apologize and make amends are objective: treating people a certain way is problematic, I have argued, for reasons accessible to anyone who attends to the facts of the case. And for those same reasons, apologies and the like are needed to work against the problematic treatment otherwise in place.

But this objectivity is challenged by the power of forgiveness. A victim can, through acts of forgiveness, *change* what her offender owes her by way of redress, even when none of the facts that call for redress have changed. She can render her offender less open to criticism, to blame and to demands for moral repair, despite the independence of the reasons that give rise to them. Forgiveness, in this sense, is extraordinary – and problematic, on all that has been argued so far. My purpose in this chapter is to explain this perplexing feature of forgiveness, by way of a general account of the phenomenon, and show why it is actually compatible and even consonant with the arguments about apology so far.

First, though, the meaning of "forgiveness" in this context needs to be specified, because the term is ambiguous. Forgiveness is often characterized as an internal mental state that we

might reach or experience. It is, on this view, either a state of mind,[54] such as a feeling or its absence (like the end of resentment or bitterness),[55] or a process or commitment toward attaining one of these.[56] Consider: "I'd hoped that I would have forgiven him by now, but it just hasn't happened." There is, however, another, equally natural sense of forgiveness, namely forgiveness as a formal communicative act.[57] This sense of forgiveness, which I will call "formal forgiveness," is the act of uttering to one's offender the speech act "I forgive you," or something the speaker takes to mean the same. This sense of forgiveness is most commonly associated with forgiving a debt or pardoning someone, as in "I pardon you for the wrong you did to me."[58] It is also what is sought by the plea: "Please forgive me." The answer "yes" is an instance of formal forgiveness, rather than an expression of an internal state. The formal type of forgiveness is less studied philosophically than its psychological counterparts. But, I want to urge, it is equally worthy of philosophical attention, and it is the one I will mainly discuss here (with the caveat that it relates to the other, in ways I will show).

It is the formal type of forgiveness that seems surprisingly capable, at times, of altering the requirements of moral repair, despite their objective basis. When we forgive someone formally, we can relieve her of at least *some* duties to make amends. Two such duties, in particular, stand out: First, wrongdoers, all else equal, have a duty to try to repair the harms they wrongfully impose on others, be it damage to their bodies, their reputations or their sense of security or comfort or trust, say. Call this duty that of "restitution." Second, as I already mentioned, there is the duty to apologize, especially when victim and offender stand in some relation to one another. Even if the victim knows that the offender feels bad and the offender

[54] Griswold 2007, 40-42.
[55] Richards 1988.
[56] Griswold 2007, 43.
[57] Austin 1962, 45.
[58] Griswold, 2007, 3.

knows this, he has a distinct obligation to offer an apology to the victim. Yet both of these duties – restitution and apology – can be relieved at times by a victim's forgiveness. A victim may decide that the cost of restitution is too high for the offender to make good on it in time to continue having a relationship, or that an understandable, if unjustified, fear of the victim might make apologizing too difficult. For these and other reasons, a victim may decide to forgive her offender in advance of restitution and apology. And her doing so can, at least sometimes, relieve him of both duties.

Formal forgiveness, so described, can vary in its scope and impact. In some instances, it can relieve the wrongdoer of being held responsible in any way for what he did, at least inasmuch as it relieves him of any outstanding duties to make amends for his wrongful behavior. Consider: "I completely forgive you – it's as though it never happened." In other instances, by contrast, formal forgiveness may be used in a narrower or partial sense. For example, a victim could formally forgive only the debt of restitution, but not that of apology, and vice versa. The power and scope of the act depends, at least in part, on the discretion of the victim.

Still, that power is considerable: a victim seems capable, at times, of forgiving debts of moral repair no matter what the offender had previously done or not done. There may be much to criticize about forgiveness granted in this way "for nothing," so to speak. But my point is only that acts of forgiveness, justified or not, sometimes *can* relieve offenders of these duties. A most dramatic example is the case of South Africa, whose black majority formally sought to forgive their white oppressors, who had barely begun to compensate their victims for decades of injustices and atrocities.[59] On a less grand scale, Reginald Denny – a victim of a violent assault

[59] See, for example, Jeremy Harding, "Picking Up the Pieces," *The New York Times* (May 30, 1999), via web at http://www.nytimes.com/1999/05/30/books/picking-up-the-pieces.html?pagewanted=all&src=pm; Suzanne Daley, "The World: Reconciling in South Africa," *The New York Times* (November 8, 1998), via web at

during the L.A. riots of 1992 – publicly forgave his assailant, before the latter even showed any willingness to apologize, much less compensate him for his injuries.[60]

The power of formal forgiveness in these cases is rendered more dramatic by its seeming irrelevance to the reasons that restitution and apology are due, in the first place. On standard accounts of corrective justice, for example, an offender should bear the burdens he wrongfully imposed on someone else.[61] The cost of having been wronged should, in other words, more rightly be borne by the person who wrongfully imposed it. As for apology, on the arguments of previous chapters, it is owed because of how a wrongdoer would otherwise treat victims if he failed to apologize (or do something morally equivalent). It is, however, unclear how forgiveness affects these reasons. A victim's formal forgiveness of his offender does not change the injustice of her bearing the cost of the wrong done to her, nor does it seem to change the way an offender treats his victims in the absence of an apology. And yet, if the victim decides to forgive his debt of restitution, then arguably he no longer owes it for the harms he inflicted – in fact, he need not even offer any recompense. In the same way, a victim can relieve him of any duty to apologize. A *victim's* forgiveness can thereby alter the moral status of the *perpetrator*: it can change what duties of moral redress he owes. As a result of a victim's forgiveness, we might say, a perpetrator may become no longer morally delinquent, no longer morally in arrears.

This is not merely to say a forgiven offender no longer has to compensate or apologize to his victim. That result would be compatible with the offender remaining blameworthy for failing

---

http://www.nytimes.com/1998/11/08/weekinreview/the-world-reconciling-in-south-africa-next-up-for-amnesty-the-unrepentant.html. See, also, Herman 2007, 319-22.

[60] Denny did not appear to forgive the Los Angeles police force, however, for having left the neighborhood of his assault insufficiently patrolled at the time; he sued the city of Los Angeles on those grounds. Kenneth B. Noble, "A Showman in the Courtroom, for Whom Race Is a Defining Issue," *The New York Times* (January 20, 1995) via web at http://www.nytimes.com/1995/01/20/us/a-showman-in-the-courtroom-for-whom-race-is-a-defining-issue.html?pagewanted=all&src=pm.

[61] Aristotle 1955, 144; Weinrib 1994, 277, 280.

to do so. A victim's death, for example, blocks or prevents these acts, even though the offender remains morally delinquent – his failure to apologize and make restitution remains morally blameworthy, or at least worthy of moral criticism. It is simply pitiable that he can do nothing about it. Forgiveness might be thought similar – blocking the ability to pay moral debts, so to speak, rather than relieving them. For example, the victim may have some overriding right to be left alone if she would rather not encounter her wrongdoer, even while apology and restitution remain criticizably absent.[62] Call this the *prevention* model of how acts like forgiveness remove moral duties. In contrast, there is also a *relief* model: on this picture, some act or event removes moral duties by removing any moral reasons they are due. For example, if I release you from a promise you made me, your duty to keep it is not only absent; you become above criticism for not keeping it. So it bears emphasis: that is the sense of forgiveness I will use here. Put differently, I mean the more dramatic claim that forgiveness can actually render the offender no longer morally blameworthy, or even subject to legitimate criticism, for failing to apologize and make restitution. She is no longer morally delinquent - and that is not because of any indifference or withdrawal on the part of either party with respect to their relationship. Again, the formal act of forgiveness changes not only the practical options available to the offender, but his moral status, as well.[63]

In this chapter I will try to account for how forgiveness has this power. How, exactly, can a victim alter the moral status of her offender by forgiving him? I will provide two different, but related, answers – one regarding restitution and one on apology. The first answer relies on

[62] For a version of this view, see Maimonides 1987. A common interpretation of Maimonides is that the victim has forfeited his right to the offender's moral repair once he ignores a sufficient number of repeated apologies and pleas for forgiveness. But the text does not support this interpretation, as it simply instructs, after enough attempts at reconciliation, "leave him be." This is compatible with further redress continuing to be due, but with the victim having a right to be left alone.

[63] David Owens takes what I think is an extreme version of the view, arguing that forgiveness can actually remove any reason to blame or criticize the offender for the *initial wrong*. Owens 2012, 56.

appreciating that restitution essentially involves bestowing something to the victim that has value to her, and to which she has a right, including the right to transfer it back to the offender. This understanding of restitution, I hope to show, explains how it can be relieved by a victim's forgiveness, without any sense that the offender is morally delinquent. As for apology, I will try to show that forgiveness, despite appearances, is not a simple unilateral act, but is intelligible only in a context in which a primary moral function of apology is already being performed.

I will begin, however, by explaining why I reject an understanding of forgiveness that either downplays its power as I have described it, or casts doubt on whether it can ever be reasonably exercised. Most important, in Section III, I will introduce and argue for a positive account of the power of forgiveness to relieve the duties of apology and restitution. First, I will explain why restitution, like certain other components of moral repair, is in fact a *directed duty*[64] -- a concept I will also clarify in greater detail. I will then spell out why this feature of the duty of restitution accounts for the power of forgiveness to release the offender from it. Next I will explain how forgiveness can make it unnecessary to apologize. This will require a brief recollection of the material in Chapter I, on the moral function of apology. I will try to show how that function can be performed, even without the formal apologetic speech act, when a victim forgives her offender and the offender accepts her forgiveness.

Finally, I will briefly consider and try to answer a question the explanations here naturally invite: even if forgiveness can relieve certain duties of moral repair, why should we ever want it to do so? Forgiveness, after all, occurs just as commonly *after* appropriate moral redress. When that happens, we also have the positive effects of shifting the cost of wrongful harms to their wrongdoers and of a meaningful and sincere apology. These steps have valuable

[64] Gilbert 2004, 83-4.

moral functions, as I have tried to show elsewhere. Why should we not wish, or at least prefer, that forgiveness occur after these steps have been taken, rather than preempting them? Why ever celebrate forgiveness that comes, as it were, too early? I will propose an answer that makes sense of the common practice of admiring, even romanticizing, forgiveness that precedes rather than follows restitution and apology.

I. Preliminary clarification: what is formal forgiveness?

This chapter, again, is not about any mental state that might be called forgiveness. It is, rather, about what I will call "formal forgiveness." To recap: formal forgiveness is a speech act whereby a victim of wrongdoing both enacts and expresses her decision to no longer hold her offender responsible for what he did. Holding him responsible, again, includes acting as though he is morally responsible – resenting, punishing, blaming and rebuking him – and calling on him to make up for what he did. Although it differs from a psychological state, formal forgiveness does have conditions of success and sincerity, which may implicate the forgiver's psychology after all.

First, formal forgiveness is an instance of what I have been calling a "stance-taking." In this it involves an expressed commitment to take and maintain a stance, which is a way to act towards someone in light of a normative position one adopts. In this case, the position is one that is expressed or implied by the words "I hereby forgive you," such as, perhaps, that the speaker should no longer be resented, blamed, punished, or called upon to make amends for the wrong that he did. To sincerely forgive, then, requires both being disposed to treat the offender as, in fact, not to be resented, punished and otherwise held accountable, say; and second, to try to

maintain the stance enacted by the first disposition – to *remain* disposed to treat the offender that way. Once forgiveness is granted, then, the forgiven offender can now hold the speaker to the stance. As a result, sincerity requires not only being in the stance taken, at the time when one formally forgives, but also becoming disposed to try to maintain the stance.

These sincerity conditions, of course, differ from the psychological states typically one might typically associate with forgiveness – an end to bitterness, feelings of hostility and resentment, say. But on closer scrutiny, they do coincide both with having and resisting intense affective states, as we saw with the case of apologies.[65] If I internally resolve to treat someone in line with having formally forgiven him, and to maintain my disposition to treat him that way, I will likely find it necessary to resist states that undermine this project – such as hostility, bitterness and resentment. Indeed, resisting those states will not only *foster* my commitment; doing so is itself an act consistent with the stance I have taken. Resisting or shunning hostile feeling is a way of acting by the normative position that my wrongdoer is, say, no longer to be blamed or resented; it is to act by the stance. So it will be highly probable, perhaps even necessary, that certain psychological states will naturally arise in the person who has sincerely forgiven another in the formal, stance-taking sense.

At the same time, affective psychological states are not sincerity conditions for formal forgiveness. Indeed, it would be excessively demanding to suppose otherwise, as forgiveness is a one-time act that is supposed to last. Once an offense is forgiven, one is presumed to have maintained forgiveness from that point on, on pain of insincerity. If sincere forgiveness required affective psychological states or their absence – say, no more feelings of hostility or resentment – that would amount to the absurd demand that a forgiver never experience bouts of these

[65] See, *supra*, Chapter I.

sentiments for periods that could last years, and that she can competently commit to this in advance. Instead, I propose that formal forgiveness has the sincerity conditions of other stance-takings, namely a disposition to act by the stance and an internal commitment to maintaining that disposition. While these naturally coincide with psychological states and with the absence of others, they do not include such conditions as requirements of sincerity.

Another important feature of formal forgiveness, whereby it differs from any private state, is that it is *bilateral*. It is among the types of stance-takings, like apologies and thanks, that put in place a way of relating *to* someone. Just as we apologize *to* someone through a speech act directed at the recipient, so too we express formal forgiveness to its recipient. As with other bilateral, commissive speech acts – consent, promises – the listener has to both apprehend and appreciate the meaning of forgiveness for the speech act to be successfully performed. In this sense, then, formal forgiveness requires acceptance or, as Judith Thomson showed with respect to promises, it requires uptake, to be successfully performed.[66]

With this picture of the constraints on the speech act of forgiveness, the problem that motivates the chapter gains sharper focus: why should this formal speech act, so described, have the power to render apology or restitution no longer required or even criticizably absent? At most, it can involve – especially if specified as involving – the forgiver's commitment not to *treat* the offender as needing to apologize and make restitution. Indeed, the forgiver may even commit to try to accept and be content with his offender's lack of repentance. But the formal speech act is done immediately, long before the speaker can begin to cultivate the psychological states that may be relevant to the impact of the offender's apologizing. And in any event, as argued earlier, apologies and restitution are owed for reasons having little to do with any

[66] Thomson 1992, 296-97.

psychological state of the victim. It may be that a victim can *block* the offender's *right* to apologize and make restitution, much the way someone who holds a grudge may tell someone, "If you don't make up in the next two days, that's it – I never want to hear from you again. Don't even bother." But that would hardly explain how, through forgiveness, she can make those acts no longer problematically absent. How, in other words, can forgiveness render the offender not only free from a *duty* of apology and restitution, but no longer morally criticizable for failing to make amends in these ways?

II. Taking forgiveness seriously

One candidate answer to the chapter's principal problem as it was just now restated – how forgiveness relieves certain duties of moral repair – is skepticism; there is no problem, perhaps, because forgiveness does not have the power to relieve those duties, after all. On this possibility, forgiveness would be defective, and so ineffectual, when offered to someone so unrepentant as to take no steps to make up for what he did.

This proposal should not be caricatured; it is *not* the claim that forgiveness would be insane or ill-motivated in the absence of apology or restitution. In fact, I proceed from the observation that victims do have reason to *want* to forgive even unrepentant offenders. One reason is that they may seek reconciliation, and they may worry that continuing to hold their offenders to duties of moral repair will serve only to prolong and deepen the conflict. Withholding forgiveness, however justified, could force the offenders to remain defensive, and thereby regard the victim negatively as a continuing accuser. Alternatively, even if a victim believes reconciliation depends on the offender appreciating what he did wrong, she may regard forgiveness as a more reliable route to enabling that appreciation.

In addition, she may resent the offender for failing to apologize and offer restitution, and she may wish to assuage her resentment by relieving him of those duties. Resenting the offender is, after all, a harmful state, over and above the harm wrought by the initial wrongful act. The second-order appreciation of its wrongfulness gives rise to what Pamela Hieronymi has called "moral pain," a sense of having been wronged that is experienced badly.[67] One reason, in particular, that resentment is so harmful is that it is a reasonable response to ongoing behavior, not just the initial wrong. As I argued in Chapter I, when an offender wrongs someone and fails to do anything to redress what he's done, he treats the victim as though he is free to wrong her as he did. This treatment, or mistreatment, gets worse with time: the longer the wrongdoing is allowed to stand without redress, which is to say the more opportunities for redress he spurns, the more definitively the offender's actions amount to an acceptance of his misdeed. The victim, if she has the same moral commitments to which she would hold the offender, will disagree with this treatment: she, of course, regards the wrongful act as decidedly unacceptable, something he's *not* free to do. Appreciating this fact forces her to revisit and reaffirm the wrongfulness of what was done to her. And that will normally be experienced unpleasantly. In short, her own moral commitments, applied to her circumstances, invite a reasonably unpleasant response on her part to the offender's failure to redress what he has done.

For these reasons, then, it would benefit the victim to be able to forgive the offender and thereby relieve him of any duty to make amends. Nevertheless, perhaps victims *cannot* properly forgive in lieu of apology and restitution, because formal forgiveness is defective, and so powerless, unless the offender has *already* made amends.[68] In particular, formal forgiveness is arguably appropriate *only* if an offender apologizes and at least *offers* to make restitution. That,

[67] Hieronymi, in discussion.
[68] Hampton 1998; Murphy 1988; Hieronymi 2001.

of course, would render the present account – an attempt to explain the power of forgiveness to relieve those duties of moral repair – hopeless.

Two points are therefore necessary in response. First, the claim that formal forgiveness should only be granted once all reasons for resentment have been removed dramatically alters our pre-theoretical understanding. In particular, it implies that the scope of the forgivable becomes much, much narrower than we would ordinarily suppose. Both examples in the first paragraph – South Africa and Reginald Denny – would be inappropriate cases for forgiveness, since in both cases significant restitution is due, especially in the case of South Africa, and in the L.A. riots case there is no apology or even expressed regret. In addition, this understanding of forgiveness – as something that must be earned – robs it of many familiar features, not least of which is the value we ordinarily place on it. Forgiveness, we might recall, is deeply and intensely sought, and sometimes thought to inspire, rather than simply follow, repentance. It is also cherished as a gift, an act of generosity, as though a victim improbably, even romantically, overcomes her justified resentment and the demands of justice. Yet on the version just raised, in which forgiveness must be morally earned, it should be none of these things. In any event, it ceases to be recognizable as forgiveness the way it is commonly understood and treated. As Meir Dan-Cohen points out in an essay on this subject: "forgiveness in the absence of repentance seems possible and indeed particularly noble. A philosopher is surely not in a position to legislate it out of existence."[69]

Second, and more importantly, I will attempt to show why the act of forgiveness itself works to remove at least some reasons to resent the offender. Indeed, it helps to enable the offender to fulfill much of the essential functions of both restitution and apology. For that reason,

[69] Dan Cohen 2007, 120.

once the victim forgives the offender, she no longer has the same reasons to resent him, at least to the extent that they involve his failure to make amends. Therefore, even on the view that forgiveness must be warranted by reasons to stop resenting the offender, the formal act of forgiveness may nevertheless be appropriate in advance of moral repair. In contrast to traditional views, then, the formal act of forgiveness may be not so much the *consequence* as the *cause* of ceasing to resent one's wrongdoer. The next section attempts to explain how this could be.

III: The "give" in forgiveness

A. Restitution

My basic claim is that restitution involves bestowing something to a victim that is of some value to her, and that her claim right to the bestowal includes a right to transfer its value back to the offender. Her right to restitution, in other words, includes a right to give it back, as well. That, in effect, is what she does when she forgives her offender, as I will try to demonstrate.

A useful model is the idea of a debt. It is uncontroversial that someone who is owed money by someone else has the power to "forgive" the debt, where that means simply relieving the debtor of the duty to pay. This is so even though, all else equal, it is morally obligatory to pay one's debts. The reasons to do so do not, at first blush, appear to depend at all on the recipient's wishes or values. Taking someone's money and not returning it is simply wrong, for reasons that can be cashed out in terms of a variety of impartial moral considerations. In a manner similar to our original question, then, we could ask: why does the lender have the power to relieve her debtor, if the reasons to pay the debt to not relate to how the lender feels about it?

Yet the question seems patently absurd in this context: the debt, we might say, is *hers* to relinquish. If she does not value it enough to want to hold it against her debtor, it seems inappropriate to seek to impose it on her anyway. Put differently, money and other goods – indeed, anything we might give to someone because they value it – are the sorts of things one is said to have, once bestowed. It is a function of a debt belonging to the lender, then, that she can take it or leave it, own it or relinquish it, as she sees fit. In particular, she is free to transfer it back to her debtor. Her doing so does not, in the end, trump the considerations of justice and property rights, or whatever, that require the debtor pay her. They simply follow from the fact that her right to own the money, once he pays her, includes a right to give it away, and she can do this in advance of payment, by forgiving the debt.

Contrast this with other types of duties, such as the duty to report a health hazard to potential victims. Here, too, there is arguably a duty to bestow a benefit on someone for moral considerations. But we do not imagine this duty being relieved if the recipients sincerely renounce their desire to know about dangers they face. The correct reply might be: who are you to give me such permission? I have the duty anyway.

The difference between the health hazard reporter, on the one hand, and the debt holder, on the other, is *not* that third-personal moral considerations justify one but not the other's receiving a benefit; *both* are justified by considerations of impartial morality. The difference is that, in a perfectly literal sense, the benefit to the lender belongs to her. It is something whose value to her she can freely determine – as with all possessions. As a result, she can decide she'd prefer that someone else have it. It is as though she takes it and returns it as a gift. Duties to bestow some good that is the recipient's right to claim or relinquish are sometimes referred to as

"directed duties," in the sense that they are duties to the recipient, rather than to society generally.[70] It is, accordingly, the recipient's right to return, refuse or transfer the benefit.

Moral repair, or at least restitution for harms wrongfully inflicted, is often described in terms of debt. We *owe* compensation to the victims of our negligence, for their pain, suffering, and damages to what they value. I am proposing that this is not merely a useful metaphor. Restitution is, rather, a directed duty – it is a duty to bestow something of ostensible value to the victim of wrongdoing, which is then hers to possess. And, I am arguing, the recipient's right to this value includes the right to transfer the same value to her offender, including by way of forgiving the debt of restitution.

This claim does not, however, follow straightforwardly from the comparison with debts. There are, in particular, two important differences between debts in general, and debts of restitution, in particular. The first is that a debt is usually for money, so once the lender receives money by way of paying that debt, he has already benefited in precisely the way the debtor intended. More importantly, the *form* of his benefit (his having the money to spend as he'd like) reflects the *moral purpose* of bestowing it to him (that he no longer be deprived of it by the debtor). In contrast, restitution seems to have a narrower purpose: to restore something specific of which the victim has been deprived by the wrongdoing. Or, to use the language of corrective justice theory, it is to undo a loss wrongfully imposed on the victim, which is not necessarily a monetary loss.[71] The use of money, then, may seem merely incidental, a non-ideal substitute for actually repairing whatever damage was wrongfully done to the victim. Paying for medical bills, then, would be a non-ideal substitute for simply curing the victim of wrongful injury, healing her

[70] Gilbert 2004, 84. Gilbert uses promissory obligations as an example of a directed duty. One reason to think it has this property is that, as Seana Shiffrin has argued, a promise transfers the promisor's right to perform or not to the promisee. See Shiffrin 2008, 517.
[71] See, for example, Martin Stone, "The Significance of Doing and Suffering," in *Philosophy and the Law of Torts*.

on the spot. The primary point of restitution, then, would be the medical benefit. As a result, it appears there may be limits to what the victim may do with her restitutionary benefits, on pain of violating the moral purpose for which they are given. Restitution, in other words, may be – to use another metaphor – "earmarked" for repairing damage. It may, then, be wrong for the victim to direct it to any other purpose, such as transferring it back to the offender .

This worry, however, ignores an important difference between the present and the past, before the wrongdoing was committed. The victim today does not face the same set of circumstances and options as before she was wronged. Earlier, her wellbeing came at no cost to someone else; in contrast, today her receipt of her rightful recompense may very well come at a difficult cost to someone else, namely her injurer. Perhaps she is so sympathetic as to disvalue his loss more than the absence of her restitutionary benefits. Both losses, then, would constitute a negative consequence of her having been wronged. Both could be aptly described as losses that the original wrong imposes on her.

If restitution is aimed at redressing the loss imposed on her by a wrong, then, its cost to the offender could qualify, as well. And for that reason, I am proposing, she can reasonably direct restitution *either* towards repairing the damage suffered or by transferring it back to the offender to spare *his* loss. The victim is, either way, freely using the value rightfully due her to undo losses incurred as a result of having been wronged.

Forgiveness is a mechanism for effecting this choice. Recall (from Section I) that the stance that forgiveness takes can include acting for the principle that the offender is no longer subject to the demand of restitution. By taking that stance, then, the victim withdraws her claim or demand for restitution from the offender. It is not controversial that we can withdraw claims

through speech.[72] What I set out to explain, rather, was the *moral* power of the victim to do so in a way that relieves the offender of the duty to make restitution; indeed, releases him even from legitimate criticism for not providing it. The answer, as I have now tried to show, is that the victim can relieve him of the duty because she is, in effect, fulfilling it for him; she is using the value of her restitutionary goods to undo a further loss which just happens also to be *his* further loss.

In short, in performing the formal act of forgiveness, a victim can use his power to return the value of restitution to the offender, by exercising it in advance of restitution – and directing, in effect, that the offender keep it. He can do this because, like retaining the restitutionary benefit, such an act of formal forgiveness involves using the value of his rightful recompense to undo a loss he has incurred as a result of having been wronged, namely the further suffering of the offender. Although the damage he himself has suffered is a loss, as well, it cannot be separated from the fact that correcting it ordinarily requires a loss to the offender, too, which the victim may disvalue even more than his own loss.

One may worry that it is only the exceptional victim who would experience the offender's payment of his restitutionary debt as her own loss. But recall that this paper is about exceptional victims to begin with – those who wish to forgive in advance of restitution and apology. It is likely that the two classes perfectly overlap: those who would forgive offenders the debt of restitution are, at the same time, those victims for whom the offender's payment would be a further loss. It is the sympathy reflected in this valuation that likely accounts for the poignancy, indeed the admiration, evoked by such acts of forgiveness.

[72] Although, in the context of the dissertation, it bears mention that coming to value something by commissively declaring that one does – through making or withdrawing claims – is perhaps the paradigmatic case of stance-taking.

This characterization of the power of forgiveness to relieve the debt of restitution should be qualified, however. First, it does not apply to all acts of forgiveness or forgiveness simpliciter. The phrase "I forgive you," even as a formal speech act, is ambiguous. A victim may mean it only partially, and wish it to communicate only matters unrelated to restitution. For example, she may wish to formally forgive her offender in the sense of rendering him no longer subject to blame, resentment or sanction for what he did, while not forgiving his debt of restitution. This possibility must be granted. All I want to argue is that a victim can, through an act of will, forgive her offender's debt of restitution and that her doing so does not merely license him refraining from restitutionary action, but renders him no longer morally criticizable for failing do to so. It changes *his* moral status in this respect.

Second, a victim's ability to determine which loss (hers or her offender's counterfactual loss) she wants to undo is not unlimited. She may, in particular, be unqualified to determine that she disvalues restitution in cases where it is of such enormous value that she cannot reasonably decline it. For example, if an arsonist rendered a victim homeless, and the trauma so demoralized the victim that he became unwilling to assert his right to a home, the arsonist may arguably owe him restitution nonetheless. His need for shelter may be so compelling that it overrides any decision on his part to disvalue it. Indeed, victims of ongoing oppression or abuse may become unreliable, as a result, in determining which loss, if any, they should correct through their right to restitution.

Third, as part of the previous point, forgiveness may not have the power to prohibit restitution altogether, or even to render it less desirable. Recall that the victim may disvalue

restitution in large part because it involves a cost to the offender. That cost, however, may be a matter for the offender to determine, just as the value of restitution – the gain or the loss undone by it – was for the victim to decide, at least in many cases. The offender may value giving restitution far more than retaining it, no matter what the victim wants. In that case, it would be a cost to the offender *not* to pay his restitutionary debt and, for the same reason, it would not be a cost to lose it, despite victim's impression to the contrary. In the case of the destitute debtor, of course, the lender's decision to return the money, or – which is functionally equivalent – to forgive the debt, seems well-informed. Repayment is a burden to the debtor that the lender deems a cost to him, too. In contrast, consider a case of a wealthy electrician who negligently caused her neighbor's house fire and cannot bear the sight – or even the continued existence – of the burnt property. She feels too guilty and ashamed, not to mention it is bad for business. Thus even when her neighbor declines to sue her and forgives her debt of restitution, she may deem non-payment a cost to herself; indeed it may be too high a cost at that. The victim would then be mistaken in disvaluing restitution, to the extent that she does so in light of the cost restitution imposes on the offender. This is one of the reasons I emphasize here that forgiveness *can* relieve restitutionary debts, at least sometimes. But it need not block an offender from paying restitution anyway.

B. Apology

The power to forgive restitution is perhaps the easier to explain, because of its obvious affinities with monetary debt. Apology, in contrast, has no such affinity. There is nothing it bestows upon the victim whose value to her she ought to be able to transfer back. True, we speak

of "owing" an apology, but it is not uncommon to say that we owe all moral duties to their rightful recipients, regardless of whether they are directed duties. While it is offered as something to be accepted or rejected, apologizing is not something whose value is to be determined by the victim. As already stated, the reason to apologize stems from how we would otherwise treat victims – whatever they think or feel about it – in the absence of a step like apology. How, then, can forgiveness relieve duties to apologize, rendering the act no longer morally required? The answer builds on the explanation just now offered for how forgiveness relieve restitutionary duties; indeed, it depends on the account in the previous section, as I will now show.

First, recall that the power of forgiveness to relieve moral debts attaches only to the formal, public type of forgiveness: the speech act of absolving someone of wrongdoing. If I mean to forgive someone but fail to communicate it to her, she arguably still has to compensate me for any damage she wrongfully inflicted and to apologize, as well. It is only if I forgive her formally, if we go through the process of performing the speech act, that she may refrain from apologizing and compensating me for damages.

What is so special about the speech act of forgiveness? Recall from Section I that, as a commissive stance-taking, it involves a commitment to the intended audience to act by its lights.[73] By declaring "I hereby forgive you," the victim commits, to the offender, to treat her as someone no longer worthy of resentment, blame or punishment, or a demand for apology or restitution. And as already discussed in Section I, forgiveness in this sense is not purely unilateral: it puts in place a relationship with its audience, as a commitment *to* the latter. As such,

[73] Austin 1962, 151.

successful performance requires uptake, or acceptance.[74] Forgiveness is not simply uttered to the wall; it is offered, or committed, to the offender who – if it is a successfully performed speech act – *accepts* it.

To see the force of the uptake requirement, consider a recalcitrant offender who rejects the victim's power to claim or relieve others of moral repair. Imagine that the victim offers forgiveness and the offender replies, "Who cares?" or "Who are you to forgive me? Take your forgiveness and stuff it!" In this case, I expect intuitions will confirm that the offender is *not* relieved of his duties of repair. He must still apologize to the victim. If so, then an important component of the power of forgiveness to relieve the duty of apology is acceptance. The offender must accept the victim's forgiveness.

Why does accepted forgiveness relieve the debt of an apology? I want to propose that acceptance of forgiveness functions morally like an apology in at least one important respect. As I argued in Chapter I, apologizing is a way of fulfilling a more basic moral requirement: wrongdoers should avoid treating their victims as people it is acceptable to wrong as they did. That treatment is constituted by wronging someone and continuing with business as usual. It is also constituted by wronging someone, paying restitution, and moving on, because this set of behaviors treats the victims as people we're free to wrong for the price of restitution; to wrong and then pay off, as it were.

Apologizing, on the other hand, counteracts this mistreatment, as I explained in Chapter I. The primary way it does this is by relating to the victim as one whom the offender owes an unpayable debt. It is not payable because anything the offender might do would amount to

[74] Thomson 1992.

treating the victim as wrongable for a price. Indeed, any subsequent action meant to make up for the wrong, by itself, would have this problematic character of a payoff. To avoid this mistreatment, then, one must come up with a way of seeking to make up for the wrong – and so not moving on – but in such a manner that it is presented as insufficient. Apology was one way; I now want to propose that seeking or accepting forgiveness is another. Accepting forgiveness treats the victim as one with the power to forgive the offense – for that is what the recipient is accepting and acknowledging. This acceptance would have no function if forgiveness had no impact, or was unnecessary. But forgiveness plays an essential role in moral repair because, absent the victim's forgiveness, there is no way to relieve one's debt to the victim over what was done. Mere restitution is not enough, because it treats the victim as wrongable for a price. Forgiveness is the only way to relieve an unpayable debt. Seeking it is a way of acting on the insufficiency of any action or restitution by itself.

Seeking or accepting the victim's forgiveness, then, reflects the insufficiency of what the offender himself can do; it places the power of moral repair squarely in the hands of the victim, instead. In that way, it treats the victim as one whom the offender owes not to have wronged as he did, and for whom no action or restitution can be sufficient. This way of treating the victim, then, ends the mistreatment – as one who may be wronged in that way – that apologies counteract as well. In this way, seeking and accepting forgiveness treats the victim in just the morally reparatory way that apologies do.

That said, the power of forgiveness to relieve the duty to apologize should not be overstated. In particular, it should not be read to preclude subsequent apologies. Accepting forgiveness, as I have described it, shares a single, albeit important, function with apologizing. But there is much more to apologizing, particularly in the case of very good or heartfelt

apologies. An apology can help the victim continue to fight or assuage any resentment that bubbles up, by persuading her of the offender's repentance or regret. It can also improve their relationship, inasmuch as it allows the apologizer to explicitly acknowledge, and thereby recommit himself to, a level of respect he owes the victim that he never considered before. An apology also allows the offender to make a better case for why the victim should continue to trust and be close to him, beyond simply forgiving the wrong.

IV. The limits of formal forgiveness

Having defended the thesis that forgiveness can both relieve wrongdoers of the duty to apologize and make restitution, and remove any blameworthiness for failing to do so, I should state some of the more obvious limits of this power. First, like forgiveness, apology and other speech acts, the acceptance of forgiveness is not a dichotomous act, either done or not done. Rather, it can be performed with varying levels of awareness, richness and depth. Recall that, on the account sketched so far, accepting forgiveness is a way of treating the victim as one we are not otherwise free to wrong as we did. This respect, however, can exist to greater and lesser extents. Arguably, it is incompatible with failing to actually regard the offender as owed the requisite level of moral repair. Put differently, a necessary condition of respecting someone as P is not disbelieving the description of her as P. For example, one who negligently injured someone, and believes that his monetary restitution is enough to obligate the victim to forgive him, fails to respect the victim as one who is owed more.

There will, therefore, be cases in which an offender simply does not believe, or sufficiently appreciate, the extent of the victim's entitlement to redress. In these cases, perhaps,

accepting forgiveness is defective or infelicitous, in much the way an apology by someone who does not believe he did anything wrong is infelicitous. Offenders who were slow to appreciate the extent of their wrongdoings or obligations to the victim, though they may have been forgiven, should therefore apologize anyway.[75] That is not to say the victim should not have extended his formal act of forgiveness. As an instrumental matter, such acts can at times spur the very respect the offender could not properly summon at the time he seemed to accept forgiveness. It is, indeed, plausible that in a great many cases offenders will not be able to reflect freely and remorsefully on what they did until and unless the victim forgives them first.

Similarly, just as accepting forgiveness can be incomplete, so can forgiveness itself. As noted in the previous section, one may forgive an offender for what he did, but not mean this forgiveness to cover the debt of restitution. Reginald Denny's poignant and public act of forgiveness, for example, did not withdraw his demand for compensation – at least from the city government that presided over the events culminating in his injury. On the account proposed here, forgiveness relieves a debt of restitution by way of the victim foregoing his claim right to it and acting on his ability to return it to his wrongdoer. Needless to say, if the victim does not willingly exercise this right, then restitution remains due him.

Further, the account so far shows only how forgiveness can have the power to relieve offenders of various duties. I argued earlier that the power to relieve the duty to apologize and pay restitution is distinct from the power to prevent them from being sensibly performed, such as when a victim dies or makes himself unavailable to his offenders. In those types of cases, while it may not be possible to apologize and make restitution, one remains morally delinquent or

[75] On the problems with premature apology, see Shiffrin 2002, 38-39.

blameworthy for not having done so. Or, put differently, there is still valuable redress and moral repair to be done. As I said earlier, forgiveness is prevented, but the duty is not relieved.

V. What's so great about early forgiveness?

I have sought to explain how forgiveness can remove the offender's duty of restitution and apology. Even when that happens, I have argued, the key moral functions of restitution and apology remain in place: the victim is respected as one with a claim right to restitution, and as one whom it is unacceptable to wrong as one did, and who is therefore owed more than the offender can offer. But, by hypothesis, those same functions are performed when restitution and apology do occur, as well. If forgiveness is attained in those cases, too, why should we not favor them over cases in which forgiveness relieves the other duties?

In some cases, it seems uncontroversial that we should favor forgiveness that comes in response to attempts at moral repair. When children wrong one another, it may be important for them to learn the standard forms of moral repair, including restitution and apology. They may also be less capable of appreciating the extent to which these acts may be functionally replicated even in early, preemptive forgiveness. Similarly, as already discussed, some offenders may take a while to appreciate the full extent of his wrongdoing.[76] Only upon appreciating why he owes moral repair, including apology and restitution, will he engage in these steps. Preemptive forgiveness, however, would block him from reaching that stage of full moral awareness and repentance.

Finally, there are cases in which the injustice is simply too great to defer to the victim's willingness to forgo restitution. Ordinarily, of course, we defer to people's own decisions about

[76] See Shiffrin 2002, 38-39.

what they would like to do with their goods and benefits. Worries about paternalism caution against overriding someone's own preferences about what is valuable to her, as if suggesting to her: "you're wrong not to want more; you're wrong to be satisfied." But there are times when something's value to someone can be a matter of objective fact: when an impoverished person is wrongfully deprived of the use of his arm, he will simply be unable to get by without some restitution, especially if his employability depends on the use of his arm.

In other cases, however, we do have reason to respect the judgment of both the victim, as to whether restitution would be as valuable to him as giving it to the offender, and of the offender, as to whether, upon accepting forgiveness, she appreciates what she did wrong and what sort of treatment she owes the victim. For example if a doctor loses his temper with an incompetent resident during an important operation and, as a result, negligently injures her patient, she could very plausibly appreciate everything she's done wrong, including why she owes an apology and the way it commits her to treating the recovering patient. Similarly, the patient could decide that he is willing to bear the cost of a few weeks' rest until he recovers. He may decide that he does not value any further restitution, at least not enough to be inclined to claim it of the physician. There is no reason not to defer to his choice.

In these cases, there is value to early or "preemptive" forgiveness not shared by the type that follows apology. First, it is evidence of greater inclination to reconcile, which – of course – is a morally welcome outcome. Although I have assumed in section I that reconciliation is not the goal of moral repair, it is definitely a welcome development. A better relationship between victim and offender is something we desire for moral reasons. And the decision not only to absolve the offender but to grant him a gift – the gift of retaining the value of his restitution – evinces and promotes a more conciliatory approach.

Second, forgiveness in advance of apology and restitution is itself an act of generosity. The victim does not, at that point, have any special reason to forgive the offender – nothing that makes forgiveness owed or warranted. The decision to absolve him anyway amounts to a gift, a benevolent act that enables the offender to rightfully retain what is otherwise owed to someone else. This benefit has a corollary: it inspires a more conciliatory approach on the part of the offender, as well. Having been spared the cost of full restitution, he now has greater reason to be positively inclined toward the victim and to trust that reconciliation will be achieved.

Finally, it may be preferable that moral repair be accomplished with less suffering or harm on the part of the offender, all else equal (including the victim's harm). Early forgiveness clearly spares the offender the cost and harms associated with restitution. We might worry that this relief, on the wrongdoer's part, amounts to the victim bearing the harm, instead. But that would amount to not taking her act of forgiveness seriously, or respecting her right to own and act freely with her rightful recompense.

As argued in Section II, forgiveness may relieve the standard duties of apologizing and making restitution, but it is not by negating or replacing those duties or the function they serve. Rather, it is by realizing them in a different form. In the case of restitution, for example, forgiveness does not correct or cancel the offender's duty to compensate his victim for the harms she wrongfully suffered. It is, rather, a result of the victim having the right to that compensation and, accordingly, the right to bestow it as a gift to his offender. The right remains in place, and exercised. Similarly, forgiveness does not relieve the offender of the duty to treat his victim as one whom he owed not to do what he did, and for which restitution is an insufficient substitute – the duty fulfilled by apology. Rather, he fulfills that duty by accepting forgiveness, the meaningfulness of which can be illustrated by scenarios in which a wrongdoer rejects

forgiveness. In such cases, he is not relieved of these duties; he is not thought to have fulfilled them. In contrast, when he accepts forgiveness, he respects the victim's rights and powers in just the way apology does.

I have urged, in other words, that the moral functions of apology and restitution are performed even when the victim forgives her offender before he's fulfilled these duties. As a result, little of moral value is lost in cases of early, preemptive forgiveness. But much is gained: the victim evinces a more conciliatory attitude, and she engages in an act of inspiring generosity. Therefore, in all but a fraction of cases, such as wrongs committed by children or those imposing severe physical loss, it is a morally welcome event when the victim forgives in advance of apology and restitution. That is, perhaps, why it is celebrated, even romanticized.

Conclusion

I set out to explain how forgiveness can relieve duties of restitution and apology. It may be noticed that, in a sense, I did something different. Rather than show how forgiveness preempts or relieves the duty to perform these acts, I argued that their essential functions are performed even in the context of forgiveness. In forgiving a debt of restitution, the victim simply uses her power to transfer restitution back to the offender, a power that is depends on the restitutionary benefit being rightfully hers in the first place. In forgiving the offender, in cases where there is uptake and the offender accepts, the victim is treated in much the way she is by a formal apology. The wrongdoer respects her as one he was not free to violate as he did.

I hope these explanations convince, at least, even at the cost of recasting the power of forgiveness as something less mysterious or magical. Indeed, on the arguments presented here,

early, preemptive forgiveness does not override the duties of restitution and apology. It simply enables the offender to perform them in a different way, one that may sometimes be a likelier and more direct route to reconciliation.

CHAPTER IV:

INSTITUTIONAL STANCES

The prior chapters sought to explain how private speech acts between people can make moral differences and resolve conflicts. Some of the most enduring and costly social conflicts, however, occur between groups, nations, and institutions, rather than mere individuals. And to resolve them, some of the grandest and most public acts of reconciliation may be required: state and corporate apologies, group forgiveness, international agreements and brokered peacemaking, for example. These acts, however, are not mere acts: they can be sincere or not, and ordinarily it matters a great deal which one. When an apology proves insincere, for example, that may be grounds to withhold forgiveness and trust and even to resent the apologizer more. In contrast, an apology that proves sincere can ground historic reconciliation, even in the absence of further repair and compensation. The apology of the white supremacist former leadership of South Africa, formally offered by its ex-premier, was arguably perceived as sufficiently sincere to motivate the act of forgiveness that followed, and it paved the way for a more unified nation to emerge.

But sincerity at the institutional level poses a problem: how can non-personal bodies, such as countries, corporations and communities, be truthful about the positions they express, if all they can do is take formal steps or make public pronouncements? Public acts and statements are, after all, the very sorts of things about which sincerity can be questioned. Yet it seems those are the only steps institutions can take. For example, suppose that a U.S. restaurant chain is exposed as xenophobic, with menu items that insult immigrant groups, jingoistic themes in its

specials that exclude non-Americans, and a hiring practice that discriminates against immigrants and their children. After enormous protest, the restaurant chain seems to recant, issuing a public apology, changing its menu and promotional themes and beginning to hire members of the offended group. The question remains: is it sincerely or genuinely repentant about its past racist practices, or is it just going through the motions? Are these acts of contrition truthful? The trouble seems to be that, *as a restaurant chain*, there isn't anything more to consider besides such official steps: new corporate practices, public pronouncements, and the like. And these can either be sincere or not. Yet whether to count the corporation as having sufficiently repented turns, arguably, on whether its professed change of heart is sincere.

The problem reaches beyond acts of conflict resolution, such as apology, forgiveness and reconciliation, to the public commitments and speech acts of institutions generally. How can an institution be in a state of, say, gratitude (presumably a requirement for sincere thanks) or resentment (expressed by acts of protest), when these states seem ineliminably mental? How can a court or legislature be thought to have intended a certain meaning by their public rulings – as in the notion of "legislative intent"? The absence of any purely "internal" states of institutions, in the way that individual human beings have inner mental lives, seems to belie the common attribution of such states to them.

Here I will argue that institutions, be they corporations, countries or nations, can in fact be genuinely forgiving, apologetic, grateful and in similar states – which I take to be required by sincere performance of acts like formal forgiveness, apology, thanks and so on. I will propose the beginnings of an account of what such states consist in at the institutional level and why, despite

prominent claims to the contrary,[77] they are not reducible to the psychological states of any particular people involved, nor do they supervene on them in principle.

In the first part of the chapter, though, I will consider that possibility -- that institutional sincerity is a function of that of certain decisionmakers or other people involved in the institution. I will argue that this proposal fails largely because institutions can take positions shared by nobody in particular. In such cases there is, crudely, no person to whose sincerity the institutional sincerity can be traced. Unless we give up on these cases as candidates for sincerity, then, some other basis – besides individual people – must be found for institutional sincerity. In the second part, I argue that – nevertheless – institutions can have all that is essential for the states in question: forgiveness, repentance, good faith, trust, and the like. Firms, nations and communities can be grateful or forgiving, say, regardless of what any of their relevant decisionmakers feel or undergo. What is essential to such states – both for individuals and institutions -- is the ability to act on certain normative claims or reasons, often identified with "policies" at the institutional level, and the ability to maintain a commitment to so acting. Institutions, I will argue, can meet these requirements, and thereby be genuinely forgiving, grateful, trusting and the like, just as – in the absence of those states – they can be grudging, greedy and suspicious, along with a variety of similar states. Indeed, many normatively important states can be undergone by institutions – and this can be shown without appeals to anthropomorphism or stretching the meaning of everyday attributes.

In Part III, I will defend the account against a variety of objections. For example, I will respond to Cass Sunstein's argument that institutional actions or statements cannot have reasons

[77] See, for example, John Searle 2010, v: "[T]he only intentionality that can exist is in the heads of individuals. There is no collective intentionality beyond what is in the head of each member of the collective"; Kutz 2000, 7 for versions of the view that only individuals can have the intentional states necessary for moral evaluation.

– as my account requires they do – because the processes that produce those actions block the attribution of one particular motivation or reason to them.[78] I will also consider the special case of groups, such as minorities or displaced peoples, whose sincerity in speech seems – contra my account – best assessed in terms of that of its members. Whether "the Aborigines" oppose a reparation plan seems to be a matter of what various Aborigines feel about the issue, rather than what the singular institution known as the Aborigine people feel. In addressing this worry, I will distinguish between a variety of often-conflated ways to speak of groups of people, be they as collectives with pooled or shared intentions, collections of separate individuals, or singular bodies, and argue that the question of their sincerity and stances on some matter turns on which form of group-talk is at work. Finally, I will briefly explore what the account implies about the speech rights, if any, of non-personal institutions.

I. Internal states

To restate the problem in more detail: institutions (countries, corporations, communities) seem to make moral differences by acts of forgiveness, apologeticness and gratitude. These acts, as commonly understood in ordinary ethical contexts, are defective if insincere. In calling these speech acts "sincere" I mean, at a minimum, that the state to which they commit the speaker obtains over some period of time starting around the moment of the act's performance. The problem for institutional sincerity is that such states as forgiveness, gratitude and the like seem to be mental or at least internal in some sense. For example, as I argued with respect to apology, they involve normative commitments and positions one has adopted (such as, in the case of being apologetic or repentant, that one did wrong and owes the victim better).

[78] See, for example, Sunstein 1996, 23-36
.

Of course, one might press the problem further, taking these states to essentially involve feelings like guilt or regret, for example (with apologeticness, perhaps), or attempts to put feelings aside (such as feelings of bitterness, in the case of forgiveness). On that view, the notion of an institution being repentant or forgiving would be as absurd as that of its undergoing a burst of rage. But I have, in earlier chapters, argued against the claim that the states of interest here – apologeticness, forgiveness and repentance, for example – must involve feelings or emotional episodes in order to obtain. That may seem to open the way for institutional sincerity after all. So it bears emphasis that large obstacles remain. Even if states like forgiveness, for example, do not essentially require emotional states (or their absence), they do require that one adopt what appear to be normative beliefs or positions. For example, to forgive someone requires treating her as no longer worthy of resentment, punishment or blame – *for the reason that she isn't, in fact, worthy of such punitive responses anymore* (if only because the victim has decided to forgive her). In other words, a forgiver is guided by a reason or normative position he has adopted. Merely treating someone as forgivable is not enough; one has to *believe*, say, that she should not be blamed or punished anymore. Similarly, being repentant arguably requires adopting the reason or belief that one did wrongly and owes the victim better. The problem for institutions is that it seems impossible for them to have reasons, or normative positions they have adopted and internalized. Again, institutions lack internal brains, and internal states more broadly understood, that could be associated with accepting and being guided by reasons or their actions. They have only the actions themselves, about which we might ordinarily ask what reason or internalized position is driving them.

Note that I have assumed that institutions can, at least, take action in the first place. For example, I assume that there is nothing artificial or stretched in saying that a corporation hired or

fired someone, and that it made a promise or apologized. This remains a matter of debate: while some, such as Philip Pettit,[79] Margaret Gilbert[80] and Peter French,[81] accept that institutions and collectives can have agency, others, such as John Searle[82] and Christopher Kutz,[83] deny this possibility, arguing that only individual human beings can take action, because taking action presupposes having intentions, which these philosophers take to be necessarily a property of individual minds alone. Here I do not engage that debate, but take the affirmative position on institutional action, as defended by Gilbert, Pettit, et al, as a premise.

Moreover, given the particular actions I grant institutions can perform – apologizing and pledging, say – I further assume that institutions can perform complex actions,[84] where the actions are done in order to do something further. For example, if an institution solicits applications from job candidates, it can and ordinarily does so *for the purpose* of (the further action of) filling a position. Similarly, if an institution issues a pay check, it is to pay someone. And so on. Apologizing and pledging are in this sense "complex actions," [85] consisting of more basic actions such as communicating a message to an audience, and an end that such actions serve – such as taking a step meant to be recognized as apologetic (apologizing), or putting in place a commitment to do something later (pledging). From this a further claim follows: institutions can act for ends. The basic actions just mentioned serve the end of the additional actions; the former is done *in order to* achieve the latter. To sum up, then, I begin with the

[79] Pettit 2003.

[80] Gilbert 2002.

[81] French 2005.

[82] Searle 2010.
[83] Kutz 2000, esp. 7.
[84] I take this term from Searle 2010, 36.

[85] Ibid, p. 37n, quoting Danto 1968, 43-59.

everyday assumption that institutions act – they may hire, fire or issue demands, for example – and that they do so for ends, such as acquiring a staff or obtaining something from another party.

The problem I take up here is, then, not with institutional action, basic or complex, but with the sincerity of a special and morally critical *class* of action, which seems to require *more* than mere behavior. Specifically, my concern is with institutional action that commits the speaker to enter a stance, such as being forgiving (for forgiveness), grateful (for thanks), or apologetic (for apologies).[86] That, in turn, seems to require that certain reasons or normative positions (such as the view that one behaved wrongly, for apologies) be at work in guiding or driving the institutions at some point during and after they perform the speech acts. Such a requirement, however, seems impossible for institutions to meet. My aim in what follows is to show how institutions can and do satisfy it after all.

A. Why not people?

There is, however, an immediately obvious candidate for solving or even displacing this problem. It might be proposed that the sincerity of an institutional expressive speech act ("We forgive you") is, or is a function of, that of the relevant individual decisionmakers who decided that the statement be issued and who decide how the institution will act on it (these need not be the same people). Institutions are, after all, organizations of people, and their actions are taken or enacted by people, with whatever degree of sincerity these people have in doing so. (Indeed, the literature on collective intentionality and action tends to lump institutions like corporations and organizations with the class of "collectives" that includes teams, clubs or simply groups of

[86] Some would argue for even more: that these speech acts "express" or give voice to an active mental episode being undergone at the time of utterance and perhaps beyond. See Chapter I, supra.

people like Aborigines, Quebecoise and vegans.) [87] So, too, the various statements by such bodies are themselves formulated and uttered (or typed, or presented) by individuals, or on behalf of individuals. Perhaps, then, institutional sincerity is just individual sincerity, where the individuals in question are those with the right role in institutional decisionmaking.

One ground for taking this line is that, it has seemed to many, what makes some action the act of an institution is that the right people made it happen; either the decisionmakers assigned the task of deciding the institution's behavior,[88] or else an overlap in the shared activity of many members or participants.[89] For example, when a community such as a university boycotts a firm, typically a designated decisionmaker – such as a board or student council – will have instituted the boycott, or a critical mass of the community's members will engage in the boycott, or both. If so, the reasons and policies behind these acts will have been those of the relevant individuals who institute or perform them. Indeed, those reasons will have been expressly devised by people.

The problem with this proposal is that institutions can have policies and practices that nobody in particular holds or supports, a fact which follows simply from the nature of institutions. Therefore an account that equates, in principle, institutional stances with those taken by individuals will be at best under-inclusive. For example, one feature of at least some institutions is rationality. Put simply, if an institution is to have multiple policies, they should cohere at least some of the time. Yet this simple desideratum opens the possibility that institutions will be committed to policies its members or decisionmakers do not support, if only

[87] See, for example, Searle 2010, and Kutz, 2000, and Velleman 2007, 29 (suggesting that what it is for a 17-member philosophy department to come to some view or other is the same question as what it is for those 17 people to do so).
[88] See, for example, Searle 2010.
[89] Kutz 2000, 7.

because these policies are entailed by others individually but not jointly supported by the members.

As Philip Pettit has argued, drawing on the work of Lewis Kornhauser and Lawrence Sager,[90] an institution's rationality depends ultimately on its ability to reach decisions or policy positions that are not shared by most of its individual members.[91] He illustrates this problem, which he calls the "discursive dilemma," with an example, which I adapt here slightly: suppose a department foresees a possible, but distant, need to install coffee vending machines at the last-minute, when there won't be time to vote on which one to buy. So a vote is taken in advance, to prepare for the possibility. The vote is on the purely hypothetical question of which company to call and patronize if they end up having to make the purchase when nobody is around. The resulting choice, by a six out of 10 majority, is Nestlé's, say. Much later, perhaps, a second vote is taken on whether to finally buy a machine after all, with six out of 10 members voting yes. The important point here is that now, given the two voted policies, the department is committed to buying a Nestlé's vending machine. Yet, this would be so even if as many as eight of the 10 voting members vehemently oppose buying a Nestlé's vending machine, not only privately but as reflected in how they voted: they're either among the four who opposed Nestlé's, in particular, due to unethical corporate practices, say, or the four who oppose the machines altogether. These various dissidents were in one or another minority in each of the particular votes, but together they'd be a majority in voting "No" had the question been: "Should we select a Nestlé's coffee machine?" Yet the department, itself, unlike its members, is now committed to taking the "Yes"

[90] Kornhauser and Sager 1986.

[91] Pettit 2003, 168-70. (The problem is referred to as the "discursive dilemma," in that it involves putatively singular group agents reaching divergent decisions; indeed, they contradict each other. At least that is what a requirement for rationality or coherence would aim to avoid.

view on that same question – due to the implications of its two prior votes. In short, the resulting institutional choice – selecting Nestlé's as the supplier – might not be that of the majority of its members. But institutional rationality requires that the department make that choice, as it follows from prior choices made by distinct majorities of decisionmakers at the time.

In these examples, of course, one can at least trace the ultimate institutional position to those of various people at different times, perhaps even attributing to them some constructive consent that their enacted position will ultimately bind the institution in ways they might not favor. But there are reasons to think that some institutional positions will emerge from none that anyone ever held, or even from a consequence of any such position. This is because of the structure of institutional decisionmaking, which can constrain the decisions made. One property of that structure, already considered, was rationality. Another (related) property is precedential authority: present-day members of institutions are bound by policies earlier enacted. But, as the cases of constitutional and common law famously illustrate, old policies can underdetermine their precise application to new cases that were unforeseen by the original policymakers. That could require interpretation and application by those who never supported the policy in the first place, and these new interpretations will, by hypothesis, outrun the intentions and imaginations of those who did. The resulting policy application will itself be a policy, of the form Policy X in case Y yields rule Z.

For example, a social media company like Facebook may prohibit the use of false personas, insisting as much in the User Agreement and admonishing members to maintain only their original "profiles," to the extent that these honestly represents their identity, rather than make new "trolls." However, the policymakers may never have anticipated the possibility of

multiple "true" identities reflecting genuine, if distinct, personas or roles taken up by people, such as one as a school teacher and another as a fierce advocate against Meghan's Law. By hypothesis, the original policymakers did not, or need not, have a view of the matter, while those currently charged with interpreting the policy will be bound by its terms. So they will face the question of whether partial, incomplete or multiple personas are by nature deceptive, and so violative of the original policy (whatever its intended target), or instead fall within its parameters as long as they accurately represent their users (however many times, in however many ways). Yet it can be allowed that the new policymakers, greedy as they are, do not themselves support the ban on false personas in the first place; indeed, they welcome fake trolls. So the ultimate decision on how to apply the original policy could, in principle, reflect nobody's intention as to how the company should treat these cases. As with rationality, then, the constraint of precedential authority enables the adoption of policies that nobody in particular supports or intends.[92]

These are just two common examples of institutional decisionmaking constraints that could lead to policies unintended and unsupported by anyone involved; "emergent policies," for short. Others can be imagined as well; a search committee chooses its favorite candidate by a "maximin" method, and, as things turn out, all the top two or three choices, out of 10, perfectly conflict, the same candidates equally loved and hated by various members, thereby canceling out. But everyone's fourth choice – about which nobody was enthusiastic and everyone hoped they wouldn't be stuck with – verges on the same candidate. So the committee issues an

[92] In saying the prior policy does not, on its face, determine the answer to the new case and that the prior policymakers did not foresee it, I do not mean to suggest there is no correct or better reading of precedent that could decide the new case. There may, in fact, be a right answer as to how to apply the old policy. But what's important for the present discussion is that such a right answer – like any answer here – can, hypothetically, be one that neither the original policymakers nor its current authoritative stewards support. For the analogous "right answers" thesis in the case of common law, see Dworkin, 1985, 119-45.

enthusiastic letter to this candidate as its "first choice" – which he is – even though he was, in fact, nobody's first choice. The upshot of these sorts of examples is that there is no reason to restrict the class of institutional policies and positions, in principle, to those that at least some individuals hold. They can be held by nobody at all, in which case there would be no relevant individual decisionmakers in whom to base their sincerity. For that reason, something besides the internal states of some individual decisionmakers must be sought as the basis of institutional sincerity in principle.

II. A positive account

(a) From individuals to institutions

There is, in other words, little hope of reducing the sincerity of institutional forgiveness, repentance and the like to that of the decisionmakers involved. One reason, argued in the previous section, is that there is no reason to think an institution's policies or reasons for action are those of any individual involved. They can be, but they need not be. As a result, a theory of institutional sincerity cannot be based on the sincerity of institutional decisionmakers. That, however, leaves us with the original problem. If institutional speech acts like apology and forgiveness require, on pain of insincerity, entering stances like apologeticness, which requires having and acting for certain reasons, then their lack of mental life may render institutions necessarily incapable of sincerely apologizing, forgiving and so on. That would render problematic or at least defective a wide class of actions ordinarily attributed to institutions, along with a familiar way of speaking about them.

To summarize the problem, then, institutions appear to be unable to meet the requirements of sincere stance-taking (forgiveness, repentance, etc.). As I have argued earlier, sincerity in taking stances through speech acts such as "I forgive you," requires (at least):

(1) Being disposed to act toward the target of the speech act in conformity with the reason or normative position expressed by the speech act;

and, because these speech acts *commit* the speakers to the stances they've taken,

(2) Being disposed to ensure that one remain disposed to act as in (1);

And, most importantly for present purposes:

(3) Being so disposed (as in 1 and 2) for the reason or in light of the normative principles expressed by the stance.

To be sincerely forgiving, then, requires at least (a) being disposed to act toward one's wrongdoer as directed by the reasons forgiveness expresses, including the position that he should no longer be resented, punished or blamed (so no acts of retribution or passive-aggressive verbal snipes, for example); (b) to try to maintain that disposition (by 2), and (c) to be so disposed *for the reason that*, in fact, he is no longer worthy of blame, resentment, etc. The final criterion – (3) – is important, because it distinguishes sincerity from mere *consistency*. We can, after all, imagine the case of the false forgiver: someone disposed to act *as though* one's wrongdoer is no longer worthy of blame and punishment, but not in fact internalizing that view at all; indeed, he may privately hope and even wish for the wrongdoer to suffer for what he did. His behavior, then, would be *consistent* with the reason expressed in the stance, even *guided* by it – but it

would not be done (nor would it be what he is disposed to do) for the right reason. And that is why he would be insincere.

The problem for institutions, however, is that (3) seems to require an inner mental life. In contrast, it would appear institutions *can* fulfill the first and second requirements of being disposed to act *in conformity* to reasons and to take steps to ensure that they stay so disposed. Take the earlier example of a racist restaurant chain that seeks to become sincerely repentant and apologetic about its past racism. Grant, further, that an apology expresses that the corporation owes the victims (i.e immigrants) not to have treated them as it did. As a result, on requirements (1) and (2) above, it must be disposed to act with special deference and respect toward them, it must seek to reverse the earlier treatment, and must avoid any repetition of it. In addition, having committed to this apologetic stance, it must take steps to ensure that these remedial acts continue and that no contrary or inconsistent activity be undertaken alongside them. Perhaps an outside consultant would be hired to review the progress and status of the corporation's repentance, and make recommendations as necessary to maintain the corporation's commitment to the mistreated group.

Still, as noted, the problem remains with requirement (3) – that all these dispositions to act (and second-order dispositions to maintain them) be for the right reason. Otherwise, all that is left for the institution is mere *consistency* with the stance it has adopted, rather than any actual internalization of it or the reasons for it. And consistent behavior – like the speech act that precedes it -- is just the sort of thing that could be said to be done for the right reasons, rendering the stance sincere, or the wrong ones, rendering it insincere. So even if institutions can be disposed to act, in just the ways that the reasons expressed by their speech acts would dictate, the problem of sincerity simply resurfaces with the dispositions and actions themselves.

(a) Reasons and ends

The problem appears to be that institutions can take actions, even those that reflect a certain reason, but they cannot take action *for* some reason. The appearance of the problem, however, depends on a *causal* reading of the requirement (3) that they be disposed to act for the right reason. On this reading, acting for reasons implicates two distinct phenomena – having reasons and performing actions – the one mental and "internal," the other behavioral, with the former causing the latter. On this view, it would be difficult to show that institutions can act for reasons, especially those that have no formal procedure in place for adopting or accepting them.

But we need not accept the causal interpretation of acting for reasons. An alternative reading is available which both accommodates institutions that lack mentality, while retaining everything important in acting for reasons. Recall that institutions can act for ends, or so we assumed in Section I. It follows, then, that they can be disposed to act for ends. Those ends can include the fulfillment of the normative requirements specified in the reason for some stance-taking. So, for example, an institution that has forgiven another will, if sincere, be disposed to act for the end of treating the other as unworthy of resentment, punishment and blame, in all possible circumstances in which it takes an action relevant to the other. A country that has pledged peace towards its northern neighbor would be disposed to refrain from attacking the neighbor, or arming her enemies, or tolerating a volatile, hostile relationship with her. Conversely, an institution that has *not* sincerely adopted some stance, such as apologeticness or repentance, will *not* be disposed to act for the end put in place by the normative demands of the stance (that one's victim be entitled not to be wronged as one did) if it is prone to repeat the offense and even sets out to do so. As I will try to show, there is no difference between being disposed to act for the right end, in these ways, and acting for the right reason – as required by sincerity. The former is

merely another way of putting the latter. And it is a way that fits nicely the capacities of institutions, which – as we assumed earlier – can act for ends.

Consider a case of an institution which performs all the right actions, each one consistent with the stance it has taken, but the institution is actually *insincere* in the adoption of the stance. On the instant proposal, that would mean it takes the right actions but for the wrong reasons. A familiar form this might take is that of acting on an *ulterior motive,* a subspecies of acting for the wrong reason. To return to the case of the restaurant chain, suppose the sole purpose of the apology and its new policies and practices is to avoid a lawsuit and any bad publicity from being exposed as racist or xenophobic. In practice, this insincerity will never come to light, since the chain will treat all employees and customers equally and respectfully in any event. But the ulterior motive of avoiding lawsuits and bad press appear to render the behavior insincere. If anyone found out, they would have a legitimate gripe with the company.

Importantly, one way of redescribing the restaurant's ulterior motive for acting is through a hypothetical: if there were no threat of lawsuits and bad publicity, the corporation would *not* be inclined to treat immigrants fairly.[93] In those sets of circumstances, unlike those that actually obtain, the company's actions would not be consistent with the stance described. Its entire set of actions and possible actions on this matter, then, serve the end of avoiding lawsuits and disrepute, rather than the (sometimes overlapping) end of treating people equally, period. Here, it does not act for those moral ends, and may, in fact, act in ways that contravene those ends wherever treating immigrants fairly does *not* serve the end of avoiding legal or media trouble. In

[93] This is not to say it will be inclined to treat them unfairly; just that it would no longer be true that it would be inclined to treat them fairly.

those cases, the restaurant chain would not be disposed to act for the end of treating people equally.

In fact, any case of having the wrong reasons can be redescribed this way: as an instance of acting for the wrong ends. If the chain aims to treat immigrants well merely for some benefit or purpose unrelated to the moral reasons to treat them fairly, then it would not be disposed to act for the end of treating immigrants better, but for some other end, yielding a disposition to act differently in cases where the ends conflict. So if sincerity requires acting for the right reason, as it does in the case of individuals, and acting for the right reason amounts to acting for the right end, then institutions can sincerely adopt the stances in question, such as forgiveness, repentance and reconciliation.

Of course, for all that has been argued so far, being disposed to act for the right ends may be merely a necessary but not a sufficient condition of acting for the right reason. The latter may also require something internal, after all, which institutions can never have. If so, then the account proposed here would be too behavioristic, eliminatively reducing states of mind to dispositions to act for certain ends.[94] But when we attend more closely to what is involved in internalizing a normative reason, it emerges that acting for the right end suffices, after all.

Consider candidates for the "right" reason in the case of improving treatment of immigrants. One might be that it is wrong to treat any class of people as inferior. This sort of claim is a normative to-do claim, which could be redescribed without loss of meaning as: all classes of people are to be treated as equal to each other, and none as inferior. This claim, in turn,

[94] This *would* be a behavioristic account if it equated typical or putatively "internal" psychological states with dispositions to act, as Gilbert Ryle does with knowledge and emotions. Ryle 1949, 83-149. Instead, my position is that not all states that speech acts express or commit one to require typical or internal psychological states, and that among those that do not are institutional stances.

amounts to a call to act for a certain end, namely the end of treating people equally, none as inferior. This reason, then, is merely another version of specifying a moral end. One could, of course, propose a richer moral reason: immigrants are to be treated equally because they are just like everyone else. Such a claim is much less easily formulated as a call to act for some end. But as Hannah Arendt points out, the beliefs in human equality of the sort that ground rights claims are never literally descriptive beliefs.[95] They do not amount to a belief that all human beings are in some morally relevant way alike; indeed, they arise prior to any empirical knowledge of the particular people involved. Rather, belief in human equality amounts to belief in the normative requirement to *treat* them equally, which can be redescribed in terms of a call to act for ends.

Put differently: the reason for which the restaurant chain must be acting is the normative position that immigrants are to be welcomed and included and treated equally. What is involved in internalizing such a reason or normative position? One answer is this: to be committed to treating immigrants equally and inclusively, period. For example, to internalize the view that friends should be trusted is to be disposed to act for the purpose of trusting one's friends and to be committed to continuing to do so. Internalizing a reason, in other words, can consist in being disposed to act on it for the purpose of doing what the reason specifies should be done. And as already emphasized, institutions – like restaurant chains, governments and political parties – can act for ends, from which it follows that they can be disposed to act for ends, including the right ends. Since acting for the right reasons is redescribable as being disposed to act for the right end, institutions can be disposed to act for the right reasons. And that, as argued earlier, is all that sincerity requires.

[95] Arendt 1973, 292.

Thus the requirements of sincere stance taking can be redescribed without loss of meaning as the following three conditions, though the third is actually implied by the first two and is separated out here only for clarification: one has sincerely taken a given stance (such as apologizing) to the extent that one is (1) disposed to act in ways consistent with the stance; and (2) disposed to ensure that one remain disposed as in (1 and 2); and (3) disposed to do so for the end prescribed by the normative position that motivates and justifies the stance-taking. And we have seen that institutions can meet these requirements. For (3), in particular, normative claims about what is to be done, or not to be done, can be adopted by institutions as an institutional reason or policy. That includes a claim about whether someone is to be treated as worthy of sanction, punishment or blame, say, as in the case of forgiveness.

(B) Recognition

I have sought to show that acting for reasons, an important requirement of sincere stance-taking, can be redescribed in a way that accommodates "mindless" institutions. If the effort was successful, would there be anything left to sincere stance-taking that institutions still cannot satisfy? At least one possibility remains: certain stance-takings seem to require that one recognize or believe certain things, over and above how one is disposed to act and for what ends. Take the stance of being apologetic or repentant. An important requirement is that, in addition to being disposed to treat the victim the right way, and for the right reason, an apologizer should appreciate or recognize that he was wrong.[96] A component of the Truth and Reconciliation Commission in South Africa was the acknowledgment of oppression on the part of the apartheid regime and its supporters, hence "*truth* and reconciliation." It would not have been enough had the former regime simply undertook to treat its black victims a certain way for all the normative

[96] Thanks to Herb Morris for raising this example, which he called the requirement of "memory."

reasons they should be so treated. The regime also had to acknowledge the fact of their culpability. "I was wrong," after all, is considered a paradigmatic apologetic statement.

This further requirement of at least some sincere stance-taking – call it the recognition requirement – seems a serious obstacle to institutional sincerity in those cases. For we have allowed only that institutions can act for ends; not that they can also entertain beliefs or be cognizant of some fact or other. Indeed, the latter possibility may be off limits or even nonsensical. I want, at least, to grant the premise that institutions cannot be in a state of believing or cognizing some truth. It is, then, incorrect to say at any point that an institution, like a court or corporation, is believing X or knowing that Y. The recognition requirement, then, casts doubt on the capacity for institutional sincerity. It does so, however, only by exploiting a vagueness in the notion of acting for the right end or reasons, as we grant that institutions can do. If we attend more closely to what is actually involved in acting for certain ends, the worry begins to dissolve. That is because acting for the right end, in the case of the relevant stances (apologeticness, say, or repentance), essentially involves the required recognition, as well.

Consider ordinary cases of action for moral ends. Suppose A accidently took a book off B's shelf, and then misplaced it, losing it for good. B calls for A to replace the lost book. Finally, A arrives with a replacement copy, in the same condition as the original, and a card with the printed words "sorry" taped to the book. The action A is performing can be accurately described as for the end of replacing the book he wrongfully misplaced. The fact that A took the book in the first place, then, figures in defining the end of his action. It would not make sense to ask of someone who is genuinely acting to replace an item he wrongfully took, "But do you concede that you took it?"

The example is meant to bring out two features of recognition that preserve its possibility at the institutional level: first, one need not be in an active or occurrent belief state to recognize something. Indeed, we see examples of this even with individual people. When we consistently defer to someone else, it might correctly be observed that we recognize her greater skills and expertise in some area – even if we never actively formed that thought that she has them. Second, if someone acts for the end of responding to some event, then she recognizes that the event took place. If I return an item for the end of making up for destroying it, I recognize that I destroyed it. With these two conclusions, we can revisit the case of institutional sincerity in stances like repentance and forgiveness. Suppose that repentance requires recognizing that one did wrong. The fact that an institution lacks the capacity to form the active or occurrent belief state along the lines of 'I did wrong' is no barrier to its recognizing such a fact. Second, more positively: if the institution is disposed to act for the end of making up for having done a wrongdoing, it does in fact recognize that it did the wrongdoing. Similarly, if forgiveness requires recognizing that one's offender did wrong,[97] an institution can fulfill this requirement by acting for the end of not punishing or blaming or resenting him for having done wrong. Since the end is a response to the event, taking the end for one's action involves recognizing the event.

It may be worried that I have helped myself to too rich a description of ends, one that implausibly builds recognition into the ends institutions can set for themselves. If there is some barrier to institutional recognition, perhaps it extends to the kind of non-occurrent or inactive recognition involved in setting ends of action. Two responses to this worry are in order: first, the assumption that institutions cannot recognize facts or events is not self-evident. Its plausibility, I think, lies in the apparent inconceivability of institutions entertaining or cognizing things (like

[97] For a defense of the claim that it does, see Hieronymi 2001.

that they perpetrated a fraud or a persecution 50 years ago). Once we disaggregate such active mental states from the recognition, the phenomenon becomes easier to ascribe to institutions. Second, the recognitional component of an end is not mere baggage; institutions who have ends with recognitional components will be disposed to behave differently from those that do not. For example, a country that resolves to make up for past oppression will be disposed to act for a different end from one that resolves only to take responsibility for some past oppression, leaving open whether it was culpably involved. True, both states may engage in reparational behavior. But as repentance requires acting to prevent repeat offenses, the one with the end of making up for *actually* oppressing someone will also act to educate and warn itself against repetition, to make public the fact of its oppression especially in response to inquiries from the victims, and – in those instances when countries issue historical records – it will report the event. The recognitional component of the end has bite, and in many cases it also does important moral work. In short, if sincere stance-taking requires recognizing something, an institution can full that requirement meaningfully.

(C) The way it feels for persons

I have argued that institutions can meet all the requirements of sincere stance-taking, including being disposed to act the right way, for the right reasons (or ends), and being internally committed to maintaining these dispositions. They can also meet requirements to recognize some historical fact or event. Nevertheless, as may seem obvious, not everything will be the same across human and institutional stance taking. As already noted, when an individual sincerely expresses remorse, she may feel guilty or badly, and this will involve a negative feeling or episode. In fact, sincere apologies by individuals may necessarily involve such feelings, so much so that if one apologizes and feels uninterruptedly ecstatic one has not apologized sincerely,

whatever reasons one has internalized. This aspect of the "feel" of an apology may be necessarily tied to being sincerely apologetic or repentant in human persons. And corporations and countries cannot share in this aspect.

Indeed, one may worry that human stance-taking, in cases like apology, repentance and forgiveness, are so rapt with feelings that a stance without them would seem impoverished, perhaps even uninteresting. That something could undergo repentance without feeling badly may be possible – indeed, it is, for all that has been argued -- but unimaginable and almost unrecognizable as a state anyone could relate to or demand.

I have already argued elsewhere that emotional states or feels are not among the individuating conditions of stances. To be apologetic or repentant requires merely the dispositions already laid out, and the right reasons or ends driving them. One reason to reject an emotional requirement, even in human beings, is that states like forgiveness and repentance are meant to last over long stretches of time and be expressed at particular moments in time. If they were also to require distinct mental episodes like affective emotional states, these would, in turn, have to be maintained improbably long and summoned immediately at will, both of which would prove implausibly demanding. For those reasons, I have claimed, individuals need not feel a complete absence of resentment,[98] say, to sincerely forgive someone, or feel negatively about what they did to apologize sincerely.

Nevertheless, the sense that feelings are very important to such stances persists. I believe this sense can be explained by a close natural connection, even a natural necessity, between meeting the essential individuating conditions of stances and certain familiar feelings in human

[98] See, for example, Griswold 2007, 33.

persons. If, for example, repentance requires having as one's end, or reason for action, that one should make up for a horrible wrong one did, and that no action can succeed at this effort, this may necessarily be experienced unpleasantly by human beings. They may feel guilt, regret, tension, frustration – all unpleasant states. That may be what it feels like to internalize the reasons involved in repentance. But the connection goes deeper. Recall that apologies, formal offers of forgiveness and similar states are put in place by commissive speech acts, which commit the speaker to maintain the stances she has taken. That has significant consequences for feelings, if the natural claim above is correct: recall that being internally committed to a stance is being disposed to maintain it. If maintaining, say, forgiveness in humans is naturally undermined by feelings of resentment or hostility, or repentance by feelings of complacency and pride, then the requirements of sincerity will lead humans to shun these feelings. They will, indeed, resist feelings that threaten the stance they are trying to maintain and welcome, even cultivate, feelings on which it thrives. All of this simply reflects the natural connection I'm positing in human persons between internalizing the reasons of some stance and having distinct affectations or feelings.

The close connection between feelings and human stance-taking, as I'm claiming is a natural fact about human beings, helps explain why it is difficult to imagine a state like repentance without guilt or regret, or forgiveness without either a calming effect or, alternatively, fighting the bubbling up of resentment and bitterness. Indeed, these feelings are not merely contingent baggage on human stance-taking. As already noted, they help bolster and maintain the stances. Additionally, they enrich them: a stance like repentance becomes more intense and all-consuming as it involves more of the stance-taker's faculties, both cognitive, behavioral and affective in the human case.

Yet institutions, too, may have unique ways of realizing the requirements of sincere stance-taking, which likewise enhance their power to enrich and maintain the stances to which they've committed, be they forgiveness, repentance or gratitude. As Seana Shiffrin points out, institutions accused of racism or a racist past, such as a university, have ways of repairing past wrongs that individuals cannot achieve on their own.[99] For example, institutions have a greater capacity for consistent behavior over time (though they often do not live up to it). When institutions resolve to do something for posterity, they can enshrine the policy in such a way that it becomes very difficult to overturn, even when enthusiasm for it wanes. A person, in contrast, has a more difficult time keeping his earlier commitments unless he persists in being privately committed to seeing it through. Similarly, tribes or communities may engage in rituals. These steps are among the sorts of resources available to non-personal institutions for manifesting the practice of their principled stances. That they differ from those of human beings is not only expected but a bit of luck, in some cases: they can do what human beings can't, as in the case of instituting public acts, inaugurating monuments and buildings, and generally involving multitudes of people. Nevertheless, at core, what makes these steps sincere expressions of some stance or other is the same as that for individual persons.

III. Objections

A. Many minds and many reasons

The previous section advanced an account of institutional sincerity based on dispositions to act in certain ways and for certain purposes or ends. A corporation sincerely apologetic about its past practices will act in distinct ways reflecting this stance: it would adopt new hiring criteria, issue apologies, treat past victims as moral debtors, and many other things. It would also

[99] See Shiffrin 2009, 336-37.

take steps to ensure that this disposition remains in place. Needless to say, these many steps depend on particular decisionmakers. For example, the personnel director will make hiring decisions; the spokesperson may issue apologies and offers to compensate victims, and the payroll department may process any compensation. The firm's board may vote on these practices and on ways to maintain and improve them in case the firm begins to lapse in its commitment. Various particular people, in other words, will carry out these actions ostensibly reflecting the firm's disposition to act apologetically.

That invites the following objection: people act for any number of reasons when they enact or institute or initiate corporate behavior. In fact, the reasons of particular agents for taking action could have nothing to do with the stance those actions might, incidentally, maintain. As a result, it would be a mistake to subsume the institutional actions they bring about into the general dispositions of the institutions; the particular whims or motivations of one or another institutional decisionmaker are only contingently related to the way an institution, qua institution, behaves. They cannot form part of its general dispositions to act. And that means that such externally motivated institutional actions cannot be guided by the types of reasons or ends that, on the analysis of the previous section, must guide institutional action if it is to constitute a sincere stance.

Put more starkly: I have just now argued that institutional sincerity depends on being disposed to act for certain ends. But where an institutional policy or decision is voted on by groups of people, the reason or end of the resulting decisions is quite plainly those of the people who enacted it. And there is no reason to believe they are doing so for the relevant end or reason specified by the stance the institution has taken, such as forgiveness or repentance. To the contrary: there is every reason to suspect that individuals have their own private, or at least

institutionally irrelevant, reasons to vote for some institutional policy or practice. Cass Sunstein, for example, argues that when people come together in institutional contexts – a court, for example, or a legislature – their inputs, in the form of the policy they vote for, all may have different personal justifications, different reasons motivating them.[100] But in agreeing with others, the relevant players acquiesce to shedding their own particular reasons for favoring the policy or decision, as contrasted with those of their fellow decisionmakers. In coming together, in other words, legislators, judges and other institutional voters leave the particularistic sides of their principled stances behind. What emerges is a kind of theoretical least common denominator: the bare decision to adopt this policy or institutional decision rather than another one. The *reasons* to do so are left behind. It is not only difficult, then, but conceptually wrong to attribute some reason or normative claim to the prevailing institutional decision, because by procedural design it emerges undertheorized, or "incompletely theorized," without a clear motivating principle that could be discerned.

On this understanding, very few collectively reached policies could be said to follow from a particular reason or commitment. In nearly all cases of institutional decisions, then, there would appear to be no way of locating a reason or policy that guided it. Even on a less radical view than Sunstein's, it would be hard to bridge the private reasons some institutional decisionmaker applied to his voting action with any putative commitment or reason at the institutional level. A congressperson who pursues certain violators of federal law may well be acting on behalf of the government, but perhaps not out of any allegiance or commitment to the government's reasons for the law. Rather, she may have found the cause a conveniently popular one in her district, and hopes to ride it to another term. Given the possibility – and apparent

[100] See Sunstein 1996, 23-36.

prevalence – of institutional action taken or carried out by people with non-institutional reasons of their own, it may be too idealistic to imagine that many institutional actions are themselves guided by some reason or commitment of the institution itself, as required by my criteria for sincere institutional stance-taking. If so, then perhaps the account itself is too ambitious, as it would render a great many institutional stance-takings insincere or ingenuine.

This objection, however, threatens the notion of institutional reasons, and therefore institutional stances and states, only to the extent that institutional action is taken *exclusively* for private, non-institutional reasons. But there are, in fact, three possibilities at work either separately or in combination: when people initiate institutional action, they either attempt to follow institutional practice and policy, or they set such policy anew, *or* they act on their personal preferences alone. In the last case, their actions will by definition be isolated from institutional practice, and will not constitute the general dispositions of the institution. Those actions will therefore not figure into an assessment of the institution's sincere adoption of some stance or other. If too many of the institution's practices have this character, then – as in Sunstein's picture – it simply will not have dispositions to act by certain reasons, and therefore, on the analysis here so far, it cannot genuinely take stances.

But that will be the rare case, because, as already brought out in the arguments of Pettit and others considered in Section I, institutions have pressures to maintain some coherence and consistency with past policies. And as long as the action of institutional "voters" fall into either of the first two categories – following or setting institutional policy – it will hardly matter what, besides that, motivated it. Consider again the example of the congressperson making a personal crusade out of a federal ban (on smoking in indoor public places, for example): she may be motivated primarily by the high incidence of lung cancer in her district, including in her own

family, say, and by a related desire for reelection. But she is *also* pursuing the ban because (she is quick to emphasize) it is *federal policy,* or because she plans to make it so. In either case, its enforcement will be *at the same time* guided by the reasons that put the policy in effect (which is a necessary condition of sincerity), even though she has her own compatible *further* reasons to pursue it. In such cases, in other words, it remains quite accurate to say that the institution (in this case the government) is disposed to act that way for the normative reasons that guide it. Those institutional actions will simply be overdetermined; policy will be one reason for the decision, but so will whatever else happened to motivate the individual decisionmaker who acted on behalf of the institution. Importantly, then, the overdetermined institutional actions (like legislating and enforcing the ban here) are still among the many acts the institution is disposed to do for the reasons relevant to sincerity, even if those aren't the only factors at work.[101] If one were to describe the institution as disposed to do X for reasons Y, one would be correct no matter what else motivated the individual decisionmakers to help it do X. Consequently, the institution would be sincere in its stance-taking.

If this seems too accommodating to multiple reasons or motivations, it might help to recall that an analogous situation is familiar in the case of individual human agents, too, when they act for reasons or are disposed to do so. They too have any number of reasons and motivations coinciding to guide a single action. When I promise to give something to my neighbor, I may be committed to keeping the promise for all the right moral reasons (a duty to do as I say, for example). But my commitment may also – even mainly – be motivated by a desire to show up my lender, who openly doubted I would keep my word. So my commitment to keep my promise will be overdetermined, even though I can accurately be described from the time I

[101] In future work I plan to argue that laws are a case of institutional stance-taking, the institution being the government (not the country or the citizenry).

commit until the time I fulfill as disposed to enable myself to pay back the lender for the reason that I owe it to him and have verbally obligated myself to do so. My further reason does not negate the one relevant to my sincere commitment or stance-taking. Nor do further reasons do so for institutions.

B. Groups and Collectives

Recall that on the preceding arguments, an institution can sincerely adopt a stance like apologeticness, forgiveness and gratitude, say, if and only if it accepts or acts in light of certain reasons: that some party is no longer worthy of resentment, punishment or blame, for example. More important, recall that in Section I, it was argued that these states are not reducible to those of individuals involved in the institution. A country like the U.S. can be sincerely apologetic towards victims of internment camps during World War II without its citizens being so apologetic at a significant level. As long as the country apologizes and remains disposed to engage in steps of moral repair for the right end or reason – say that they owe a great moral debt to the victims of this horrible wartime policy – it has entered a sincerely apologetic, even repentant, state.

This may invite the following counterexample. Suppose a group that has been unjustly treated in the past continues to demand reparations. Say, a labor union has demanded compensation for exploitatively low wages in the past. This demand has come to be issued formally by the union secretaries and delivered each year, already enshrined as routine procedure, with mechanisms for adjustment due to inflation. Yet this year, an independent poll

uncovers that, as it turns out, none of the members individually want the reparations anymore, if they can even remember why it was demanded. Not a single one is interested in receiving it; though, equally, they do not endorse changing the procedure or amending the policy that results in the union's persistence in seeking reparations. The indifference of the members would seem to falsify the claim that the union demands or genuinely seeks reparations. As individuals, not a single one wants or seeks it; how, then, could a union that consists of them be said to want or seek it? This kind of worry is even starker in the case of groups like minorities or communities; a displaced group of people seeking the right of return would seem in a kind of internal tension if all its members turned out not to want to return home.

I share the intuitive reactions to these cases, but I do not take them to be counterexamples to the claim that institutions can genuinely and sincerely take stances of their own, for reasons that none of their members have internalized. What the union and refugee cases track, rather, is that the groups in question are ambiguously understood – both as singular institutions and as collections of individuals demanding things for themselves. The union doesn't demand a lump sum for the union; rather it demands compensation for its members and on their behalf. Similarly, the refugees seem to demand the right *for each of them* to return to their homeland, rather than for the group as a whole to be recognized as entitled to the right. Consider, in contrast, the state actions taken by Qadaffi's Libya over the years. We would ascribe those actions to the Libyan state even if we found that not a single Libyan official or citizen besides Qadaffi endorsed and approved of them. States are structured so as to act for themselves, as a whole. While it is tragic and unjust when a state acts undemocratically, we still attribute those acts to the state. In contrast, a union or a group demanding rights seems to act or speak on behalf of the individuals who make up the group in question.

In other words, it is the nature of the union's or refugee group's *demand*, not the demander, that belies the ostensible stance whenever the members do not share it. If the union demanded, instead, that its president be given one of the executive parking spaces reserved for managers in the corporation, it would seem to matter much less whether most members were invested in this demand.

A similar sort of falsification occurs when the group is being used as shorthand for the individuals who make it up. Thus it might be said that the Armenians resent the Turkish refusal to acknowledge their genocide. That state of resentment, then, is not attributed to the organized Armenian community or its official institutions. Rather, it is a generalization made about many individual Armenians. Therefore, it will be unsurprising that such statements would be falsified by a finding that very few actual Armenians resent the Turkish refusal, even if the tiny minority of flag bearers includes, say, the executive director of the most powerful Armenian rights organization.

There are, in short, three different ways of describing groups: first, as a collection of individuals (as in saying, "Holocaust survivors resent the casual use of the terms genocide and ethnic cleansing today"); second, as a collective of individuals acting together or in unison, as in saying, "The class of 1965 reunited every year to honor her"; and third, as a single agent – as though the group is a being in its own right that makes decisions, enters agreements and takes action – as in saying "The Armenian People applied for representation at the United Nations." In the last example, where group talk is talk of a single agent, the claim is not that some number of Armenians sought to obtain UN representation, much less that all of them did. Rather, the example refers to an action taken by the single entity or agent that the Armenian group happens to constitute. This way of talking implies that a group can constitute an institution, over and

above being a collection of people, even one that acts collectively. It also implies that this institution can take actions for reasons that not all its members share at the time. For example, if union members become fed up with paying dues to a lawyer who has neglected their case, and therefore fail to press the relevant parties to pay back legal fees on time, they may become delinquent, and an outside administrator may be empowered to apologize for all late payments including this one. In such a case, the union will have apologized – and it can, for all that has been said, apologize sincerely – even where many union members are not themselves apologetic about it, or even aware of it.

Importantly, my account of institutional sincerity applies to groups only in the institutional sense. Yet that is an important sense of group talk. When the black community in South Africa is described as having forgiven its white oppressors, this refers to talk of the blacks as an institution, acting by way of its representatives. The description does not *prima facie* entail that some amount of blacks have privately adopted a stance of forgiveness. Notice I deny only the *prima facie* entailment claim, in this case. That is because certain institutions may, in the end, be structured in such a way that in order to act or be disposed to act for certain ends, some number of individuals must do so as well. A sports team, for example, may be like that: although we may say that a team forfeited the game, and mean something different from saying that all its players stopped trying to win, the latter may just be a necessary condition of how teams, in fact, forfeit games: their players stop trying to win. That's just how teams are structured, perhaps. And I suspect the same is true of communities like the South African black or white community mentioned above: for the community to forgive, it may be that some people (particular members) may have to do so, in light of the particular way the community and its means of taking action are structured. All I mean to deny is the logical *equivalence* of statements about group-level

stance taking – 'the black people forgave,' 'the Armenian People withdrew its application' – and statements ascribing the same action to some relevant amount or combination of people.[102] The institutional sense of groups is, in other words, a real sense, distinct from that of collections and collectives.

The availability of this institutional description of groups – to which the present account is exclusively relevant – is a reason to criticize the dominant treatment of group agency and group minds in the philosophical literature.[103] That treatment analyzes group agency or collective intentionality solely in terms of the shared or overlapping intentionality of individual members. To that extent, it neglects the reality of groups as institutional agents, and stands to misdescribe what is involved in, say, group-level remorse, resentment or forgiveness.

VI. Institutional rights?

I wish to turn now to what may seem a dramatic consequence of what has been argued so far. In particular, I have argued that institutions can genuinely and sincerely take certain normative positions, such as that one is to be treated as having acted with undeserved generosity towards oneself (gratitude) or as no longer worthy of blame, resentment or punishment (forgiveness). All they need is the capacity to act in light of certain reasons, which itself requires only that they be disposed to act certain ways and to maintain that disposition in the face of potential lapses. It follows that an institution can genuinely and sincerely endorse a political

[102] I say "logical equivalence" because it may be possible to substitute a claim like 'Institution A acted' with the claim that some complex of members did what is necessary to instantiate that 'Institution A acted,' but that substitutability would be entirely contingent; the same institution could conceivably be structured differently so that the substitution doesn't work.

[103] See, for example, Searle 2010, v ("the only intentionality that can exist is in the heads of individuals. There is no collective intentionality beyond what is in the head of each member of the collective").

candidate, express support or opposition to a cause, or pledge loyalty to another. In short, they can genuinely and sincerely take positions.

This may be read to suggest that institutions – including corporations – have some claim to protection for these expressed positions, at least if ordinary people do.[104] If people have the right to free expression, so, perhaps, do institutions, which in turn implies the broader claim that they have rights. But such a reading is not warranted by anything argued here so far. It depends, rather, on an additional premise: that the right for X, in this case free expression, follows from the capacity for X. But rights may plausibly depend, not only on doing what the rights protect, but on whose rights they are. It may matter, for example, that *humans* are the ones claiming the rights we call "human rights." On some views of rights, there is something special or value-worthy in being human that grounds or inspires protecting certain human interests with rights.[105] For example, human beings are thought to have a right against quick, forcible eviction from a place they take as their homes. But this right is not generally believed to be held by most animals, even though both humans and, say, mice have the ability to make homes. The view that human interests merit protection is based, at least in part, on the assumption that there is something about being human that adds value to those interests. Or, alternatively, it is based on an understanding of those interests as distinctly human, even if the interests of other beings may share the same objects: the *human* interest in a home, or in free expression, may be different from that of other beings. Similarly, the reason to respect such human interests may not generalize to others.

[104] The U.S. Supreme Court recently argued that corporations, for example, have the same right to free speech that individuals enjoy because corporations contribute to the public discussion and have valuable expertise or perspectives to express. *Citizens United v. Federal Elections Commission,* 558 U.S. 1, 27, 38-39 (2010).

[105] See, for example, Locke 1988, 330.

One obvious example of such a disconnect may be the interest in not being destroyed simply because one has become useless to others. Both humans and institutions can lose their instrumental value to others. But it is entirely uncontroversial to treat useless institutions as no longer worthy of existence. The League of Nations was popularly dissolved, once it seemed to have outlived its usefulness, and replaced more than a decade later by the United Nations. This dissolution involved no moral violation of the League, as far as I can tell, even though it could well be argued that the League had an institutional interest in going on.

One reason to doubt that institutional interests, as such, have the sort of value that human interests have is that institutions come in infinitely many forms. A bowling team, a fan club, and a street gang all share the capacity that countries and corporations have to take stances and genuinely maintain them. But the similarities end there. On the theory that rights are grounded partly in features of the claimants they protect, there can be no corresponding grounding in the case of institutions simpliciter, because they need share almost nothing in common.

This same observation, of course, leads in the other direction: *some* institutions may be sufficiently like human rights holders, in all relevant respects, so as to have rights of their own, too. Or else they may have other characteristics that render their interests worthy of protection. For example, it has been proposed that peoples or nations may have political rights, known as "group rights,"[106] such as a right of self-determination. Similarly, peoples in the same sense have been described as having rights against annihilation. Genocide is considered wrong over and above the killing of all individuals in the targeted group. It may similarly appear that peoples as such have the right to free expression. Nothing that has been argued here counts against that claim. But if peoples have the right to free speech, among others, it may come from

[106] Waldron 1993, 364.

characteristics of peoples that other institutions do not share. It may, alternatively (or additionally), stem from the rights of individuals to organize and express themselves as peoples, peoplehood itself being a form of expression. But if such a right exists for groups, there is no reason to infer it can be held by other institutions, as well. Indeed, the differences among institutions, coupled with the connections – if such there be – from rights to the salient characteristics of their claimants, suggests that the basis for institutional speech rights may vary from case to case.

This last point about institutional rights, then, tracks a larger question that runs through this entire discussion: why care about institutional sincerity? Recall the argument presented in this chapter: an institution can sincerely take a stance if it meets the necessary and sufficient individuating conditions of ordinary individuals doing the same. That includes, primarily, being disposed to act consistently with the stance and to take steps to ensure that this disposition persists. Moreover, as I argued at length in section II, there is no reason to think that the psychological states of particular people are necessary to realize these stances, genuinely or sincerely, at the institutional level. So institutions can sincerely take stances none of its participants believe in. If, however, institutional stance-taking is a phenomenon independent of the views and attitudes of particular people involved, then why should we – as people – care about them all that much? If an institution expresses an insulting view of someone, but all the participating decisionmakers are known not to harbor it, then even if – as argued here -- the institution can be sincerely insulting, why should anyone *feel* insulted? Why should anyone care?

As already noted, if institutions are objectively valuable, they do not necessarily have this value merely *as* institutions: it takes too little to be an institution, and institutions take too many forms. Some may be valuable, but not all of them. One reason that some do have value – to

people, at least – is that they carry out projects we value but cannot fulfill on our own.[107] International fora like the UN serve as neutral ground for state actors to translate conflict into dialogue, argument and negotiation on common terms, and possibly resolve conflict. As such, the UN as institution best serves its purpose to the extent it is neutral, tolerant, friendly to dialogue and as apolitical as possible. These stances, in other words, matter. But no particular person or party can fulfill them, as people by definition belong to the partial perspective we want the UN to transcend. Similarly, governments are the exclusive facilitators of many people's security and status as equal citizens in a polity. They are therefore necessary to accord people a certain kind of respect that comes from treating them fairly, equally and as final reviewers, which fellow citizens cannot adequately bestow in the same way. Institutions also wrong people in ways that matter for moral repair, even when no individuals participated in the wrong. And these are just some of the ways institutional stances matter even when those of their decisionmakers do not.

Conclusion

Perhaps it bears restating that institutional stances already do seem to matter to people, as the examples at the start of this chapter illustrate. That could, of course, be an association of institutional stances with those of peoples, leaders or key human players. But there is no reason to assume that. Instead, people may simply care about the stances institutions take quite apart from any people involved. If so, the arguments advanced here offer some basis for identifying, assessing and morally scrutinizing those stances – and for the apparently widespread assumption that they are real.

[107] See Shiffrin 2009; Herman 2007, esp. 322.

In particular, the core argument proceeded in two steps. The first was to clarify what sincere forgiveness or repentance and similar states essentially require for individual humans, and the second was to show that institutions can meet those requirements merely in virtue of being able to act for certain ends. More specifically, I argued that sincere stance-taking consisted in being disposed to act for the right reasons, and that doing so is a matter of being disposed to act for the right ends. Institutions can, in fact, meet those conditions.

This analysis may invite the charge of behavioristic reduction, in that the sincere stance-taking of institutions – and individuals -- have been reduced to dispositions to act. But I reject the "reductionist" charge, because the essential requirements of sincere forgiveness, say, or repentance or gratitude do not exhaust what may necessarily coexist with how these requirements are met in particular cases, by particular bodies. It may not be necessary to sincere repentance that there be sadness or remorse, but human beings who do meet the requirements of sincere repentance – being disposed to act a certain way towards their victims – may, as a consequence, feel a deep sense of remorse. Indeed, it may be impossible for human beings to adopt and act on the right ends without a range of pangs and feelings, even if these are not conceptually necessary to the type of sincerity in question.

But it bears emphasis that the same can be said of institutions. As I began to suggest at the end of Section II, institutional sincerity in, say, forgiveness may involve events and activities that exploit the unique advantages of institutions, as contrasted with people. For example, institutions can be more consistent, over many more instances, in acting for the ends that sincerity requires. Their activities are also necessarily public, in ways that, arguably, amplify their effectiveness. If anyone doubts the sincerity of the institutional stance, they need only consult the record, or visit any relevant monuments or public symbols, or review the recorded

ceremony that could have constituted part of acting for the end. Indeed, with the increase in political apologies and reparations programs, we may soon witness the rise of novel and heretofore unimagined forms of sincerity, drawing on the unique resources of mindless, impersonal institutions.

CONCLUSION

The dissertation began with an air of mystery. Speech acts like apology, forgiveness and others seemed to do impossible things, giving rise to a sense of puzzlement. How could apologies act against past wrongs? How could they make a difference when they have nothing to reveal and no tangible benefit to bestow? How could forgiveness undo a moral trespass – how could it change the past or the moral status of a wrongdoer, absent any changes on his part? And how can a college or a company manage somehow to apologize, forgive and reconcile when it has no thoughts or feelings to reveal or invest in such expressions?

While I accept that these are difficult questions, I believe the sense of magic they evoke comes from a mistaken, if standard, picture of the utterances in question. These speech acts – such as "I am sorry," "I forgive you," or "We recognize your right" – are equated with expressions of a speaker's occurrent state of mind. In that form, they seem to have a transcendent power, suggesting that sharing one's feelings or beliefs with another can counteract the impact of genuine, concrete violations and harms. That standard picture also suggests that the very idea of an institution sincerely apologizing or forgiving is a fiction, albeit a useful and perhaps legally significant one.

The power of these expressions becomes more plausible and understandable, however, when we begin to see them less as revelations about oneself and more as ways of treating others, much like the non-communicative actions at which they are directed. How, on this alternative picture, do apologies respond remedially to past wrongs? I argued that they do so by treating one's victim as someone to whom the apologizer owes a debt he cannot repay, in effect owing

her not to have violated her as he did. This treatment ends the mistreatment of the victim otherwise put in place by wronging her and doing nothing about it – which takes her to be violable, and himself free to wrong her in the way he did.

How does forgiveness relieve wrongdoers of moral debt, like their duty to apologize and make restitution? It does so, I argued, by actually enabling them to fulfill those duties after all, but in a different form: in forgiveness, a victim effectively acts on his right to transfer restitutionary goods back to his offender, and enables his offender to treat him as a moral debtor, much as apologizing does, as well.

Still, I tried not to downplay the communicative component of these acts. In particular, the treatment put in place by apology, forgiveness and the like depends heavily on their being commissive speech acts – verbal commitments, to their audiences, to continue and maintain a certain way of relating to them. But unlike other commissives – promises and consent, for example – speech acts like apologies are themselves ways of acting on the treatment to which they commit. These sorts of utterances – which both constitute and commit to ways of treating their audience – I called *stance-takings,* and their double-duty as action and as verbal commitment is essential to the moral differences they make.

Not only did I seek to play up the active role of the speech involved in stance-takings, but I also tried to bring out the communicative role of morally significant action, in its own right. Wronging someone, or even merely harming her blamelessly, and then moving on without attempts at redress amounts to more than harmful sets of behaviors. Together, these actions also become a meaningful way of treating people, one that I called "objectively insulting" – treating them in a manner to which they could rightly take offense (even if they do not). When we harm others blamelessly, for example, and then move on as though nothing has happened, we treat

them as people it is acceptable to harm. It is this treatment that speech acts like apology and forgiveness help mitigate, and even reverse.

The communicative role of action also emerged in the discussion of institutional stances. Even without words, an institution that compensates victims for the end of repairing a past injustice thereby acknowledges the injustice, along with the victim's right to recompense for it. The ability to act for ends, such as that of treating people equally or respecting an aggrieved minority, enables institutions to act on principles they need not express or even believe. It enables them to enter stances. As a result, when they take stances verbally – apologizing publicly, forgiving and reconciling with other institutions, or taking responsibility, for example – they can do so sincerely.

The move away from an expressive picture may appear to deprive these utterances of a certain mystique. But viewing them as stance-takings, instead, also opens up possibilities that may seem equally magical, if at the same time within reach: through taking stances, we can restore relationships and transcend the momentary ups and downs of our affective attitudes about each other and our past differences. We can decide how we want to relate to another and, through verbal commitment, put that relationship in place immediately, to dramatic moral effect. And institutions can do even more: in apologizing, repenting and reconciling, they can take advantage of their infinite records, publicity and firm policymaking procedures, so as to heal relationships in ways that far surpass the capacities of individual human beings. With the recent explosion in public apologies and historic reconciliation between countries and corporations, we have only just begun to see how this power might play out.

BIBLIOGRAPHY

Anderson, Elizabeth and Richard H. Pildes. 2000. "Expressive Theories of Law: A General Restatement." *University of Pennsylvania Law Review* 148: 1503-75.

Arendt, Hannah. 1973. *The Origins of Totalitarianism*. Harcourt Brace.

Austin, J. L. 1962. *How to Do Things with Words*. Oxford University Press.

Bovens, Luc. 2008. "Apologies." *Proceedings of the Aristotelian Society* Vol. CVIII:3: 220-239.

Bratman, Michael. 1993. "Shared Intention." *Ethics* 104 (1): 97-113.

Dan-Cohen, Meir. 2007. "Revising the Past: On the Metaphysics of Repentance, Forgiveness, and Pardon." In Sarat and Hussein, eds. *Forgiveness, Mercy, and Clemency.* Stanford University Press.

Danto, Arthur. 1968. "Basic Actions," in A.R. White, ed., *The Philosophy of Action*. Oxford University Press.

Driver, Julia. 1997. "The Ethics of Intervention*." Philosophy and Phenomenological Research* LVII, 4: 851-70.

Dworkin, Ronald. 1985. *A Matter of Principle*. Harvard University Press.

Frankfurt, Harry. 2008. "Inadvertence and Moral Responsibility." *The Amherst Lecture in Philosophy*. Via web at www.amherstlecture.org.

French, Peter A. 2005. "Corporate Moral Agency (revised)." *The Blackwell Encyclopedia of Management: Business Ethics,* Volume II Second Edition. Blackwell.

Gardner, John. 2001. "Obligations and Outcomes in the Law of Torts." In *Relating To Responsibility: Essays For Tony Honore*. Hart Publishing, London, England.

Gilbert, Margaret. 2004. "Scanlon on Promissory Obligation: The Problem of Promisees' Rights." Journal of Philosophy, 101 (2): 83-109.

Gilbert, Margaret. 2002. "Collective Guilt and Collective Guilt Feelings." *Journal of Ethics*, Col. 6 (2) : 115-143.

Grice, Paul. 1957. "Meaning." *The Philosophical Review*, Vol. 66 (3): 377-388.

Griswold, Charles L. 2007. Forgiveness: A Philosophical Exploration, New York: Cambridge University Press.

Hampton, Jean. "Forgiveness, Resentment, and Hatred," in Forgiveness and Mercy, Hampton, Jean, and Jeffrie G. Murphy (eds.), Cambridge: Cambridge University Press, 1998, pp. 35–87

Helmreich, Jeffrey S. 2012. "Putting Down: Expressive Subordination and Equal Protection." *U.C.L.A. Law Review Discourse* 59: 112-27.

Herman, Barbara. 2007. *Moral Literacy*. Harvard University Press.

Herman, Barbara. 1993. "What Happens to the Consequences?" In *The Practice of Moral Judgment*. Harvard University Press: 94-112.

Hieronymi, Pamela. 2001. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62 (3): 529-555.

Kornhauser, Lewis and Lawrence Sager. 1986. "Unpacking the Court." *Yale Law Journal* 96: 82–117.

Kutz, Christopher. 2000. *Complicity: Ethics and Law for a Collective Age.* Cambridge University Press.

Locke, John. 1988. *Two Treatises of Government*. P. Laslett, ed. Cambridge University Press.

Maimonides, Moses. 1987. "Laws of Repentance." In Touger, ed., *Mishneh Torah*. Moznaim Press, Brooklyn, New York.

Martin, Adrienne. 2010. "Owning up and Lowering Down: The Power of Apology." *Journal of Philosophy* 107 (10): 534-553.

Morris, Herbert. 1988. "Nonmoral Guilt." In Ferdinand Schoeman, Ed. *Responsibility, Character and The Emotions: New Essays in Moral Psychology*. Cambridge University Press. 220-240.

Murphy, Jeffrie and Jean Hampton. 1988. *Forgiveness and Mercy*. Cambridge University Press.

Nagel, Thomas. 1979. "Moral Luck." In *Mortal Questions*. Cambridge University Press. 24-38.

Novitz, David. 1998. "Forgiveness and Self-Respect." *Philosophy and Phenomenological Research* 58 (2): 299–315.

Nusseibeh, Sari. 2007. *Once Upon a Country.* Douglas &McIntyre Ltd.7

Owens, David. 2012. *Shaping the Normative Landscape*. Oxford University Press.

Pettit, Philip. 2003. "Groups with Minds of Their Own." In F. Schmitt, Ed., *Socializing Metaphysics: the Nature of Social Reality.* Rowman and Littlefield: 167-94.

Rees, D.A. 1955. *Aristotle: The Nicomachean Ethics, a Commentary*. Clarendon Press.

Richards, Norvin. 1988. "Forgiveness." *Ethics* 99: 77–97.

Roth, Philip. 2000. *The Human Stain*. Houghton Mifflin Company.

Ryle, Gilbert. 1949. *The Concept of Mind*. University of Chicago Press.

Scanlon, T. M. 2008. Moral Dimensions: Permissibility, Meaning, Blame. Harvard University Press.

Searle, John R. 2010. *Making the Social World: the Structure of Human Civilization.* Oxford University Press.

Searle, John R. 1979. "A Taxonomy of Illocutionary Acts," in Searle, *Expression and Meaning*. Cambridge University Press: 1-30.

Shiffrin, Seana Valentine. 2010. "Incentives, Motives and Talents." *Philosophy and Public Affairs* 38: 111-42.

Shiffrin, Seana Valentine. 2009. "Reparations for U.S. Slavery and Justice over Time." David Waserman and Melinda Roberts, Eds. *Harming Future Persons.* Springer. 336-37.

Shiffrin, Seana Valentine. 2008. "Promising, Intimate Relationships and Conventionalism." 117 *Philosophical Review*. 481-524.

Shiffrin, Seana Valentine. 2002. "Caution about Character Ideals and Capital Punishment; A Reply to Sorell." Criminal Justice Ethics. 35-39.

Smith, Adam. 2002. Haaokenssen, ed. *The Theory of the Moral Sentiments*. Cambridge University Press.

Smith, Nick. 2008. *I was Wrong: the Meanings of Apologies*. Cambridge University Press.

Sunstein, Cass. 1996. *Legal Reasoning and Political Conflict.* Oxford University Press.

Tannenbaum, Julie. 2007. "Emotional Expressions of Moral Value." *Philosophical Studies* 132: 43-57.

Thomson, Judith Jarvis. 1992. The Realm of Rights. Harvard University Press.

Velleman, James David. 2006. "Don't Worry, Feel Guilty." In *Self to Self: Selected Essays*. Cambridge University Press. 156-69.

Velleman, James David. 1997. "How to Share an Intention." *Philosophy and Phenomenological Research* LVII (1): 29-30

Waldron, Jeremy. 1993. *Liberal Rights.* Cambridge University Press.

Weinrib, Ernest. 1994. "The Gains and Losses of Corrective Justice." *The Duke Law Journal*. Vol. 44: 277-97.

Williams, Bernard. 1982. "Moral Luck." In *Moral Luck.* Cambridge University Press. 20-39.