Matters of Trust

By

Kelly Ann Cunningham

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Philosophy

August 9th. 2024

Nashville, TN

Approved:

Diana B. Heney, PhD

Robert B. Talisse, PhD

Matthew Congdon, PhD

Jason D'Cruz, PhD

To my grandfather, Bill Cunningham

Acknowledgements

"So many names, there is barely room on the walls of the heart" -Billy Collins

And since there are so many names, and so little room, I'll omit them—you'll know who you are. First and foremost, I'd like to thank the sum of my efficient causes—my parents, grandparents, and great-grandparents and so on. Without them, I (literally) wouldn't be here, nor would I have had the self-trust they instilled in me. I also want to express thanks to my sister, and cousin for believing in me and for making sure that I don't take myself too seriously, and to my aunt for her unwavering support and unfiltered advice.

Next up, my advisor and committee chair. I left each of our meetings with my fears assuaged and my ideas clarified. She is deserving of thanks for so much more, but critically for telling me when to stop reading and start writing and for passing on her mentor's advice that the best way to be a woman in philosophy is to simply *be* a woman in philosophy. Now would is also a fitting time to thank all the women in the cohorts before and after me, for modelling the many ways it is possible to do so without compromising their values or personality. More thanks still is owed to my only cohort-mate to graduate with me for sharing five years' worth of drafts, conversations, and spilt tea. Within my department, I also want to thank one of the only people who can help me understand Hegel for her encouragement, rigorous and exciting seminars, and advice. A fellow Murdochian, for introducing me to her work, catching my typos, and for transformative conversations about ethics. And the third faculty member on my committee for his leadership within our department and for setting high standards when it comes to arguments and writing. Acknowledgements for those within my department would not be complete without mention of our Spinozist, for reminding me I still have what it takes to do metaphysics, and our resident Eastern philosophy expert for conversations that kept my interest in Chinese philosophy alive. Beyond my department, I am grateful to my external reader for his thoughtful commentary on my earliest work on trust and for his own work which has deeply informed my views. Also, the professor still at DePaul, who is responsible for getting me into all this philosophy business in the first place. I also would like to thank the director and staff at the Curb Center for supporting the efforts of Open Book Collective and my mentorship pilot co-pilot for her dedication to FGLI community we've grown together.

Beyond Vanderbilt. I also owe thanks to many of the staff and volunteers at The Nashville Food Project and to my former co-workers at Trader Joe's Store #667. This list would not be complete without mentions of the wonderful friends I've had the pleasure of keeping in my communities back home in East Dundee and Chicago, and of making in my communities here in Nashville. I've gotten by, tried, and the rest with more than a little the help from my friends, and I am lucky to have known and been known by many of you through many stages of life. I would also especially like to acknowledge the friends I've made through union organizing for keeping alive my hope in the possibility of a better future during challenging times. .

Finally, I want to thank three others who've turned a house into a home. My partner and comrade for his encouragement and for helping me keep things in perspective. Our small black cat for showing me it is possible to earn an animal's trust. And, last but certainly not least a certain sneaky pup for being by my side every step of the way. If it had not been for our many walks together, I doubt that many of the ideas in this dissertation would have made it to the page.

Table of Contents

"Some degree of trust in the social world is the starting point and very basis of morality."

-Annette Baier

Preface

I first became interested in the concept of trust while writing a term paper in which I argued that managing the harms of political polarization requires cultivating a climate of trust. I found that employing the idea of "a climate of trust" was essential to understanding how trust operates in broader social contexts (i.e. in contexts where trust is not interpersonal). In a climate of distrust, I argue, attempts to foster moral respect and signal trustworthiness are liable to backfire. As I continued to develop this paper (a longer, heavily revised version of which is soon to be published), I came to appreciate the difference between a mere lack of trust and the presence of distrust, as well as the extent to which both can underlie and exacerbate a variety of wicked social problems. I borrowed the idea of a "climate of trust" from Annette Baier's influential essay "Trust and Antitrust", which I would re-read a few months later (for the first of many times) alongside several other essays on trust in her 1994 book, *Moral Prejudices*, during the preparation for my field exam.

This extended encounter with Baier's work affirmed my earlier judgement that trust is central to both the applied and theoretical dimensions of ethics. I chose trust as a dissertation topic because of its ability to bridge these two sides of my research interests. As I began exploring more of the literature, I was shocked to find that although "Trust and Antitrust" was cited by nearly every philosopher with something to say about the topic, many of her most compelling arguments were outright ignored, often resulting in a distortion of Baier's critical engagement with the history of moral philosophy. It seemed to me that the issues with

philosophical explorations of trust that Baier identified were being repeated rather than rectified. The following excerpt summarizes the problem:

> The more we ignore dependency relations between those grossly unequal in power and ignore what cannot be spelled out in an explicit acknowledgement, the more readily we assume that everything that needs to be understood about trust and trustworthiness can be grasped by looking at the morality of contract (Baier 1986, 241).

Over the course of the next five chapters, I will show that the core problem identified here by Baier persists: existing philosophical accounts of trust do not sufficiently attend to how trust operates in relationships of asymmetrical power nor to the latent forms of trust that support its more explicit, contract-like forms. In response to that problem, I articulate an alternative account of trust that starts with, rather than ignores, dependency relations and the implicit agreements that emerge as individuals and the societies in which they live develop. In doing so, I demonstrate that the ethical significance of trust and trustworthiness can only be fully understood when we look at how trust contributes to the flourishing of individuals and their communities. My goal is to offer an account of trust that is well-suited for interventions to problems in the domain of applied ethics. In other words, my account of trust is designed to be useful for inquiry, aimed at the moral ideal of growth, and to recognize the power of both individual habits and enduring social practices.[1]

Before I begin, allow me to define a few keys terms I'll use throughout and provide a roadmap for each chapter.

[1] In this way, my account of trust is designed with a pragmatist methodology in mind. Jane Addams's work has been influential in shaping my understanding of each of the three items on this list. *Newer Ideals of Peace, Democracy and Social Ethics,* and *The Long Road of Woman's Memory* have been particularly formative in my understanding of pragmatist methodology, as well as my approach to ethical theorizing more broadly.

Key Terms:

One of the qualifiers I frequently use to describe my account of trust and the aspect of other accounts I draw inspiration from is "embodied". By using this word, I intend to draw attention to our body's constitutive vulnerability and to the fact that our bodies are both the product of and means by which we develop habits. Our corporeality renders us susceptible to being harmed by other people, as well as by other beings and objects found in our environment. Our bodies are always *in* and *a part of* reality, and at the same time our bodies also what we experience reality through. Differences in the appearances and abilities of our bodies have an impact on how we perceive the world and how we are perceived by others. Though there is a sense in which all experience is "embodied" (if we exclude reports of "out of body" experiences), by using this phrase, I aim to draw attention to our full range of learning modalities and the visceral ways in which they contribute to our judgements. This includes visual and auditory phenomena that we resort to describing with figurative language, tactical know-how, and multifaceted experiences that we can only describe by referencing that status of our whole body. For example, as we do when we are "feeling unsafe" or "on edge". Finally, although I will keep philosophical debates about what emotions are at arm's length, I do believe that our emotions have a physiological dimension that is particularly relevant to our experience of trust and distrust.[2] In sum, I want to draw attention to the way our biology mediates and sets the parameters for our interactions with our environment.

This emphasis on the body and lived experience also appears in my use of "embrace" in the definition of trust I offer in chapter four: *trust is a response to one's situation that embraces*

[2] See William James's essay "What is an Emotion?" for an account of the emotions that ties them directly to our physiological sensations. I do not endorse everything about his argument, but I do think he gets something basic about the nature and origin of our emotions right.

vulnerability. There are many kinds of embrace and ways in which one might embrace another—imagine the difference between the tight squeeze you might give a loved one after a long time apart from the side-hug you cautiously extend to an acquaintance who has just received bad news as you try to console them. This range of kinds of embrace and ways of embracing introduces some ambiguity to my definition, however because of its association with the body it resists collapsing trust into a purely cognitive or affective phenomenon. "Accepting" or "acknowledging", run the risk of such a collapse and can place limits on what sorts of entities can trust and be trusted.

It is important, also, to think of the other ways in which the term "embrace" is used—embracing one's flaws, embracing a new challenge, embracing a new culture –all imply something *more* than mere acceptance and carry with them a connection to courage and to acting openly or honestly. I touch upon the relationship between trust, honesty, and courage throughout the dissertation and plan to explore these connections in more depth in my future work. To embrace something is to welcome it. Acceptance or acknowledgement does not necessarily entail this optimistic or positive attitude. While this may seem a very minor point, it is essential for differentiating trust from both reliance and distrust. One may acknowledge vulnerability and still take steps to mitigate or eliminate it and such a mindset is antithetical to trust. So, it is the optimistic valence associated with this second use of embrace that that I have in mind when I articulate my definition of trust as involving an embrace of vulnerability.

Another term that requires clarification is "perception". My use of perception is tied to what I want to emphasize by using "embodied"—the fact that we gather information about reality through multiple sensory avenues that are unique (in the sense of being non-trivially physically distinct) to our individual physical forms. Perception is sometimes used as a success

term to imply that we correctly apprehended something about the world. We also use the term as praise in character evaluations. A perceptive person is one who notices or can identify the nuances of a given situation well. However, the way I intend to use it in the context of this dissertation is to refer to how what we take in from our environment through our senses impacts our judgement about our situation. What a person can perceive will depend partly on their habits and character as well as their talents and skills. Using perception in this way allows us to describe it with a rich array of adjectives: one's perception can be clear, but it can also be warped, distorted, clouded, and so forth.

Finally, a pair of terms it will be helpful to clarify from the outset are "doxastic" and "affective", which are the descriptors I use to classify different kinds of accounts of trust. There are several ways, and additional terms, I could have used to carve up this territory. One reason I settled on these two is that they are both broad enough to encapsulate several more specific accounts that, despite important differences among them, share key commitments. Accounts I refer to as "doxastic" share the key premise that trust is primarily a matter of belief. It follows from doxastic accounts that trust and distrust can be justified using the same epistemic norms that we have for belief. Though some of the accounts I group into this category include an affective component, they all define trust in terms of belief, which indicates that the belief-component ought to have priority. Affective accounts, on the other hand, define trust as an expression of an attitude or emotion. Because the affective component of trust is placed at the forefront, whether trust is justified does not depend on evidence in the same way that doxastic accounts suppose. Affective accounts do not deny that beliefs matter but insist that beliefs alone cannot fully explain what trust is or how it operates.

Roadmap:

Our starting point is the contemporary trust literature, and our destination is an account of trust that does not neglect dependency relations. The first three chapters set up the account I offer in chapter four and elaborate on in chapter five.

In chapter one, I situate my dissertation historically by describing how my account of trust builds on Annette Baier's work. Baier is a significant source of inspiration for the developmental account I offer for three reasons. First, her moral philosophy blends virtue-based and care-based approaches to ethics. This combination of theoretical commitments draws our attention to the body, particular relationships, and external circumstances—three variables that I believe have received insufficient attention by most extant accounts of trust. Second, Baier's moral philosophy strikes a balance between emphasis on the individual and emphasis on community that coheres with the pragmatist goals that guide my account of trust. In particular, Baier encourages us to think intergenerationally by considering the discrepancy of power between past and future generations, and the limits of our voluntary choice in our relationships. Third, and perhaps most obviously, this chapter has a corrective aim. Baier's work has not been given the kind of attention that it deserves. This chapter aims to give credit where credit is due and make the ways in which I see myself following in Baier's footsteps clear. Baier's work engages critically with the history of moral philosophy in a way that I believe is necessary to correct its sexist, racist, ableist, and anthropocentric tendencies. I strive to emulate her mode of engagement with the history of philosophy by contributing to the discipline in a way that disrupts and challenges the narratives that dominate the history of philosophy and the assumptions that those narratives often take for granted.

Chapter two establishes requirements for an adequate account of trust. Drawing on the existing literature, I propose that an account of trust should: 1) be able to account for distrust 2) make a meaningful distinction between trust and reliance, 3) attend to how trust operates in relationships with asymmetries of power. An account that can do these three things will be well-equipped for application to real-world problems. In chapter three, I assess doxastic and affective accounts of trust with respect to these three requirements. I find that although they meet the first and second requirements to varying degrees, both types of accounts fall short of meeting the third requirement. In this chapter, I also identify two pitfalls that give us additional reasons to doubt the approach to defining trust that affective and doxastic accounts share.

In chapter four, I return to the task of describing my alternative, Baier-inspired, developmental account of trust. Here, I explain the similarities and differences between my account and the ordinary language accounts that are currently the primary alternatives to the more prominent affective and doxastic accounts examined in chapter three. In addition to defining trust as a response to one's situation that embraces vulnerability, I explore three ways in which trust contributes to human flourishing: by facilitating cooperation, by enabling and sustaining intimate interpersonal relationships, and by playing a constitutive role in the formation of our personal identity. In chapter five I offer further support for why I believe trust contributes to human flourishing by elaborating on the developmental account of trust and by indicating how insights from the trauma studies literature can benefit further explorations of trust and distrust. The fifth chapter is more promissory than the first four, as it indicates the directions in which I plan to take my future research.

Chapter One: Revitalizing Baier

"Trust is a fragile plant, which may not endure inspection of its roots, even when they were, before the inspection, quite healthy" –Annette Baier

Contemporary accounts of trust acknowledge Baier as a foundational, influential figure in the philosophical work on trust. Although her 1986 essay "Trust and Antitrust" has earned its place as required reading for any scholar interested in trust, what is typically quoted and emphasized from this essay is not representative of her work as a whole.[3] Therefore, to ensure that the ways my account expands on Baier's are clear, I will offer a developmental interpretation of her work on trust—one that foregrounds the role of perception and the importance of embodied experience. I argue that most interpretations of Baier focus too narrowly on two passages from "Trust and Antitrust" and that in doing so, they obscure the general understanding of Baier's account and fail to appreciate its nuances. As a result, many of the available theories of trust reproduce a bias in favor of ethical theories based on contract or universal moral principles, which is precisely what Baier thought an account of trust ought to avoid.

I describe my interpretation of Baier and my account of trust as developmental for two reasons. First, Baier is clear that some varieties of trust arise out of or depend on other, more

[3] "Trust and Antitrust" was first published in *Ethics* in 1986 but was later reprinted in Baier's 1994 book *Moral Prejudices* alongside several other essays about trust. Most citations of the essay use the 1986 version, but because I also reference the other essays in *Moral Prejudices* in this chapter, the page numbers I use correspond to the essay as it appears in the 1994 reprint. There are three non-consequential differences between the two versions. The first is that in the 1986 *Ethics* version, the introductory section has the subheading "Trust and Its Varieties", whereas in the 1994 *Moral Prejudices* version, the subheading is omitted. The second is that a line in the 1984 version reads "Infant trust that normally does not need to be won…" (Baier 1986, 242) and in the 1994 version the same line reads "That infant trust normally does not need to be won…" (Baier 1994, 107). The third and final difference is that in the 1986 version Baier capitalizes the first letters in "Prisoner's Dilemma" (Baier 1986, 232,252) and in the 1994 version they are in lower case (Baier 1994, 96, 118).

basic forms. Although Baier explores many different varieties of trust, she insists that they all share a common element, which she frequently describes as "risk" or "acceptance of vulnerability". This implies that the various forms of trust can be traced back to a shared root.[4] Second, her understanding of interpersonal trust corresponds to human biological development, and her understanding of social and political trust is historically and culturally informed (and thus responsive to the development of societies, communities, and states). In the process of offering this developmental interpretation I will pull into focus the commitments and inspiration she takes from both care and virtue ethics.

In the first section of this chapter, I revisit "Trust and Antitrust" with an eye toward contextualizing the most cited passages from the essay and substantiating my claim that it should be read as a collection of Baier's earliest reflections on trust as opposed to her definitive account. I do this first by showing that what *is* typically cited from the essay largely misses Baier's point and second by drawing attention to her arguments about trust that are typically *not* cited. In the second section, I demonstrate that Baier continues to build on these neglected ideas and arguments throughout both "Trust and Antitrust" and in the other essays on trust in *Moral Prejudices*. My aim is to show which of the ideas in "Trust and Antitrust" remain central to Baier's account as it is fleshed out in her later work. In the third section, I jump forward to Baier's penultimate book: *Reflections on How We Live* (2009). The two essays I focus on in this section offer an illustration of Baier's mature account of trust, wherein we can see how she builds on and departs from her initial reflections in "Trust and Antitrust". I bring the chapter to a

[4] Baier is not strongly committed to any one description of this shared element. She also refers to it as a "special vulnerability", and as a "special danger" (Baier 1994, 104). Later she states that different types of trust have the same "main dimensions of fragility" (Baier 1994, 137). She also states that the discretionary power relinquished by the one-trusting and possessed by the one-trusted is what introduces this shared element (Baier 1994, 104).

close by making explicit the ways in which my developmental account of trust is an expansion of her work.

Revisiting "Trust and Antitrust"

Prominent accounts of trust have largely been inspired by or framed as a response to Baier's work. In addition to being recognized as the first philosopher to seriously theorize trust, Baier has been cited by those interested in ethical and epistemological questions about trust (Pettit 1995, Faulkner 2007, D'Cruz 2020, 2015, Simpson 2012) as well as those concerned with trust's more practical side (Govier 1992, McLeod 2002). However, the majority of these references to her work can be traced back to two claims she makes in the opening pages of "Trust and Antitrust": 1) her observation that betrayal is a distinctive emotional response to failed trust and 2) the observation that trust is a three-place predicate.[5] Although these two passages have inspired now widely accepted assumptions found in prominent accounts of trust, neither of them is representative of Baier's central argument in the essay.[6] Both are found in the introductory section, where Baier is only beginning to reflect on the ways in which we trust. Furthermore, unlike other observations she makes in the same section, neither of these points are ones that she returns to later in the essay or in her work beyond it.

Contextualizing these commonly cited remarks by looking at their placement and function within the essay will reveal how the emphasis on these two passages obscures her

[5] In the literature I have seen "three-place predicate", "three-part relation", and "entrusting" all used by different philosophers to describe the same passage. Baier herself uses "three-place predicate" and "entrusting" to describe the oft-cited framework. The exact quote will be made clear later, but what I am referring to here is the *A trusts B with respect to C* formula that has become standard for modeling trust. I use the terms interchangeably.

[6] According to PhilPapers, "Trust and Antitrust" has been cited 524 times as of April 12th, 2024. Influential citations that build on the trust-betrayal connection include Kriton (2020, 584), Mullin (2015, 317), Hawley (2014, 2), Holton (1994,), Faulkner (2018 3; 2014, 191), McGeer (2006, 240), Jones (1996. 2019) and Simpson (2012, 552). Those who persist in modelling trust on the entrusting framework include the accounts offered by Mullins (2015), Hawley (2014; 2017), Pettit (1995, 217) and Bennett (2021). Karen Jones follows Baier in seeing and making explicit the limits of the entrusting model (Jones 1996, 17; 2019, 957).

argument in the essay and will also serve to bring her less frequently cited contributions into relief. One of them is her use of the phrase "climate of trust". Another is the connection she makes between trust and the trajectory of human's biological and evolutionary development. The third is the attention she pays to the role of relative power and voluntary abilities, which is largely due to the way her account is informed by the concerns raised by the tradition of care ethics that was emerging during this time. What becomes clear when "Trust and Antitrust" is read with these three points in focus is that Baier's goal in the essay is to show that the difficulties that dominate moral philosophies have when it comes to accounting for trust can be explained by the underrepresentation of women in the discipline. For this reason, Baier argues, moral philosophers should continue to explore trust while keeping in mind the blind spots of prominent moral theories.

Baier begins "Trust and Antitrust" by reflecting on the Sissela Bok quote that serves as the essay's epigraph: "*Whatever* matters to human beings, trust is the atmosphere in which it thrives" (Baier 1994, 95, emphasis in original). She states that there is something "basically right" about Bok's claim but adds that immoral and undesirable things can also thrive in atmospheres of trust (Baier 19994, 95). A few paragraphs later, when describing the many unnoticed ways in which we trust others, Baier uses the imagery again to argue that "we inhabit a climate of trust as we inhabit an atmosphere and notice it as we notice air, only when it becomes scarce or polluted" (Baier 1994, 98). These opening lines are significant because the image of trust as an atmosphere or climate is essential to Baier's understanding of how trust operates throughout her work. Baier also returns to the idea that the most basic forms of trust are those that are latent. So, from the outset Baier is clear that trust enables the thriving of what matters in our lives, so much so that it is compared to the air we must breathe to live. Those who cite this

essay typically ignore this opening context where it is clear that Baier takes trust to be a fundamentally moral concept, despite the fact that trust is sometimes abused.

Baier then proceeds to frame her essay as a "start on the large task" of distinguishing among different forms of trust, identifying trust's morally relevant features, and determining when trust should be preserved (or destroyed) (Baier 1994, 95-96). Baier is explicit that her inquiry into trust is something *new*: "It is a start, *not a continuation*, because there has been a strange silence on the topic [of trust] in the tradition of moral philosophy" (Baier 1994, 96, emphasis my own). She finds the philosophical silence strange because despite sharing an interest in cooperation (which requires trust) moral philosophers have yet to offer "even a sketch of a moral theory of trust" (Baier 1994, 96-97).[7] The forms of trust that canonical figures in the history of Western philosophy *have* given attention to are trust in God, trust in governments and officials, and "what might be called obsessive trust in contracts and contractors" (Baier 1994, 97). She then states that this "selective attention" in theorizing trust is what she aims "to remark [on], tentatively explain, and *try to terminate* (Baier 1994, 97, emphasis my own).

In the process of describing several cases in which we trust one another, Baier raises the question of what makes trusting others different from merely relying on them. Here is where she makes the claim that proponents of affective accounts often draw on, writing: "we can still rely where we no longer trust" and "trusting can be betrayed, or at least let down, and not just disappointed" (Baier 1994, 98-99).

The contrast between trust and reliance is an important moment in the text and I find Baier's interest in what is revealed about trust by our emotional responses to failures of it compelling. However, accounts that present this distinction as the defining characteristic of trust

[7] She mentions a particular frustration with Aristotle who, despite condemning tyrants for sowing seeds of distrust, omits a discussion of trust from his account of wisdom (Baier 1994, 97).

overlook the function of this passage with respect to the essay as a whole. Baier uses the difference between the emotional responses that follow from a failure of trust and a failure of reliance *not* to define trust, but rather to introduce an element that all varieties of trust share: acceptance of vulnerability. This is evident not only because she allows for some kinds of trust to be mixed with reliance, but also because her discussion of this difference is immediately followed by her first definition of trust: "Trust then, on this first approximation is accepted vulnerability to another's possible, but not expected ill will" (Baier 1994, 99). This first definition focuses on how trust exposes one to a distinct kind of vulnerability and not on the emotional response associated with failed trust. Furthermore, Baier repeatedly returns to and expands on the idea that vulnerability and risk characterize trust in later sections of the essay. Therefore, on my developmental account, Baier's discussion of betrayal is interpreted as a clue about what makes trust meaningfully distinct from reliance and not the full answer.

In the page and half that separates this discussion from her introduction of the oft-cited entrusting framework, Baier insists that to understand trust we need to emphasize the vulnerability of the person doing the trusting rather than the potential benefits or risks that accompany the act of trusting. She observes that we often make ourselves and the things we care about vulnerable out of necessity: "we need help in creating and then in not *merely guarding* but *looking after* the things we most value" (Baier 1994, 100, emphasis my own). The distinction between guarding and looking after is one place where we can see how care ethics informs Baier's understanding of trust. Allowing someone to "look after" a person or other entity that we value involves a higher degree of intimacy and likely physical proximity, both of which invoke

the notion of caring.[8] She makes this implicit connection to care explicit a few lines later when she writes:

> The simple Socratic truth that no person is self-sufficient gets elaborated once we add the equally Socratic truth that the human soul's activity is *caring for* things into the richer truth that no one is able by herself to look after everything she wants to have looked after, nor even alone to look after her own "private" goods, such as health and bodily safety (Baier 1994, 101, emphasis in text)

Baier's point is that the lack of attention to our interdependency in the history of moral philosophy is a part of the reason that philosophically rich accounts of trust were lacking at the time she was writing.

A few sentences later, Baier introduces the formula for the entrusting framework: *A trusts B with valued thing C* (Baier 1994. 101). This modelling of trust as a three-place predicate is taken for granted by many trust theorists who attribute it to Baier, and yet, Baier does not directly take credit for the model herself.[9] Immediately before introducing it, she is clear that in trying to distinguish among different forms of trust by looking at the differently valued goods we entrust she is following Locke "in analyzing trusting on a model of *en*trusting" (Baier 1994, 101 emphasis in original). Though she briefly uses the model in this section and suggests that "it will prove more of a help than a hindrance", she is upfront about its limitations. She notes that using this framework "will involve some distortion and regimentation of some cases, where we may have to strain to discern any definite candidate for C" (Baier 1994, 101). Furthermore, a few pages later, she makes the reasons she is drawn to the model explicit:

> To entrust is intentionally and usually formally to hand over the care of something to someone, but trusting is rarely begun by making up one's mind about trust, and it often has no definite initiation of any sort but grows up slowly and imperceptibly. What I have tried to take from the notion of entrusting is **not**

[8] For example, think of the difference in the kind of relationship between the knight who *guards* the princess in the tower and the kind of relationship between the princess and the handmaiden who *looks after* her.
[9] In a footnote in her later essay "Sympathy and Self-Trust", Baier remarks on how she "tended to drop this construing of trusting as entrusting" and credits Karen Jones for observing this shift in her conceptualization.

> **its voluntarist and formalist character** but rather the possible specificity and restrictedness of *what* is entrusted, along with the discretion the trustee has in looking after that thing. Trust can come with no beginnings, with gradual as well as sudden beginnings, and with various degrees of self-consciousness, voluntariness, and expressness. (Baier 1994, 105, emphasis in italics is in original, emphasis in bold is my own)

From the excerpt above and the way Baier introduces the entrusting framework I draw three conclusions: (1) though Baier is the first to present the entrusting framework in a tidy and helpful way, she does not claim to be not the original source of this idea, (2) Baier does not think that trust can always be modelled on this three-part relation, and (3) she believes the tendency to think of trusting as entrusting seen throughout the history of philosophy is at least partly responsible for the inattention to essential features of trust. It is my hypothesis that continued employment of the entrusting framework has reproduced the problem Baier identifies—the inability to account for trust in relationships with asymmetries of power—in contemporary accounts of trust. Furthermore, in this passage Baier foreshadows her own account of trust by noting features of it that she will elaborate on in the next four sections of the essay.

Baier ends the opening section by indicating that her remarks up until this point have been about the forms of trust where the one-trusting is aware of their trust and has conscious control over whether they want to continue trusting. In such cases, both the one-trusting and the one-trusted are aware of what discretionary powers are being entrusted and the one-trusting expects the one-trusted "to use their discretionary powers competently and nonmaliciously" (Baier 1994, 104-105). She transitions to the next section by concluding that while what we trust one another with is one variable that affects trust relations, another "crucial variable…is the relative power of the truster and the trusted" (Baier 1994, 105). Without understanding how trust alters positions of power (i.e. how the position of being in the trust relation is different from the

position one is in when the trust relation is absent) we cannot properly judge when a "form of trust is sensible and morally decent" (Baier 1994, 106).

The second section of the essay, 'Trust and Relative Power', shifts away from focusing on dimensions of trust that are most apparent in trust relations where the two parties are relatively equal and toward those where power differences are most extreme. Baier describes approaches that take equal power relations as the paradigm for trust as a "myopia" in moral philosophy. She attributes moral philosophers' lack of attempts to theorize trust in relationships of asymmetrical power to the prevalence of contract-based moral theories and aims to reposition "where trust in contracts fits into the picture we get" on an account which takes trust to be essential to the thriving of morality (Baier 1994,106). It is clear by the time Baier transitions into the next section that if our account of trust is to capture the full range of trust relationships, contracts (and trust relations that share such features) must be moved to the periphery. To initiate this shift, she considers two alternative relationships with extreme power asymmetries—infant trust and trust in God— both of which she sees as relationships of total dependence.

Despite noting that both trust in God and infant trust are cases where the one-trusting is a state of total dependence, she finds infant trust a more promising form of trust to adopt as a starting point. A major reason she favors starting with infant trust is because coheres well with what is easily observable about human development and biology. Baier writes that "some degree of innate, if selective, trust seems a necessary element in any surviving creature whose first nourishment (if it is not exposed) comes from another, and this innate but fragile trust could serve as the explanation both of the possibility of other forms of trust and of their fragility" (Baier 1994, 107). In this passage, Baier suggests again that many forms of trust depend on this more basic form. While she does not commit to the argument that infant trust is the *most* basic

form of trust, passages later in the essay confirm that infant trust plays an explanatory role in her account. Without this form of trust to "pave the way", she argues, "it would appear a miracle that trust ever occurs" (Baier 1994, 107).

Aside from its connection to biological development and its explanatory potential, Baier finds the key difference between infant trust and trust in God to be another reason to favor the former. Though infants begin powerless and completely dependent on their caregivers, their power and degree of dependency changes over time. Trust in God, unlike infant trust, can never transform into mutual trust. Instead, the trust remains one-sided and static because God (at least an omniscient and omnipotent one) is, unlike a parent, invulnerable. The kind of vulnerability we expose ourselves to when we trust in God is simply too far-removed from the kind of vulnerability we are exposed to when we trust in any human relationship—even those where our degree of dependency reaches its peak. Baier takes infant trust as the starting point because, relative to the two extremes ends of dependency relations seen in trust placed in God (total dependency) and trust placed in contracts (equal or neutralized dependency), it occupies a middle ground that, statistically, captures the kind of dependency that characterizes most of our relations with others; that is, dependency that is mutual and in flux. Even though infants eventually grow into adults and become less dependent on their caregivers over time, Baier insists that contractual forms of trust fall short of capturing the trust seen in parent-child relationships. She ends the section on 'Trust and Relative Power' by asserting that "Parental and filial responsibility does not rest on deals, actual or virtual, between parent and child" (Baier 1994, 110).

In the next section of the essay, 'Trust and Voluntary Abilities', Baier further justifies her decision to start with infant trust, rather than trust in contracts or trust in God, by exposing the falsity of another assumption contract-based ethical theories make about trust relationships:

that they are always voluntary. She begins, again, with infant trust, referring to it this time as both the "essential seed" and as "primitive and basic" (Baier 1994, 110). She states that her account "has been designed to allow for unconscious trust, conscious but unchosen trust, as well as for conscious trust the truster has chosen to endorse and cultivate" (Baier 1994, 110).[10] She suggests that one advantage of starting with infant trust is that it becomes "relatively easy to tell" a story about how we come to trust ourselves, leaving us to explain only how trust ceases, transfers, and changes in scope. She doesn't imply that these explanations are easy, but only that they are easier than the reverse: if trust were consciously and voluntarily chosen in its most basic form, rather than unconscious and unchosen (as it is in infant trust) we'd also have to explain why another's invitation to "trust me" often fails to illicit trust.[11] I take her point to be that, given what we know about how human brains and the reflective capacities required to make choices develop, it is easier to explain the how conscious forms of trust grow out of unconscious forms. The fact that we are not always able to will ourselves into trusting others, even when instructed to, serves as proof for Baier that latent forms of trust are more explanatorily basic.

This observation leads Baier to a discussion of the form of trust we see in promises. Here, Baier invokes Hume to argue that promises (and by extension contracts) are "artificially contrived and secured cases of mutual trust" (Baier 1994, 111). While she praises promises for being "an ingenious social invention" and a "complex and sophisticated moral achievement," she chastises moral philosophers of the past for trying to force all forms of trust and all kinds of trust

[10] It would take an essay itself to prove, but the categories of trust Baier offers here resemble the way Aristotle categorizes different forms of action. This parallel seems helpful as it suggests that the degree of voluntariness in the trusting relationship is an important factor to consider when judging when trust is sensible, morally decent, or blameworthy. It also points to a way in which her early work on trust is influenced by virtue ethics.

[11] Though Baier thinks that infant trust is of the unconscious and unchosen kind, she acknowledges that even infant trust is not indiscriminate. Her point is that even adults cannot always trust at will, so it would be implausible to assume that infants can. She argues that the invitation to "trust me" is either reassurance (if it is extended by someone we already trust) or encouragement to not *di*strust.

relationships into this special, artificial mold (Baier 1994, 111-112).[12] Though promises appear to possess a "verbal magic" that can initiate trusting, Baier is concerned that this feature has been mistakenly taken as proof that promises are a good model of what most instances of trust are like.[13] The main problem with using promises as a paradigm is that they too presuppose more basic forms of trust. Citing Hume, Baier explains that promises are an established social custom that have the power to reverse presumptions about trustworthiness. She appeals to climates of trust to explain what is missing from the picture: "once the social conditions are right for it, once the requisite climate of trust in promisors is there, it is easy to take it for a simpler matter than it is and to ignore its background conditions" (Baier 1994, 112). Here, we see Baier return to the image of a "climate of trust" to identify what the (already limited) discussions of trust in the history of moral philosophy are missing. Promises and contracts are not basic forms of trust, rather they are only possible because some other, more basic, and less voluntary forms of trust support them.

These more basic and less voluntary forms of trust are on display in the relationships and experiences of those who historically were denied the right to participate in the practice of promise-making. Baier considers the status of women who, historically, found themselves in relationships that "were not entered into by free choice or by freely giving or receiving promises" and "like the infant...found themselves faced with others to trust or distrust and found

[12] Baier reaffirms her interest in promises later in the essay: "Promises *are* morally interesting, one's performance as party to a promise is a good indicator of one's moral character, but not for the reasons contractarians suppose (Baier 1994, 118, emphasis in original).

[13] It is worth briefly noting that Baier also believes "awareness of what is customary, as well as past experiences of one's own" are factors that can impact climates of trust (Baier 1994, 111). This acknowledgement anticipates Baier's later work in *Reflections on How We Live,* where she considers how both features of our environment (including social structures) *and* our inner lives have the power to shape the climate of trust we inhabit.

themselves trusted or not trusted by these given others" (Baier 1994, 113).[14] Promises and contracts fail as a paradigm because the form of trust they embody portray trust relationships as ones that are voluntarily entered. Baier ends this section with what I would argue is one of her most compelling claims in the essay:

> ...the liberal morality which takes voluntary agreement as the paradigm source of moral obligation must either exclude the women they expect to continue in their traditional role from the class of moral subjects or admit internal contradiction in their moral beliefs. Nor does the contradiction vanish once women have equal legal rights with men, as long as they are expected to take responsibility for any child they conceive voluntarily or nonvoluntarily…Voluntary agreement and trust in others to keep their agreements must be moved from the center to the moral periphery once servants, ex-slaves, and women are taken seriously as moral subjects and agents. (Baier 1994, 113-114)

What we can infer from Baier's claim by transforming it into a conditional statement is pivotal: if servants, ex-slaves, and women are taken seriously as moral subjects, *voluntary agreement must be on the periphery, rather than at the center of philosophical accounts of trust*. The entrusting framework is modelled on the kind of trust seen in contracts and promises—it directs our attention toward the entrusted object and away from the relationship between the parties involved. When the object and conditions under which we can expect our trust to be met are spelled out explicitly, the character of the vulnerability taken on by the one-trusting is altered. Sometimes the promises we make (contractually or otherwise) overlap with conscious, endorsed, and chosen varieties of trust, but sometimes promises and contracts can function as a substitute for trust. The trust we have in contracts and promises is not rooted in an expectation tied to the relationship, but instead in the confidence we have in our ability to enforce sanctions on those who break them. For these reasons, Baier classifies contracts and promises as limit cases. We still entrust things to others in relationships where our obligations are not spelled out in a

[14] She goes on to add that the only "promise" of any significance women were allowed to make was "one vow of fixed and non-negotiable content, the marriage vow" which itself was "often made under duress" (Baier 1994, 113),

contract or promise, but in our intimate interpersonal relationships the expectation that the object entrusted will be properly cared for is not the only way such trust contributes to our flourishing. The entrusting model, which presents the just the two individuals in the trust relationship and the object or domain of the trust as the only variables we need to consider reproduces the tendency to focus only on trust's instrumental value. It also, as Baier points out in this passage, tends to overlook the extent to which the relationship between A and B is voluntary. Accounts of trust that continue to assume that all instances can be adequately modeled on the entrusting framework are thus prone to excluding those who voluntary abilities are limited in physical or psychological ways (such as infants, the ill, and the dying) as well as those whose range of voluntary choices are constricted by oppressive social norms and institutions.

Though the target of Baier's critique in the fourth section of her essay— 'The Male Fixation on Contract'— is still the limitations of contract-inspired moral theories, in this section her tone shifts to that of a historian of philosophy. Here, she attributes the neglect of "webs of trust" in part to biographical facts about canonical figures in the Western philosophical tradition.[15] Baier argues that, with a few significant exceptions, most of the great moral philosophers in the Western tradition were "a collection of clerics, misogynists, and puritan bachelors" (Baier 1994, 114). As a result of having little interaction with women, philosophers such as Hobbes, Butler, Bentham, and Kant, could more easily "ignore the virtues and vices of family relationships, male-female relationships, master-slave, and employer-employee relationships" and "confine their attention to the rights and duties of free and equal adults" (Baier 1994, 114-115). She contends that the work of these canonical figures, like the possibility of trust in promises, presupposed a background of domestic trust that they did not comment on or see as

[15] For those who might find Baier's claims uncharitable or insufficiently proven it should be noted that she explicitly states that she is aware that these claims need more substantial defense (Baier 1994, 115).

morally significant. Consequently, they did not find the trust in these relationships worth theorizing, even though they made the forms of trust they deemed worthy of philosophical investigation possible.

The inspiration Baier draws from the emergent care ethics literature is at the forefront of this portion of "Trust and Antitrust". She draws on the work of Carol Gilligan, to argue that more women's voices are needed to course correct philosophical conversations about trust because "equality of power and a natural separateness from others" are "alien to a woman's experience of life and morality" (Baier 1994, 116). According to Baier, "a moral code designed for those equal in power will be at best nonfunctional and at worst an offensive pretense of equality as a substitute for its actuality" (Baier 1994, 116). Modelling all our actual moral relations on voluntary agreement between relative equals is statistically problematic, as most of our relationships do not fit this mold.

It is also ethically problematic. Baier reminds us that equality of power is not ideal or desirable in many of our important moral relationships, including those we have with animals, the ill, the dying, and young children. Though the explicitness of contracts is part of their "functional excellence", it can also be a limitation (Baier 1994, 117). Explicit provisions can only be made for contingencies that we can anticipate or imagine arising. When we truly trust—even in conscious, chosen, and endorsed ways—we give the person we trust discretionary power. With this discretionary power comes the essence of trust Baier repeatedly identifies throughout the essay: vulnerability. Contracts have the advantage of being able to redistribute risk and vulnerability in ways that minimize it for both parties. However, Baier argues that this feature provides another reason to classify them as "a limit case of trust, in which fewer risks are taken for the sake of lesser goods" (Baier 1994, 118). The goods that come from trust are not ones that

can be exhausted by commodity exchange. A more complete moral theory, Baier argues, "would tell us how and why we should act and feel towards others in relationships of shifting and varying power and shifting and varying intimacy" (Baier 1994, 120). To this, I would add that a more complete account of trust also will acknowledge the uncommodifiable goods we gain from trust.

The sections of "Trust and Antitrust" I have examined so far all explicitly push back against contract-inspired ways of conceptualizing trust. Baier's final section in the essay, 'A Moral Test for Trust', targets another characteristic feature of dominant moral theories—their attempts to articulate universal rules and absolute moral principles. She argues that judging whether a trusting relationship is healthy or unhealthy is not amendable to this sort of evaluation. Instead, she offers some guidance by proposing a test that we could, in theory, apply to any particular instance of trust or to any trusting relationship. The test is summarized as follows: "trust is morally decent only if, in addition to whatever else is entrusted, knowledge of each party's reasons for confident reliance on the other to continue the relationship could in principle also be entrusted" (Baier 1994, 128). In other words, in a healthy trust relationship both parties could become aware of the other's motivating reasons for participating in the relationship without any adverse effects. In rotten or unhealthy cases, such an awareness would undermine or destroy the relationship.

Baier is upfront in acknowledging that her test may fall prey to the prejudices she has been arguing against. She worries that it "ignores the network of trust" (i.e. how different forms of trust and their respective climates affect one another) and is concerned by the fact that it can only be applied to "two-party trust relationships" (Baier 1994, 126). Although the test is limited in these ways, it does affirm that judgements about the appropriateness of trust require a

significant amount of attention to both of the individuals involved in the trust relationship and their circumstances (Baier 1994, 125). The emphasis Baier places on attending to interpersonal relationships and individual psychologies, as well as on taking social, political, and historical contingencies into account shows that even in her earliest reflections she is concerned with the embodied experience of trust. Furthermore, her preliminary test points back to the vulnerability she finds essential to all forms of trust. She warns that application of the test might affect the trusting relationship in unexpected ways. If we try to look for a deeper explanation for our trust by inspecting its roots, the trust may not always survive. For this reason, Baier argues that deciding whether and when to use the test in our daily lives is an entirely different question. She ends with the advice that it is "better to take nonsuspect trust on trust," even when doing so is a "very risky bet" (Baier 1994, 129). As we will see, these last few lines speak to Baier's persistent optimism about trust and her belief that it is an essential moral concept.

In this section I have shown that what has been claimed to be central to Baier's account of trust is actually closer to its periphery with respect to the images, themes, and questions that she engages with in her groundbreaking essay. In the next two sections of this chapter, I demonstrate how Baier continues to develop many of the ideas typically not seen in citations of "Trust and Antitrust" in the essays found in *Moral Prejudices,* and in her later book, *Reflections On How We Live.* [16]

Growth from Early Reflections

[16] These two books are the first and last monographs where Baier writes extensively about trust. I have no doubt that her other books, lectures, and articles on Hume, trust, and the mind, could further fill out her account of trust, but I must draw the line somewhere and since I am interested in how her thoughts on trust develop, focus on these bookends of her work on trust felt appropriate.

Moral Prejudices contains three other essays where Baier reflects extensively on trust: "Trust and Its Vulnerabilities", "Sustaining Trust", and "Trusting People". Throughout them, she reaffirms trust's central role in morality. In addition, these three essays elaborate on how our social practices contribute to climates of trust. Rather than work through each of these essays individually, I will focus on how they collectively build on the themes and ideas seen in "Trust and Antitrust". My aim is to outline an argument about the nature of trust that runs through all three essays, namely that trust is a response to our situation and, therefore, is informed by our perceptions. I start by looking closely at two anecdotes Baier shares in "Trust and Its Vulnerabilities" and discuss how they highlight the importance of attending to lived and embodied experience. Then, I briefly discuss Baier's observations about trust and language found in "Sustaining Trust" and "Trusting People". I end this section with an overview of Baier's observations about how to foster, maintain, and repair trust, which appear across these three essays.

In "Trust and Its Vulnerabilities" Baier states that trust is a feeling response to our situation (Baier 1994, 131). She is clear from the start that this response is not reducible to a set of beliefs the one-trusting has about the one-trusted, nor to an emotion or feeling, but rather has to do with our judgement, which is impacted by a variety of factors. She describes trust as "one of those mental phenomena attention to which shows us the inadequacy of attempting to classify mental phenomena into the 'cognitive', the 'affective', and the 'conative'" (Baier 1994, 133).[17]

[17] For some, the term "judgement" might imply that reflection or deliberation has taken place. However, I take it that Baier believes we often make unreflective judgements based on our initial sense-impressions of the world around us. Going forward I assume that our judgements can be sorted by their degrees of reflexivity. There are cases where suspending judgement until we've had time to reflect and deliberate is desirable, e.g. when one occupies the official role of a "judge" within a court, but there are also times where being too reflective or taking too long to deliberate about our judgements can be a liability—we simply do not always have time to step back and reflect before we respond. Furthermore, there are some cases where reflecting or deliberating on our judgement might undermine its moral worth. For example, when we see something that is clearly unjust, stopping to reflect or deliberate about it instead of addressing or responding to it right away opens the door for further harm.

The two anecdotes she offers in this essay simultaneously illuminate what these factors are and why accounts of trust inspired by contractarian systems of morality fail to capture them. Both stories involve young women who find themselves the object of unwanted sexual attention. Baier offers these anecdotes with the intention of illustrating not only "the vulnerabilities incurred by trust" but also "the difficulty of formulating any useful rules about how not to misplace our trust or to misuse our capacities for being trustworthy" (Baier 1994, 142).

In the first story, a young woman accepts her professor's offer to spend the summer at a remote cottage, which he had frequently rented in past years.[18] Over the course of her stay she becomes familiar, but not friendly, with the family that owns the cottage. They teach her the ways of their quaint town and make regular visits to deliver supplies. On the last night of her stay, the cottage landlord lingers longer than usual and persuades[19] her to walk with him from the cottage back to his family's home by claiming that the locals will get the wrong idea if they see him walking back from her cottage alone late at night. She has some hesitation, but unfamiliar with the local customs, she agrees. On their way, he attempts to rape her. She is fortunately able to escape and run back to the cottage. The next day she says her goodbyes to the landlord and his family as if nothing happened. Baier describes this as "a story of one forgiving too many, as well as a misjudgment of the extent of 'real kindness', of the undefined normal limits of friendship,

[18] It should be noted that the professor is not the owner of the cottage and does not accompany her during her visit. It is difficult to summarize every detail of this example concisely, but I think Baier included this complication to the scenario because it speaks to her point that trust operates via networks. The student ends up in this scenario because she trusted her professor's judgement in a few different ways. First, she trusted him not to recommend an unsafe place, and second, she trusted him as a mentor (the getaway was offered as an opportunity to focus on dissertation writing). This trust in her professor seems justified from what we know about their relationship. The "mistake" in her judgement was in extending her trust in her professor to the cottage landlord.

[19] These anecdotes are presented in much richer detail in the essay. While it could reasonably be said that "pressured" would be an appropriate synonym, it is less clear if "coerced" would be an appropriate or too strong of a substitute here.

especially of cross-cultural friendliness" and as one where a terrible outcome was avoided only by "plain good luck" (Baier 1994, 143).

The second story is about an unmarried woman hired into a junior level position in a philosophy department at a time when people like her were not commonly seen in the profession.[20] The woman receives "amorous attention" from two of her married male colleagues, which disrupts what she is has been told was previously a harmonious, drama free (and entirely male) department. The chair of the department, who is both her boss and the person who hired her, publicly accosts her for causing trouble. He reveals to her that the hiring committee had discussed how risky it was to hire an unmarried woman and shouts "*why don't you get yourself married and out of circulation*" at her after a meeting one day. Embarrassed, the woman does her best to gather herself and leave. When she gets home, she reports the incident to the administration and calls for the chair's resignation, only to be encouraged by the Vice-Chancellor to withdraw her complaint. She does, but only after thinking about how tough the current job market is and how following through might hurt her chances of getting hired elsewhere. She does her best to keep her head down the rest of the school year as she applies for other positions. When she leaves at the end of the year, the chair gives her an embarrassingly extravagant gift "as if to make it difficult for her to keep her grievance alive and well" (Baier 1994, 144). [21]

[20] While times *have* changed since Baier was writing with respect to women in philosophy, the question of to what extent are they different today remains open. Lots has been said about how habitual patterns of sexism and misogyny are re-inscribed. It appears to me that Baier's example might need a few key details changed to be updated, but that the point she makes here still holds. Empirical research on this matter conducted by Conklin, Artamonova, and Hassoun (2020) shows that while the percentage of women in tenured or tenure-track positions in philosophy departments has increased, women still make up significantly less than half of almost all philosophy department faculty. Furthermore, women are better represented in lower-ranking positions, which shows "there is a clear pyramidal shape to the discipline" (Conklin et al. 2020).

[21] Autobiographical footnotes in Baier's *Death and Character* and *Reflections on How We Live* reveal that this second example is pulled directly from her experience at her first faculty job.

What Baier is bringing to the forefront with these two stories is the extent to which our interpersonal trusting relationships are entangled with a variety of other factors that can only be understood by attending to social, historical, institutional, and cultural contexts as well as the embodied experience of the one-trusting. Attempting to make judgements about trust without attending to these particularities is both descriptively problematic (it doesn't capture what making judgements about trust is actually like) and prescriptively questionable (rules about when to trust that do not take these factors into account are likely to yield bad judgements). In the first case, there are several asymmetries of power. In addition to her gender, the woman was a guest who was unfamiliar with local customs and, therefore, she was dependent upon the family in several ways. In the second case, the fact that the woman was unmarried and occupying a professional role typically held by men is clearly the central issue. However, her situation is worsened by several other factors, including a lack of institutional support from the university administrator, the state of the job market, and the chair's performative gift-giving. Through these two richly detailed narratives, Baier demonstrates that our embodied experience of trusting is relevant when it comes to making judgements about when trust is appropriate.

Furthermore, these two cases are difficult to model on the entrusting framework because precisely who the women trusted, and with what is difficult to discern (i.e. they are both cases where we must strain to discern a definite candidate for 'C', and to some extent 'B' as well). Furthermore, Baier argues that in both cases there was no reasonable alternative to the trust extended by the victims. Not trusting, or actively distrusting in either case would have come at a high cost—without trust in those around us, our lives become solitary and bereft of the many goods that come from living with or around other people. She ends this section by reaffirming that while cultivating our powers of judgement can improve our ability to determine when trust

and distrust are appropriate, there are "no useful rules to tell us when to trust or even when we should have trusted" (Baier 1994, 151).

Baier continues to build on the argument that judgements about when to trust and distrust cannot be determined fixed by rules and universal principles in both "Sustaining Trust" and "Trusting People." In "Trusting People", the developmental aspects of her account of trust resurface. Baier makes the same inference that some willingness to trust must be innate and expands on what this means for her account (Baier 1994, 195). According to Baier, we first learn about the norms for trusting and judging trustworthiness from those who take care of us in infancy and childhood, and then from the communities or societies in which we grow up. There is, however, an important caveat—some capacity to trust is a prerequisite for cultivating the ability to make judgements about trust (Baier 1994, 195). Some may see this condition as unsubstantiated and argue that making judgements about whether to trust requires a certain level of cognitive abilities and, in particular, the ability to form a belief about whether or not a person is trustworthy (either full stop, or within the relevant domain). However, I believe Baier's account—which acknowledges that our judgements about trusting others are often made in the moment and are shaped by our upbringing— proves helpful for developing a deeper understanding for how the experience of trauma, abuse, and neglect can damage not only one's capacity to trust, but also one's ability to judge when trust is appropriate or warranted. [22]

In "Sustaining Trust", Baier responds directly to the view that judgements about trust can be made solely based on beliefs about others' motivations. In particular, she raises doubts about theories that assume trust is a matter of accurately discerning whether the one-trusted has altruistic or non-egoistic, rather than egoistic or selfish motives (Baier 1994,152-159).

[22] In chapter five, I explore these ideas further and argue that my developmental account can illuminate this aspect of trust better than existing accounts.

Furthermore, she argues that such accounts of trust do little to explain the short-term, impersonal kinds of trust (i.e. trust in normal appearances, trust in uniforms and in badges, trust in framed certificates) that are extended subconsciously.

Even if we *could* become experts at discerning others' motivational sets with precision, there is a deeper assumption Baier takes issue with: the notion that human motivations can be easily divided along the lines of egoistic and non-egoistic.[23] Her critique here is reminiscent of the complaints seen in "Trust and Antitrust", only this time aimed at traditional Western philosophers' understanding of human psychology: "The ego's boundaries are less clearly marked than are most nations' boundaries, but in the absence of clear boundaries we cannot be sure when our concern is for ourselves alone, when for others" (Baier 1994, 156). Rather than try to reclassify our motivations using some other binary standard, she suggests we move away from speculating about differences in kinds of motives and instead look at how "our attitudes and actions in our dealings with persons standing in all degrees of closeness and distance from us fall into pretty regular patterns of habitual behavior" (Baier 1994). When we acknowledge that our assumptions about others' motives are based on behaviors we observe, we realize that we often "take many appearances on trust, as we would go mad if we did not and could not" (Baier 1994, 159).

After demonstrating that judgements of trust cannot and should not be based solely on the one-trusting's beliefs about the one-trusted's motivations, Baier turns to the kinds of cues that she believes do factor into our judgements about trust. In addition to reminding us of the

[23] In this essay, Baier begins by directly responding to an essay on trust by Bernard Williams. However, Baier is clear that she is taking his view as representative of a category of philosophers who all make the same assumption that there is a sharp contrast between egoistic/ non-egoistic as well as micro/macro motivations. Micro-motivations are sensitive to particular contexts and people and macro-motivations are more general, directed at recurrent and recognizable types.

importance of attending to historical and social context and the relevance of personal past experiences, she also considers how we communicate to be another factor that contributes to our trust judgements.

Baier refers to speech as "our cooperative and trust-facilitated activity *par excellence*", describes it as "one of the tough pervasive webs of trust", and observes that "We do cooperate in speaking, even in our uses of speech to wound and insult" (Baier 1994, 175). However, she believes that this social practice builds on "primitive cooperative practices" such as nonverbal bodily expression and gestures, which we typically interpret without worrying about what the person's motivation might be (Baier 1994, 176). We learn these forms of communication through the body before our cognitive abilities become more sophisticated. We learn the norms of primitive cooperative practices along with language from those who raise us, and we pass them on to future generations. The social habit of trusting these forms of communication is justified because "like all habits [it] increases its strength, wears thick, not thin, by constant use" (Baier 1994, 176). Furthermore, because speech increases our ability to deceive one another, we also have developed the ability to synthesize non-verbal cues and speech. According to Baier, it is these perceptual capacities, which depend on an ingrained awareness of what is customary, that are essential to the formation of our judgements about who to trust when, with what, and under what conditions.

The form of trust that is required for the social practice of language is, like infant trust, a basic form out of which other more complex forms of trust arise. Other forms of broader social trust (i.e. trust in contracts, trust in institutions, trust in governments) develop out of the already established practice of language.[24] The connection she sees between language and trust lead her

[24] Baier makes this point in "Sustaining Trust" when replying to Scanlon's account of trust. What she objects to in his account is the idea that general or social trust is secured by a moral principle that does not presume the existence

to further elaborate on her earlier observation that many forms of trust go unnoticed (often precisely because of their prevalence in our everyday lives).

In every one of the *Moral Prejudices* essays on trust, Baier describes trust as a climate, atmosphere, web, and/or network. As she develops this concept, she offers more insight about what contributes to the health of a climate of trust.[25] According to Baier, understanding how trust functions demands that we look at how different forms of trust interact and affect one another:

> There are interesting differences between the trust of intimates and what is good about it and the nature and value of more impersonal trust, and each is prey to some sickness peculiar to that type, but the main dimensions of fragility are the same, and there are interdependencies between a healthy climate of impersonal trust and the likelihood of a strong trust relationship of a more personal sort (Baier 1994, 137).

At the interpersonal level, Baier notes that trust is difficult to get started. She sees gestures of mutual disempowerment as well as the social practice of promise-making as potential ways to initiate a trust relationship (Baier 1994, 195-196).[26] She speculates that in relationships where there is no trust, or where there is distrust, a mutually trusted third party may be the only way to establish or repair trust. Trust can be destroyed at the personal level by a craving for security, and the desire for power can corrupt trust on both global and interpersonal scales (Baier 1994, 160). Additionally, she suggests that we can place trust in procedures, as well as people (Baier

---

of a concrete social practice, such as promise-making or oath-taking. She also objects to the idea that trust can be reduce to a set of universal principles. In particular, she disagrees with Scanlon's argument that trust can be explained by a "fidelity principle", which he argues is the basis of agreement, on the grounds that it would undermine trust by depriving those we trust "of the freedom to do what they judged best when the time came to act" because "the principle would require them to do what they had given others reason to believe they would do" (Baier 1994, 169). Genuine trust, she argues, is trusting" without *special assurances"* (Baier 1994, 170, emphasis in text).

[25] In *Moral Prejudices* these different ways of describing trust are used more or less interchangeably and appear roughly the same number of times. However, as we saw earlier, Baier used "atmosphere" and "climate" first, and as we will see shortly, by the time she is writing *Reflections on How We Live,* climate becomes her standard term.

[26] Her example of a gesture of mutual disempowerment is the handshake, which she says, "is learned in almost all societies as the symbol of some mutual trusting, and it grows out of more asymmetrical trusting of small hands into larger ones, offered to support or to guide" (Baier 1994, 196).

1994, 201). With these additions, Baier indicates that healthy climates of social trust can encourage those within them to take risks, which, by extension, promotes interpersonal trust. [27]

The picture of trust we're offered by the end of *Moral Prejudices* is complex, but that is part of what makes the climate metaphor she relies on throughout these essays so apt—climates, like trust, are affected by a number of variables. In the next section, I show that Baier's climate imagery not only becomes central to her understanding of how trust operates, but also that it gives her account of trust several distinct advantages.

Baier's Mature Account of Trust

While Baier's account of trust in her later work is still clearly informed by concerns associated with care ethics, her mature account of trust is articulated in the vernacular of a virtue ethicist.[28] The way she navigates between the two ethical frameworks contributes to the dexterity[29] of her account.[30] In this section, I examine two essays from *Reflections on How We Live* that show a number of developments in Baier's account of trust. The first is the role of trust in Baier's moral philosophy. In her earlier work, we saw that Baier took trust to be essential to facilitating cooperation and caring for the kinds of things that contribute to a flourishing life. In these two later essays, Baier describes the material preconditions that are necessary for a climate of trust and makes an explicit connection between trust and morality writ large. The second

[27] In his forthcoming essay, Jason D'Cruz introduces the concept of "trusting overtures" which expands on the connection between courage and fostering trust. The concept suggests a way that individuals can transform or alter the climate of trust they inhabit and share with others.

[28] Many of the essays in *Reflections on How We Life* expand directly on topics that Baier only gestures at in *Moral Prejudices.*" Why Honesty is a Hard Virtue" and "The Moral Perils of Intimacy", for example, further explore the same questions about the relationship between language, perception, relative power, and vulnerability that were seen in her earlier work. In other words, despite not having the word "trust" in their title, the concept is consistently employed and operating in the background.

[29] By characterizing her account as having dexterity I aim to draw attention to the ways in which Baier's account remains aware of how the manipulation of different variables can drastically alter when an instance of trust or distrust is warranted or justified.

[30] What I can say about how she blends these two normative frameworks is constrained by the nature of this endeavor. For a helpful discussion of Baier's take on care ethics and on the relationship between care and virtue see pages 19-20 and 56-57 in Virginia Held's *Ethics of Care* (2006).

development is in Baier's increased use of the phrase "climate of trust". The third, related development is that Baier takes her reflections about how to foster trust inward—on what we can change about ourselves to improve the climates of trust in which both human beings, and the things that matter to them, thrive. While "Trust, Demoralization and the Virtues" touches on this point, this inward turn is Baier's focus in "Sympathy and Self-Trust", which adds another layer of complexity to her account.

Baier asserts that "some degree of social trust in the world is the starting point and very basis of morality" (Baier 2009, 179). She then offers evidence to support this claim by looking at how trust underlies relationships between individuals, local communities, and nations. At the individual level, she appeals to scientific knowledge about how infant brains develop, noting that regions of the brain associated with emotional regulation and moral reasoning are underdeveloped in those who were traumatized, neglected, or otherwise lacked trusting relationship as infants or children (Baier 2009, 179).[31] On the societal scale, Baier argues that it is unreasonable to expect trust and moral virtues from those living in demoralized conditions. She characterizes demoralized conditions as ones where people lack basic necessities (like food and shelter) or as ones where one's physical safety is in jeopardy (such as battlefields and disaster zones) (Baier 2009, 178).

It should be noted that *feeling* and actually *being* secure can come apart on Baier's account, and that both are important preconditions that must be met if an individual is to enjoy the benefits that come from living in a climate of trust. One may continue to feel unsafe or

[31] Baier cites a 2002 report from Jo Carlowe that covers the work of Bruce Perry at the Child Trauma Academy in Houston and Peter Fonagy at University College in London to support her claims here. More recently, Bessel van der Kolk explains in his book *The Body Keeps the Score*, that both the limbic brain, which is mainly organized during the first six years of life, and prefrontal cortex are both impacted by trauma exposure (van der Kolk 2015, 59). Saying more would take me beyond my area of expertise, but the science seems promising.

insecure even after the conditions have changed and therefore one may remain distrustful or weary of trust. This is especially pronounced in individuals who have never experienced security or those who endured extended periods of insecurity. What these observations demonstrate is that Baier's mature account of trust is consistent in seeing the material conditions and embodied experience of individuals as important factors to consider when theorizing trust, neither of which fits into the formula offered in the entrusting framework.

Baier begins "Trust Demoralization and the Virtues" by looking at a critique of virtue ethics that was popular at the time she was writing. The critique takes the form of a doubt, a doubt that individuals have dependable character traits. Skepticism about the possibility of cultivating lasting character traits, such as the virtues, led some ethicists to believe that one's circumstances, not one's character, offer a more accurate explanation as to why individuals act the way they do (Baier 2009, 173). Those who adopted this position complained that the virtues that a person appears to act from seem to depend too heavily on what circumstances they find themselves in. While Baier admits that external conditions like war, plague, famine as well as personal shock and misfortune can "threaten to rob a person of the good qualities she has been reputed to possess", she balances it with the claim that: "there will be normal conditions in which good habits of the heart can be cultivated and more or less survive" (Baier 2009, 174). This line of critique, prompts her to defend virtue ethics by suggesting a new method for identifying virtues: assessing traits on the basis of how they contribute to or detract from a climate of trust.

On Baier's proposed revision to virtue ethics, all virtues share a good moral quality that lies "in the mental attitude a person has, either on a particular occasion or on a succession of like occasions, to an ever-present fact about our human situation, namely, our mutual vulnerability"

(Baier 2009. 174).[32] We can determine whether a person's attitude about our mutual vulnerability is virtuous by looking at "its contribution to the climate of trust within which the person lives" (Baier 2009, 174). While one-off acts of bravery in the face of great danger might make a large contribution to the climate of trust by preventing panic, lasting traits—such as humility—do most of the work when it comes to maintaining climates of trust. Humility is a virtue that allows us to accept our human fallibility with grace. A readiness to admit to our own mistakes and acknowledge the limitations of our knowledge makes it easier to forgive small betrayals of trust against us and to acknowledge when we fail to meet the trust others have placed in us. What makes actions virtuous or not, on her proposed account, is the thought of our power over each other. Moral virtues, Baier argues "regulate, sometimes by increasing the volume of, sometimes by silencing, some variant of the mother thought of our power over each other, for good or ill, and the point of such attempted regulation is improvement and maintenance of a climate of trust" (Baier 2009, 175). With this test in place, Baier sets out to re-evaluate a list of traits traditionally understood as virtues based on how their presence contributes to or detracts from a climate of trust.

An outcome of this analysis is that Baier identifies a pair of virtues that become central to her account: considerateness and thoughtfulness. Though she considers several others, I'll restrict my discussion to these two, which reaffirm Baier's earlier argument that social and emotional forms of intelligence contribute greatly to one's ability to discern when trust is healthy and when it is rotten. For Baier, "the considerate person is appropriately aware of how her attitudes and actions affect those around her and if necessary, will alter them to avoid causing fear, hurt, or other unpleasant feelings in others, especially those who she wants to cooperate

[32] Baier defines mental attitudes by referring to what Hume called "lively thoughts", those that give content to desires, emotions, and intentions.

with or receive help from" (Baier 2009, 176). What is important to note about this virtue is that it requires awareness of relative power dynamics, a willingness to listen to others, and the ability to converse without forcing one's views onto others.

Baier identifies thoughtfulness as the second primary virtue in her climate-of-trust rendition of virtue ethics. In doing so she reaffirms that social and emotional forms of intelligence contribute to the climate of trust one inhabits. According to Baier, cultivating the virtue of thoughtfulness requires us to refine our perceptual capacities so that we can attend to others in a way that acknowledges their individuality. When this virtue is properly cultivated, a person will "notice the particular vulnerabilities of those around her" (Baier 2009, 176). A thoughtless and imperceptive person is not the same as an inconsiderate person because a considerate but thoughtless person can still treat others as she would like to be treated. However, without the other half of this complementary pair of virtues, an otherwise considerate person is at risk of acting in ways that will unwittingly cause harm to others. These observations serve as additional evidence that Baier's account of trust remains deeply concerned with embodied experience, even at its most mature stage. To cultivate the virtues of considerateness and thoughtfulness we must pay attention to the physical cues we receive from others' body language and be able to imagine how their experiences may differ from our own.

Baier argues that one advantage of evaluating the virtues with respect to how they contribute to or detract from the climate of trust is that it gives the virtues a "loose unity" which makes them "more than a mere bundle" without "reducing the comprehensive variety to any one virtue" (Baier 2009, 185). It also allows us to see that some virtues are best understood in pairs and that many virtues take new forms as technology and the conditions of life change (Baier 2009, 186). She ends by adding that she "takes the virtues to include good judgement, but also to

include many other qualities that affect such judgement and such willingness, all of them attitudes to mutual power and vulnerability" (Baier 2009, 188). What this demonstrates is that what Baier understands to be at the core of trust—the distinct kind of vulnerability it exposes us to—remains consistent over time. Though she adopts virtue ethics terminology in this chapter, concerns about relationality, vulnerability, and care remain essential to her understanding of trust.

It is also clear that the use of climate imagery remains essential to Baier's account of trust. Baier understands a climate of trust to be affected by the past experiences of individuals as well as shaped by the kinds of institutions and practices in the society they find themselves in. This opens the possibility that trust can be improved (or harmed) in two directions: by changes to structural features of society and material conditions, as well as by individuals' attempts to focus their attention on themselves and others. In other words, both ends of the spectrum can function as starting points for altering a climate of trust. In describing Baier's account of trust as dexterous, I want to draw attention to this aspect of her account and the implications it has for applied ethics. If both interpersonal relationships and structural or institutional features of society can impact trust, interventions aimed at fostering, maintaining, or repairing it must look at the way they interact. Another advantage of Baier's climate metaphor is that it allows her to take her account of trust in a global direction: "a decent climate of trust demands some measure of equality, not just among citizens of one nation, but among nations" (Baier 2009, 181). On Baier's account, trust's social and interpersonal forms are interwoven, not separate phenomena. Attempting to analyze them as separate will result in a misapprehension of how trust operates. We might miss, for instance, the fact that more localized climates of trust often exist within

regional or global climates of distrust. Baier's suggestions for fostering healthy climates of trust remains tied to material conditions even at the global level:

> The individual virtues we need to cultivate in order to get a great equality are not merely a sense of fairness and the willingness to protest (and relinquish) unfair advantages but also the vision to design workable institutions, both national and international, or to extend existing ones in ways that improve our overall climate trust. (Baier 2009, 182)

Attending to the fact that climates of trust and distrust themselves admit of scales can help us model the networks by which trust operates more accurately. It also provides us with an ethical impulse to cosmopolitanism by drawing attention to the fact that climates of trust within a nation or community that depend on the unfair or unjust treatment of other nations and communities actually exist within (and are responsible for) a larger climate of distrust which deprives those living outside the local or regional climate of trust of a flourishing life.

The use of climate as a metaphor is also helpful for developing an account of trust that can be applied to contemporary questions in applied ethics because of what it implies about the power individuals have and the limits of that power. If we think of improving trust as analogous to addressing the climate crisis, it becomes clear that both require the collaborative efforts of communities. However, it also demands that those in positions that afford them the power to alter practices and structures causing harm use that power appropriately. Individuals can make a small difference and these small differences can have a cumulative impact. However, the success of such efforts depends on individuals being disposed to intentionally cultivate virtues that contribute to a climate of trust and such virtues may need to be cultivated before individual cooperates, collaborates, or advocates for structural adjustments if their actions are to create lasting and impactful change.

In "Sympathy and Self-Trust", Baier explores the concept of self-trust, which was not prominent in her earlier work. Baier writes that she felt compelled to revisit the concept after

years of setting it aside because she realized that in her earlier work she had "foolishly assumed that self-trust was the norm" (Baier 2009, 190).[33] Much of what Baier discusses in this essay is exploratory and relies heavily on her interpretation of Humean sympathy, which I will not attempt to explain at length here. What I will say is that her investigation of self-trust through Hume's conception of sympathy allows her account of self-trust to remain relational, even though she shifts to focus on trust within a single individual. She considers at length the ways in which our self-trust or self-distrust can spread to others, noting that "sympathy, as well as trust and friendship, connects us closely with others, and like love makes us vulnerable to their troubles" (Baier 2009, 190). An interesting implication of this idea is that it also gives others power to affect the degree of trust we have in ourselves.[34] She adds that "all cases of sympathy with another's trust or distrust will be cases where we think we know and share the other's reasons for their trust or distrust" (Baier 2009, 201).

It is clear that Baier sees some links between trusting our own perceptions, the judgements based upon them, and the formation of our trust-related attitudes and beliefs. Though her claims about self-trust in this essay are preliminary and her death a few years later unfortunately prevents us from seeing how she might have developed this line of thought further, what is expressed in "Sympathy and Self-Trust" reintroduces the question of how it is possible to determine what warrants trust. When read in conjunction with "Trust, Demoralization and the Virtues" it confirms the idea that cultivation of virtues is essential to creating a climate of trust.

Conclusion

[33] Baier adds that this realization was sparked by work of two other philosophers: (1) Jennifer Whiting, who argues that self-distrust is often warranted and questions Baier's assumption that we must trust our friends and (2) Aristotle, who describes friends as "second selves" (Baier 2009, 190).

[34] Baier offers an anecdote about finally learning to drive as an adult. She argues that accomplishment would not have been possible without the self-trust she developed as a result of the trust her driving instructor placed in her. This leads her to reflect on the difference between trust and hope, which gets developed further in other essays included in *Reflections On How We Live*.

At the start of this chapter, I stated that my reason for beginning with Baier was to situate my project as an expansion of her work. By way of conclusion, I think it is prudent to be explicit about how what follows in the next four chapters constitutes such an expansion.

The first way in which my dissertation can be seen as an expansion of Baier's work is with respect to what she is reacting to. As I mentioned in section one, the entrusting framework Baier offers early in "Trust and Antitrust" has remained popular among philosophers who work on trust. Much of the extant literature on trust takes for granted that trust is a three-part relation, especially when it comes to the way it functions in interpersonal relationships. While this framework has generated valuable insights, the fact that it has become a stable assumption indicates that Baier's complaint about contractarian bias in philosophical conceptions of trust is still relevant. Therefore, the critique she offers in "Trust and Antitrust" informs my argument that many of the available accounts of trust fail to accurately depict how trust works in relationships with asymmetries of power.

My work is also an expansion with respect to what it attends to. As I've demonstrated, Baier's account of trust pays attention to the embodied, contextual, and phenomenological dimensions of trust. Such dimensions, I argue, are essential to an account of trust that can accommodate trusting relationships where there are stark or subtle asymmetries of power. Furthermore, the developmental account I offer also focuses on the interaction between individuals and their environments and among various forms of trust.

Finally, I also aim to expand on Baier's work by drawing out the explanatory remit contained within her account. Because Baier attends to particularities and lived experience, the developmental interpretation of her account that I offer is better suited than most for application and practical interventions. What I mean by this is that Baier's account provides a helpful

regulative ideal—a healthy climate of trust— as well as insights on into how we can foster, maintain, and repair trust. In addition to capturing how trust works in a wider range of relationships, I will go on to demonstrate that my developmental reading of Baier's account has the additional advantage of being able to connect trust to insights about trauma and identity.

Chapter Two: What Should an Account of Trust Be Able To Do?

"Few hurts which human beings can sustain are greater, and none wound more, than when that on which they habitually and with full assurance relied fails them in their hour of need."

-J.S. Mill

In the last chapter, I argued that many of Baier's insights about trust have been overlooked and that inattention to them has led to a misrepresentation of her work. I then offered an alternative interpretation of her account of trust, one that I characterized as developmental. In this chapter, my aim is to answer the question posed by this chapter's title: What should an account of trust be able to do? I propose three requirements.

Before I proceed to identify and explain my reasons for the items I've included on this list, I want to say a bit about why I chose a quote from Mill as this chapter's epigraph. Baier's work on trust, though otherwise deeply informed by canonical figures in the Western tradition of philosophy, does not say much about either utilitarianism or Mill.[35] Yet, Baier's observation that there has been a strange silence on the topic of trust is applicable here, where we see Mill going to great lengths to avoid even using the word "trust". I think it is worth considering both what is

[35] No variation of the word "Utilitarianism" makes the index of either *Moral Prejudices* or *Reflections on How We Live*. Baier's references to Mill in *Moral Prejudices* total to four, three of which are passing mentions. Her only substantive engagement with Mill in either of these books is a comparative discussion of Mill's and Hume's epistemologies: "If Hume gives us an early capitalist social epistemology, Mill gives a high capitalist version" (Baier 1994, 91). Though Baier is critical of both theories' capitalist leanings, she finds Hume's epistemology advantageous compared to Mill's because she sees it as capable of transforming the concept of reason "from being a quasi-divine faculty, something that we share with God" to "a natural capacity and one that is essentially shared with those who learn from experience in the way we do, sharing expressive body language, sharing or able to share a language, sharing or able to share our sentiments, sharing or able to share intellectual, moral, and aesthetic standards, and sharing or aspiring to share in the setting of those standards" (Baier 1994, 94). The picture of humanity we get from Baier's channeling of Hume is a cooperative one in which the progress we make in various areas of culture and science contributes the flourishing of society overall. It mirrors the values, questions, and methods typically found in pragmatist and virtue ethics traditions. All this is to say, again, that there is more work to be done when it comes to situating Baier in the history of philosophy and that doing so would further fill out the developmental interpretation of Baier's account of trust I proposed in the last chapter. I won't be doing that here, but I do want to "set the stage" with these thoughts, as my reasons for choosing the requirements I offer in this chapter radiate from this background.

implied and what questions are raised when we make following substitution: "Few hurts which human beings can sustain are greater, and none wound more, than when *trust* fails them in their hour of need". The depth of the wound, kind of harm, and the connection to justice we can associate with failed trust captures something about why we value trust that I think many contemporary accounts of trust have not emphasized enough.

If we assume Mill's statement here is correct, and if we agree my substitution of "trust" for "that on which they habitually and with full assurance relied" is acceptable, I could offer a more concise answer to the central question of this chapter.[36] That answer would be: an account of trust should be able to explain why failures of trust can result in the deepest wounds and are among the greatest hurts human beings can sustain. Framed in this way, the requirements I offer in this chapter could be thought of as attempts to explain why my revised version of Mill's claim is true.

The structure of this chapter is as follows. I will begin by describing the case of a recovering addict which I will refer to throughout the chapter. Then, I offer three answers to the question: what should an account of trust be able to do? I propose that it should be able to 1) account for distrust; 2) make a *meaningful* distinction between trust and reliance; and 3) attend to the way trust functions in relationships with asymmetries of power. Each section includes reasons of my own and draws on reasons that have been offered in the trust literature. What I

[36] For those who find themselves wanting more to support my claim here, consider the context in which this quote appears. In the sentence immediately before Mill claims that disappointments of expectations, and, in particular, breaches of friendships and promises, are "highly immoral acts" that hold "an important rank among human evils and wrongs" (Mill 2001 [1863], 60-61). Immediately after, he adds that "fewer wrongs are greater than this mere withholding of good; none excite more resentment either in the person suffering or in a sympathizing spectator" (Mill 2001 [1863], 61). This discussion occurs when Mill is elaborating on the notion of "just deserts", the seed of Mill's conception of Justice, which itself could be interpreted as developmental. My diagnosis here is that Mill recognizes that the prudential value of trust, which lies in its ability to facilitate cooperation, does not exhaust its moral significance.

offer should *not* be interpreted as an exhaustive list of what an account of trust should be able to do. While it would be accurate to describe them as a set of necessary, but not sufficient conditions, my reasons for avoiding the use of such terminology stem from my doubts about the fittingness of a conceptual analysis approach to trust. I believe there is more that an account of trust should be able to do, and I will describe some of these things in chapters four and five when I discuss additional advantages of my developmental account. I also think there are things that an account of trust should *not* do. I hint at some of them throughout this chapter, but I'll defer most of these considerations to chapter three.

A Case for Consideration: The Recovering Addict

After a close encounter with death stemming from her drug and alcohol abuse, a now-recovering addict spends several months in rehab and moves back in with her parents. Her parents, few remaining friends, and other members of her family—who she lied to, stole from, and failed to make good on promises with prior to getting help for her addiction—continue to exhibit distrust toward her in the weeks following her release from the rehabilitation center. We can imagine that, at least at first, the recovering addict (who has apologized, expressed regret, and is still coping with the guilt from her past actions) would find this distrust understandable, even if it pains her emotionally. These feelings might prompt her to prove she is worthy of trust by attempting to make amends, right past wrongs, and by behaving in ways that demonstrate trustworthiness.

Several years pass. The recovering addict has not relapsed; she has taken up new endeavors where she has had the opportunity to prove herself to be reliable, trustworthy, and dependable; and she has worked hard to try to reestablish trust with her family and friends who knew her during the worst periods of her addiction. Despite her efforts, she can sense that her

parents often still distrust her—they call repeatedly when she is not home exactly at the time she said she would be, they are apprehensive about loaning her money, or allowing her to house-sit when they go on vacation, etc. Her friends, too, seem suspicious when she is in a good mood, they are weary when she attends parties where others are drinking, and they are visibly shocked when she remembers the kinds of things she used to forget when she was abusing drugs. When she calls them out on the distrust she perceives, they deny it. The recovering addict's experiences of distrust seem to fall into a middle ground—they are not entirely without warrant or unreasonable, but they also seem to be inappropriate given the progress and efforts she has made to become trustworthy. We can imagine that, after years of being treated this way, the recovered addict might fall back into her old habits— ones that merit the distrust she experiences. Her thought process might be something like: *It doesn't matter if I drink again/ fail to pay her back/ don't show up to the party. It's what they expect of me anyway.*

Cases like this show the difficulty of attempting to distinguish 'warranted' or 'appropriate' trust and distrust from 'unwarranted' or 'inappropriate'.[37] They do so by highlighting the fact that the moral and non-moral criteria for making these distinctions are tough to pull apart. When we make judgements about when and who to trust, we appear to use a mix of both and doing so seems both inevitable and advisable. Non-moral facts often provide essential guidance about who to trust when. Our experiences, which admit of both moral and non-moral dimensions, inform our understanding of what is at risk when we trust. Furthermore, differences in individual experiences can also affect how moral and non-moral criteria are weighted. We can imagine that the bar the recovering addict must clear to regain the trust of a childhood friend who

[37] I am using these terms, as well as "justified" and "unjustified" and "reasonable" and "unreasonable" interchangeably. I have seen all of these used to describe trust and distrust and I have not discerned any consistent pattern. When I get to my own account, I will use "healthy" and "unhealthy", by which I aim to distinguish forms of trust and trusting relationships that contribute to flourishing from those that detract from it.

has a similar relationship with drugs and alcohol might be lower than the one she must clear to regain the trust of her sister whose abusive partner recently overdosed. The same might be true for the different healthcare providers that the recovering addict interacts with.

Another reason I chose the recovering addict case is because addictions present a persistent challenge for ethicists. They are a type of habit—a concept central to the theories of many ethicists and moral philosophers—and the notion that our habits (and their formative activities: practice and education) are important for moral development is generally accepted. Habits are particularly salient in ancient virtue ethics theories in both Western and Eastern traditions (e.g. Aristotle and Confucius)[38] and among virtue-ethics leaning theorists pre-and-post enlightenment (e.g. Hume and Mill). Furthermore, interest in habits is a common thread in the moral philosophies of those in the American pragmatist tradition.[39]

Addictions are often the culprit when it comes to explaining why people act badly or fail to cultivate virtuous characters. They are identified as an obstacle that prevents proper upbringing and education from resulting in good habits. Mill sees them as destructive not only of an individual's physical health and moral development, but also the potential of moral progress in society (Mill 2001 [1863], 10-11, 14-15). For Aristotle, addictions to the wrong kinds of pleasures can spoil one's chances of becoming truly virtuous (hence his emphasis on cultivating temperance from an early age). Korsgaard's chapter "Defective Action" i*n Self-Constitution* makes a compelling argument to show that Plato and Kant thought similarly about one form of addiction—obsession—which can deprive a person of their agency and personal integrity

[38] See Wee (2011) for an account of how trust relates to the Confucian virtue of *Xin* (信). Interestingly, the character emphasizes the forms of trust and trustworthiness associated with testimony—the radicals that comprise it are of a person standing (亻) next to their word (言).

[39] For a contemporary account of the role habit plays in John Dewey's work and its potential for social transformation see Carolyn Pedwell's book *Revolutionary Routines* (2021).

(Korsgaard 2013). Further complicating the matter, modern biology and psychology have revealed that the development of addictions can often be explained by complex interactions between genetics, one's past experiences, and one's environment.

The case of the recovering addict makes a good litmus test for an account of trust: it involves multiple, fluctuating, and intersecting relationships, portrays a psychologically complex character, and engages with other pressing ethical questions. Additionally, it raises questions such as: what constitutes trustworthiness? And what counts as a "good" reason to trust or distrust? Knowing that someone has an addiction might count as a good reason to avoid trusting them, at least in situations when they are likely to be tempted. However, in the case I've just outlined, not extending trust after a person seems to have really changed seems harsh and unfair. Overcoming one's own addictions requires one to be vigilant about when they can and cannot trust themselves. Being distrusted by others can be hurtful or insulting, and it can undermine self-trust, which may lead to more lasting harm. An adequate account of trust should be able to provide an explanation of the recovering addict's experience and assess when the distrust she experiences is and is not warranted. An account that meets the following three requirements will have the resources to do both.

Requirement #1: Accounting for Distrust

The attachment of a prefix can only tell us so much about how the concepts or trust and distrust are related. Clearly, they oppose each other, but the exact nature of this opposition is hard to pinpoint. Fortunately, discussions of distrust are not as rare as they used to be. In this section of the chapter, I rely heavily on a handful of influential essays that demonstrate different ways in which the concept of distrust deepens our understanding of trust. From Hawley, we see how the structural and formal elements of distrust complement and complicate the concept of

trust. D'Cruz's work brings the epistemic and ethical harms of distrust to light and, in doing so, gives us additional insight into the various ways trust adds value to our lives. Finally, the work of Trudy Govier and Meena Krishnamurthy allows us to see the practical advantages and disadvantages of distrust, giving us a deeper sense of how trust and distrust function, and why they matter, in applied contexts. Combined, these philosophers' contributions give us a robust sense of what accounting for distrust involves.

Hawley's 2014 essay "Trust, Distrust, and Commitment", begins with the observation that distrust has been largely undertheorized and the assertion that any adequate account of trust must also be able to account for distrust. She writes that trust is "intriguing" because, despite the popular claim that we should have more of it, trust can be dangerous when misplaced (Hawley 2014, 1). A better understanding of distrust, therefore, is useful not only for understanding what trust is not, but also for understanding when trust is and is not appropriate. Hawley offers a commitment-based account of trust, which aims to avoid treating distrust as an afterthought by using the limits of each concept to explore the conceptual space between trust, reliance, non-reliance, and distrust. Clarity about trust and distrust, Hawley argues, is necessary "if we are to understand the different ways in which trust can go wrong, the reasons both trust and distrust are sometimes unwanted, the nature and limitations of trustworthiness, and the difference between unpredictability and untrustworthiness" (Hawley 2014, 1).

According to Hawley, one reason that we need to understand the relationship between trust and distrust pertains to the requirement I will explore in the next section: doing so gives us a clearer picture of how trust and reliance can be distinguished from each other in a meaningful way. Hawley notes that while the appropriateness of some reactive attitudes (i.e. betrayal v. disappointment) can help us see the difference between "rich trust" and reliance, distinguishing

between the two in this way often does not map on to our ordinary usage of the terms (Hawley 2014, 2). Despite this discrepancy, she maintains that trust and mere reliance should remain conceptually distinct because trustworthiness, unlike mere reliability, "is a significant category for normative assessment" (Hawley 2014, 2). Trust, therefore, is related to our assessments of others' characters in a way that reliance is not. Hawley's argument touches on an important point: when we trust our expectations take on a quality that is different in kind from the quality they have when we rely.

Hawley then argues that motivation-based accounts of trust make the mistake of conceptualizing the distinction between trust and reliance as one where trust equals reliance plus some positive expectation of the one-trusted's motive (Hawley 2014, 5)[40]. She uses Jones's (1996) and Hardin's (1993) accounts as examples of this type. For Jones trust is reliance plus an expectation that the one-trusted's goodwill will extend over the full range of relevant domains or actions that pertain to what they are being trusted with. On Hardin's "encapsulated interest" account, trust is described as reliance based on a belief that the one-trusted is motivated to encapsulate our interests within their own (sometimes simply out of a desire to maintain good social relations with the one extending the trust).

Though neither model offers an explicit account of distrust, Hawley derives one for each of them by inverting the shared formal structure of their accounts. If trust is reliance plus some positive expectation about the one-trusted's motivations, it follows that distrust is non-reliance plus some negative expectations about their motivations. However, according to Hawley, this

[40] Philosophers who offer commitment-based accounts of trust use different categories to sort the various kinds of account of trust available in the literature. Hawley's use of "motivation-based" is an example of one category commonly used to group accounts of trust that take it to be defined by the motivation the one-trusting ascribes to the one-trusted. On the classificatory system I described in the preface the two motivation-based accounts Hawley discusses at length, Jones's affective attitude account and Hardin's encapsulated interest account, fall into the affective and doxastic camps respectively.

would be a mistake because our reasons to distrust one another often go beyond having an expectation regarding their motivations. For instance, anything that makes a person untrustworthy (e.g. knowledge that one's guest is a kleptomaniac or that they tend to tell white lies when it is convenient for them), can merit distrust. It would be irrational not to factor knowledge one has about another's habits (that they steal things, that they are dishonest.) into our judgements about how whether to trust them, even if we believe that they bear us goodwill. To put Hawley's view simply: belief that someone bears us goodwill should not automatically override evidence of untrustworthiness.

Hawley adds that the expectation of ill will also does not necessarily warrant a stance of distrust. To illustrate this point, she asks her readers to imagine they are running for office against a candidate in a close election. In this case, we would be aware that our opponent does not bear us goodwill and, given that our interests are mutually exclusive and opposed (because both of us cannot occupy the same office), that they are not encapsulating our interests. Hawley points out that neither of these facts, on their own, indicate that our political opponent is an untrustworthy person. In fact, although the absence of goodwill and knowledge that our goals directly oppose one another do give us reason to not trust our opponent, our opponent may very well still display traits that are indicative of trustworthiness, such as openness and honesty (Hawley 2014, 6).[41] While we should refrain from trusting them, we also do not have a valid reason to distrust them. In other words, Hawley believes that motivation-based accounts of trust

[41] Hawley makes this claim, but I have some reservations about embracing it myself. To me it appears that are two senses in which we can understand the motivation of our political opponent, and only *one* of them seems to preclude goodwill in the minimal sense I describe in chapter four. "Trying to win the election" and "trying to make sure my opponent loses the election" describe aims that may involve the same actions, but it does not follow that they have the same intention. We associate the former (the opponent trying to win) with characteristics we'd attribute to a tough, but worthy competitor and associate the latter (the opponent trying to make sure that we lose) with a vindictive or inappropriately aggressive opponent—someone we might call a "bad sport". In sum, I think there is more nuance here than Hawley lets on.

come up short when it comes to their ability to explain cases where neither trust nor distrust is an appropriate stance to adopt. The two motivation-based accounts she considers differ with respect to what must be added to reliance to transform it into trust. However, Hawley concludes that despite these differences the form shared by motivation-based accounts of trust collapses the conceptual space between trust, reliance, non-reliance, and distrust.

Hawley then turns to other accounts of trust to demonstrate why they too are incomplete when it comes to accounting for distrust. She focuses primarily on Richard Holton's (1994) account, which distinguishes between trust and reliance by borrowing the concepts of reactive attitudes and the participant stance from P.F. Strawson.[42] Hawley thinks that Holton is on the right track because both trust and distrust involve adopting the participant stance toward the one-(dis)trusted; however, she argues that Holton's lack of attention to distrust leads him to a mistaken implication: "[Holton] seems to suggest that where our trust in someone is limited, then so too is the extent to which we adopt the participant stance to that person" (Hawley 2014, 7). However, as Hawley points out, when we distrust someone, we still hold reactive attitudes toward them. In fact, she argues, we must, otherwise we would be merely not relying on them.[43] Since both distrust and trust require one to adopt the participant stance toward the other, there appears to be something missing. Furthermore, she argues that we can choose to not rely on someone without necessarily distrusting them, and in many cases, she argues, that is the moral thing to do.[44]

[42] Much has been written about Strawson's complex essay "Freedom and Resentment". Hieronymi's recent book *Freedom, Resentment, and the Metaphysics of Morals* (2023) provides a detailed breakdown and analysis of Strawson's argument. In it, Hieronymi argues that expectation and response are the basic units of reactive attitudes.
[43] Hawley, like many others, does not believe that we can trust or distrust inanimate objects on the grounds that they cannot be properly (un)trustworthy (Hawley 2014, 2) Further discussion of this debate occurs in the next section of this chapter.
[44] Hawley uses the example of not relying on her co-workers to buy her champagne throughout the essay.

On Hawley's account, we can see that trust is more than mere reliance and that distrust is more than mere non-reliance When we examine the concepts of trust and distrust together, it also becomes clear that trust and distrust both involve expectations about the one-trusted's willingness as well as their competence. While we can conclude that the participant stance is a necessary condition for both trust and distrust from Holton's account, Hawley argues that it is not sufficient if we want to understand when trust, reliance, non-reliance, and distrust are appropriate stances to adopt toward others.

With this critique in place, Hawley offers her own account of trust: the commitment-based model. On this model, trust is a three-part relation (A trusts B to Φ), where "trusts" means believing that the one-trusted has a commitment to Φ-ing and relying on them to meet that commitment.[45] Conversely, distrust amounts to believing that the one-distrusted has a commitment to Φ-ing, but not relying on them to meet that commitment (Hawley 2014,10). Hawley suggests that her model has the advantage of offering clear criteria for determining when trust and distrust are appropriate. She argues that it is appropriate to (dis)trust another only if they actually have the commitment the one-(dis)trusting believes them to have and contends that this feature of her account allows us to explain how trust can fail without necessarily entailing that the one-trusted behaved badly (Hawley 2014, 13-14). This is possible because others can mistakenly attribute commitments to us that we may not actually have. Though I agree with Hawley that a complete account of trust will also be able to account for distrust and can see the

[45] *A trusts B* to Φ, *A trusts B in domain D* and *A trusts B with valued thing C* are all views are accredited to Baier's observations in "Trust and Antitrust" (1986) and overlap in many ways. When Φ or *domain D* is used, it tends to indicate a slightly broader category—the full range of actions that pertain to what is trusted—whereas *C* indicates a narrower category— particular object or specific action. The broader domain referred to in *I trust him with the kids* could mean the same thing as *I trust him to make sure the kids eat dinner,* or *I trust him in all domains related to caring for the children*. However, the reverse direction is not always true, as it could be the case that *I trust him to make sure the kids eat dinner, but not to care for them otherwise.* For more support of the view that the domain-specificity of trust is explanatorily basic see D'Cruz's "Trust Within Limits" (2018).

appeal of an account that offers clear criteria for determining when trust is appropriate, I have doubts that any rule could be universal or that any set of necessary and sufficient conditions will accurately map on to every case.

One reason for this doubt is that Hawley's account describes trust, reliance, non-reliance, and distrust as if they are on a spectrum that varies along a single axis. This way of mapping the relationship between them ignores another dimension on which forms of trust and distrust vary—their degree of conscious awareness. As we saw in the last chapter, Baier's account allows for unconscious unchosen trust, conscious unchosen trust, and trust that is conscious, chosen, and endorsed. Hawley's account of trust is well-suited for describing the latter—the conscious, chosen and endorsed varieties of trust—and it seems like the same would be true of distrust. While the definition of commitment she uses is a broad one, it would still be a stretch for it to accommodate unconscious, unchosen varieties of trust. On her account, the one-(dis)trusting's trust or distrust is only warranted if the commitment ascribed to the one-trusted is one that they actually have. But in cases where there is no definite object or where the domain of (dis)trust is indeterminate, pinpointing the commitment may involve distortion or be impossible. By indexing trust and distrust to beliefs about the one-(dis)trusted's commitments, Hawley flattens out the multidimensional network via which trust operates. In other words, her account obscures from view the ways in which latent forms trust support the conscious, chosen forms of trust that her account focuses on.[46]

Something else that Hawley's account leaves out is a discussion of how vulnerability factors into both trust and distrust. For Baier, trust is an acceptance of risk—the risk is either latent and never really occurs to us (think everyday forms of trust, infant trust), or we trust in

[46] Thanks to Robert Talisse for pressing me to clarify this point.

spite of the risk, by embracing it and trusting anyway (we'll see a few examples of this later when I discuss Holton's account). This means distrust involves a non-acceptance of or resistance to risk. When we distrust someone the possible ways in which they can harm us are at the front rather than the back of our minds.

Finally, despite claiming to be equipped with criteria for determining when trust and distrust are appropriate, Hawley's commitment-based account does not seem to have a clear answer about how to judge the following case: Your friend agrees to help you move, so he has a commitment, and you are relying on him to meet it. However, when moving day comes, he finds out that his sister has been rushed to the hospital, so he drops everything he had planned for that day (including helping you move) to be with her. Now, Hawley says her account can explain how trust can fail without moral culpability, but only offers examples where a commitment is misattributed. In this case, the commitment is not misattributed, yet it is incorrect to say that your friend is not trustworthy or that you should have distrusted him. Things get worse if we reverse the scenario: everything is the same, except in this case your friend *does* show up to help you move. You find out later that he chose to help you move rather than visit his sister in the hospital. In this variation, it seems like your friend is actually someone you should *not* trust—he clearly has bad judgement about what one ought to prioritize when commitments come into conflict. In short, on Hawley's account it is difficult to discern when trust is appropriate or inappropriate when the one-trusted has multiple, competing commitments. I will say more about the limitations of commitment-based accounts, a subset of the doxastic branch, in the next chapter.

So far, we have seen that the concepts of trust and distrust inform one another—locating their boundaries tells us more about what trust is by clearly differentiating it from what it is not.

However, the fact that the two are different and appear to be in opposition does not give us a robust understanding of what is at stake when we trust or distrust. In addition to providing insights about the conceptual boundaries of each other, another reason why an account of trust should be able to account for distrust is that doing so gives us a clearer picture of the epistemic and ethical significance of both concepts. D'Cruz's extensive work in this area offers several helpful insights. I explore four of them here: the nature of distrust, distrust's impact on interpersonal and community relations, how distrust functions, and how distrust influences our actions.

Like Hawley, D'Cruz maintains that distrust is more than just the mere absence of trust or reliance. He believes that both trust and distrust have normative dimensions and that trust and distrust are mutually exclusive (D'Cruz 2020). He observes that distrust is not as amenable as trust to being modelled on a three-part framework. Typically, distrusting another is a global judgement, or at least one cannot be as narrowly confined in the way that trusting often is. Though it may be possible for such a restriction to exist when the reason for distrust is rooted in beliefs about a lack of competence, what is generally implied about a person's character when they are distrusted makes it more difficult to confine distrust to a narrow domain.[47] Furthermore, our reasons for distrust are complex and often context-dependent (D'Cruz 2020, 43-46).

[47] To illustrate, one might distrust their fifteen-year-old nephew to babysit their infant for a weekend, not because they believe he bears ill will toward the infant, but simply because they do not believe he is competent to care for someone so young for such a duration of time. This seems consistent with trusting him both in general, and even with trusting him in other domains related to caring for the infant (i.e. holding and feeding the baby, or even babysitting for a couple of hours). One might argue this is a case of non-reliance, but I think the point still holds, especially when contrasted with a case of distrust that *is* rooted in a belief or expectation of ill will—we can imagine distrusting an adult who we believe hates children to watch the infant, but it is hard to imagine distrust rooted in this reason not "spilling over" into other domains. It also worth noting that, if forced choose between the two, the one distrusted because of suspected incompetence might be the preferable or reasonable choice.

Unlike Hawley, D'Cruz (2019) understands distrust not as a belief, but rather as a construal.[48] The nature of the construal of the distrusted party impacts the manifestations or phenomenology of how the distrust is experienced. D'Cruz offers the following as potential ways distrust can manifest in the one-distrusting: as skepticism or doubtfulness, contempt or moral disgust, and as fear (D'Cruz 2019, 937). These different manifestations explain why distrust is often felt as insult by the distrusted party. D'Cruz argues that Hawley's commitment-based account does not fully capture what the stance of the one-distrusting toward the one distrusted is like. To illustrate this point, he offers the example of a financier who buys credit defaults. The financier believes the borrowers have a commitment (paying back the loan) and he does not rely on them to meet that commitment. In fact, the financier purchases the credit defaults precisely because he expects the commitment to go unmet (D'Cruz 2020, 45). In addition to expecting the commitment to go unmet, the financier expects to profit as a result. What happens in this case is that we get a weird result: the financier appears to be trusting (or at least relying) on what Hawley's account would describe as his distrust. Though her definition does sometimes describe distrust, it can also accurately fit cases, such as this one, where distrust is not apt description (D'Cruz 2020, 44-45). It doesn't seem right to say that the financier distrusts the borrowers and this fact, D'Cruz suggests, shows that Hawley's account is missing the ethical salience of what the experience of being distrusted is like.

We can start to appreciate its salience by examining how distrusting another impacts the way we interact with them. D'Cruz notes that we will often attempt to hide our distrust, even when it is warranted, due to social pressure. The social pressure to suppress our distrust is related

[48] D'Cruz is careful to note that this construal is *affective*. While there may be cases where its cognitive and affective elements overlap, the affective nature of the construal allows for one to distrust another *without* holding a corresponding belief (D'Cruz 2019, 936-937).

to how revealing it might harm the distrusted person: "distrust…has the power to insult and to wound, [by] sending a signal to the distrusted party and to witnesses that one regards the person one distrusts as incompetent, malevolent, or lacking in integrity" (D'Cruz 2020, 47). The desire to avoid insulting or wounding another might stem from benevolence, but one might also avoid doing so for self-interested reasons, particularly in group settings where their reputation or other social relationships could be adversely impacted. Even though masking our distrust can be difficult, we will often attempt to do so because awareness of distrust can adversely affect the person distrusted and increase social tension.[49] Overt distrust, in other words, carries a distinct social risk.

Other concerning features of distrust includes the fact that it is characterized by an inertia which makes it self-confirming and recalcitrant.[50] Distrust of another can obscure our perception and lead us to interpret the one-distrusted's words and actions in a way that confirms and strengthens our distrust. Distrust is also impacted by looping effects that have been more carefully examined in the context of belief and affect polarization.[51] Discovering that someone distrusts *you* often prompts you to distrust *them*, especially if you interpret their distrust as rooted in prejudice, bias, or some other unjustified reason. What counts as an "unjustified" reason, is of course, a complicated matter. In the case I described at the start of this chapter we saw that both moral and non-moral reasons can justify trust and distrust. Consider the different responses the recovering addict might have to the continued distrust of her friend, who has overcome similar issues with drugs and alcohol, and the continued distrust of her sister, whose partner died from

[49] It is interesting to consider; however, that humans are generally pretty bad at hiding this and good at seeing through others' attempts to do so. For a discussion of nonverbal cues and how they convey trust and distrust, see Baier's essay, "Why Honesty is a Hard Virtue" in *Reflections on How We Live* (2009).
[50] D'Cruz cites Jones (2013) for noting that distrust is recalcitrant and tends to "spillover".
[51] For more on looping effects, political polarization and how they affect trust see Adam Pham and Clinton Castro's work on autonomy and the attention economy in Pham and Castro (2020) and Pham, Rubel, and Castro (2022).

an overdose after unsuccessfully battling his addiction for years. The recovering addict may feel that her friend is culpable for her continued distrust. After all, her friend knows from her own experience that sustained recovery and improved trustworthiness is possible. The recovering addict may have also expected her friend to sympathize with her frustration and disappointment about being distrusted since they share a similar experience. On the other hand, the recovering addict may excuse her sister's continued distrust because her sister has been betrayed in tragic ways by someone like her in the past.

One of the elements I find most compelling about D'Cruz's investigation of the ethical and epistemic hazards of distrust is the attention he gives to how distrust impacts social groups and communities. Though we might distrust in order to minimize risk to ourselves and what we care about, D'Cruz observes that misplaced distrust "[threatens] to put us out of harmony with others in our community deeply and irrevocably" (D'Cruz 2019, 934). Being the object of unwarranted distrust—defined by D'Cruz as distrust not rooted in evidence of ill will, a lack of integrity, or incompetence—often serves to further exclude or marginalize people. It has the undesirable consequence of "planting the seeds of alienation expressed in behavior that does warrant distrust" (D'Cruz 2019, 934). D'Cruz is primarily focused on these passing, non-intimate interpersonal encounters; however, what he says seems applicable to what has been called "social" or "general" trust—the kind we have in strangers and exhibit through our daily habits and interactions without consciously thinking about it. An advantage of D'Cruz's account of distrust is that it lends itself to thinking about how these different forms of trust and distrust affect one another.

Additionally, D'Cruz argues that trust and distrust promote different kinds of dispositions, which, when taken up or expressed habitually, prompt different kinds of actions. In

this way, his approach fits nicely with Baier's later, virtue-centered account. Trust disposes us toward approaching others, toward being receptive and open, whereas distrust disposes us to avoiding them, toward being suspicious or cagey (D'Cruz 2019, 936). We often find ourselves in positions where we must rely on others, regardless of whether we trust them. However, D'Cruz's point that the risks of distrust are often underappreciated, still holds. Even fleeting encounters where one is unwarrantedly distrusted can have a lasting impact "on a person's sense of honor, self-respect, and moral agency", especially if such experiences accumulate (D'Cruz 2019, 934). It is in such situations where the ethical and epistemic harms of distrust are most clearly on display. When we have to rely on someone we trust, D'Cruz argues, we are not only already disposed to rely on them, but we are also willing to "interact closely and to allow ourselves to be vulnerable more generally"(D'Cruz 2019, 935). We risk the harm of being betrayed, humiliated, or dishonored when we trust, but as D'Cruz notes being made a fool of "does not require having made a foolish decision to trust" (D'Cruz 2019, 936). When we distrust, we not only avoid the reliance, close interaction, and vulnerability that we otherwise would have embraced if we trusted, but we also deprive ourselves of discovering how the person would have acted if we had chosen to rely on them (D'Cruz 2019, 936). In other words, when we distrust, we close more doors and make our world smaller. Distrust is accompanied by feelings of wariness and uncertainty that undermine the development of meaningful relationships. This in turn can impede progress on shared projects and goals, and adversely affect the environment in which they are conducted.

D'Cruz's seeming optimism about the goods of trust might lead one to worry that he does not pay sufficient attention to the risks and dangers of misplaced trust. The stakes for misplaced trust can be high, and in many cases where trust is the only reasonable option this is certainly

true. However, even trust that seems ill-advised on the surface typically has a net positive effect on one's community. D'Cruz's forthcoming essay on trust and courage explains this idea through the concept of "trusting overtures". Trusting overtures are extensions of trust that, though they may appear misguided, have the power to restore the capacity or ability to trust in the one-trusted. By acting in a way that presumes the one-trusted to be trustworthy despite a lack of evidence to support the assumption (or even substantial evidence to the contrary), the one making the trusting overture can encourage the one-trusted to see themselves as worthy of trust and respect. This in turn gives the one-trusted an opportunity to respond to that trust, thus making it possible to ignite a trusting relationship.

Furthermore, D'Cruz does acknowledge that trust can be risky, and that appropriate distrust can be empowering. However, he maintains that there are reasons to believe the hazards of distrust outweigh those associated with trust. For instance, D'Cruz concurs with Onora O'Neill (2020) that it would certainly be better if we trusted the trustworthy more and the untrustworthy less, but he argues that her framing of trust's risks may send us down the wrong track (i.e. to thinking about trustworthiness rather than trust).[52] While O'Neill is right that distrusting the untrustworthy and trusting only the trustworthy is certainly an ideal state of affairs, D'Cruz notes that it does not follow that aiming directly at this target is the best way to hit it. Without extensive interaction with others, judgements about trustworthiness, including the trustworthiness of testimony, are typically made on the basis of fleeting first-impressions of faces (D'Cruz 2019, 938). Given what we know about how quickly our judgements about trust are made, how susceptible such judgements are to bias, and the extent to which they correlate with irrelevant features (e.g. the placement of someone's eyebrows, the shallowness of someone's

[52] I offer reasons for rejecting this move in chapter three, when I consider the pitfalls of affective and doxastic accounts of trust.

cheeks, or the color of someone's skin), it is not clear how we could properly signal trustworthiness. Rather than defaulting to the skeptical position of withholding trust until there is sufficient evidence to support the belief that trusting is appropriate, D'Cruz advocates for an alternative: defaulting to a position of "humble trust".[53]

Humble trust, according to D'Cruz is a social practice that "manifests a spirit of skepticism, curiosity and moral commitment" (D'Cruz 2019, 947). It should be noted that the skepticism here is about one's own attitudes of trust and distrust, not about the trustworthiness of others. Someone who practices humble trust "must remain open to modifying their stance in the future when their epistemic position may improve, or their vulnerability may be less" and will appreciate the true weight that accompanies withholding trust (D'Cruz 2019, 948). Humble trust is not the same as full trust or as reliance according to D'Cruz, but rather requires one to be internally "in continuous dialogue with distrust" (D'Cruz 2019, 948). Adopting a stance of humble trust has epistemic benefits too, as it encourages us to gather information and continually reassess our own biases as well as the merits and biases of those we produce knowledge with. All this allows us to see why trust is valuable from a new angle and tells us more about the risks and vulnerabilities associated with both trust and distrust.

[53] Though D'Cruz does not articulate his concept of humble trust using the vernacular of a virtue ethicist, it is worth taking a moment to see how a virtue ethics account offers some explanatory clarity here. "Aiming for the mean" is Aristotle's mechanism for cultivating virtue, but his metaphor of the archer provides further insight. Archers need to be aware of their own tendencies. For instance, if your shots tend to pull to the right, you'll want to aim slightly left of the bullseye. The same advice applies to the cultivation of virtues—you have to know your own habits and correct your aim with respect to them. If you know you tend to be rash, you should veer toward what appears to you as cowardice in order to cultivate the virtue of courage. D'Cruz's concept of humble trust seems to imply something similar. If the goal is to trust (and avoid distrusting) the trustworthy and to refrain from trusting the untrustworthy, and we know that our distrust in others tends to pulled by certain biases, "humble trust" *is* an appropriate way to take their influence into account so that we can better guide our actions so as to increase our chances of hitting our target. Thanks to Diana Heney for raising a question about virtue ethics and trust during the Q&A portion of one of D'Cruz's talks on trust and courage.

While D'Cruz's and Hawley's work makes significant progress in exploring the theoretical side of distrust, it is those who have traversed the practical side of distrust that raise some of the most challenging questions about how to account for it philosophically. Trudy Govier's (1992) essay "Distrust as a Practical Problem" and Meena Krishnamurthy's (2015) "(White) Tyranny and the Democratic Value of Distrust" each complicate the picture by adding additional features of distrust that an account of the concept must consider. In terms of method, these philosophers share the same approach: they look at distrust by attending to it in concrete cases and examine how it has played out in history. However, Govier focuses on the disadvantages of distrust, whereas Krishnamurthy highlights its potential advantages.

Trudy Govier (1992) was among the first philosophers to examine distrust. Her view aligns with D'Cruz insofar as she acknowledges that distrust can serve valuable functions while maintaining that trust is preferable to distrust. She notices many of the same social and emotional harms of distrust we've already discussed but chooses to focus on how distrust poses a practical problem. Her essay is guided by the following question: Given the existence of distrust, which is in many cases warranted, and given that trust is essential for communication and effective cooperative action, how can we move from warranted distrust to well-founded trust?" (Govier 1992, 52). This question is one that a complete account of trust and distrust must be able to answer. Govier is clear that the practical problem runs deep. It threatens to destroy the foundation of expert knowledge[54] and when taken to extremes can "corrode our sense of reality" to the point where the trusting and distrusting effectively live in different worlds (Govier 1992,

[54] See Hardwig's (1991) essay "Epistemic Dependence" for more on why we must trust and not merely rely on the knowledge of experts.

55). This sentiment resonates with the state of many contemporary political climates, and gives us reason to consider how social, historical, and cultural differences impact distrust in practice.[55]

Govier rules out additional surveillance and contracts as possible means for addressing the problem of distrust rather quickly. Surveillance is at best a partial solution because one must still trust the person or technology doing the surveilling. Furthermore, if the one-"trusted" is aware of even the possibility that they are being surveilled, it is likely to increase the uneasy feeling associated with being distrusted. She concurs with Baier that contracts presuppose some level of trust. Furthermore, they may destroy some forms of trust, as writing down or making explicit what is being trusted can sometimes further alienate the distrustful party (Govier 1992, 54-55). Given the limitations of these options, Govier considers alternative proposals for how to move from distrust to well-founded trust. They range from enunciating principles that should be mutually respected (an approach suggested by Bok and Hardwig) and adopting the teachings of Ghandi (who argues we should approach individuals with optimism about their trustworthiness on the grounds that 'brave people disdain distrust') to Osgood's GRIT and the Harvard negotiation approach (Govier 1992, 56-62). Her discussion of these approaches considers the advantages and plausibility of each strategy in depth, but her ultimate conclusion is that while they each may have "something to offer", even when taken together they "fail to constitute a solution" (Govier 1992, 62).[56] None of these methods, she argues, work across the board. By

[55] For more on the effects of distrust in polarized political climates and on what we might be able to do about it, see Talisse (2019).

[56] Most of the essay consists of a survey of different proposals about how to remedy distrust, what follows in this footnote is a summary. Govier argues that Bok's and Hardwig's suggestion of having each party in the distrusting relationship enunciate ethical principles that need to be respected has the advantage of being systematic but tell us nothing about how to get the consensus we need on these principles, nor what to do when we have little to no power over the other party in the trusting relationship. Ghandi's teaching that "brave people disdain distrust" and practice of optimistically approaching all interpersonal interactions with the assumption of trustworthiness seems to ignore the serious harms that can result from misplaced trust. Furthermore, one must share the spiritual values in order to ground this stance, making it difficult to practically defend. She sees therapeutic trust as a way of providing "moral space" but has reservations about the extent to which therapeutic trust involves pretense, as it can easily be

trying to offer an abstract set of rules that can be applied to every instance these proposals fail to appreciate the historical and social context as well as the particularities of the relationship that shape the distrust in each case. She concludes that to find a solution to the practical problem of distrust we must "gain a real appreciation for what distrust and trust amount to in concrete terms" (Govier 1992, 62). Her recommendation to turn to practice suggests that an adequate account of trust and distrust must be flexible enough to offer guidelines that can be tailored to the concrete situation at hand.

Meena Krishnamurthy (2015) looks at distrust from a different angle. She argues, contrary to prevailing views that distrust is a threat to democracy, that distrust actually has an important democratic value, which lies its ability to temper tyranny.[57] Krishnamurthy's account of distrust is deeply informed by the history of racial injustice in America. Her target, accordingly, is white tyranny, which could be characterized as a set of entrenched and interlocking practices that systematically privilege white people. Her account of distrust builds on the work of Martin Luther King Jr. and construes distrust as a narrow normative concept where "in order for A to distrust B, A must have a confident belief that B will not act justly"

---

construed as manipulation by the distrusted party, which would then deepen distrust. The GRIT approach proposed by Osgood involves announcing a goal of reducing tension and reliably carrying out a conciliatory act. She argues that it takes a step in the right direction, as voluntarily making oneself vulnerable and displaying openness can promote the attribution of trustworthiness. However, while this method might be fit for collectives, it does not seem as applicable to personal relationships. Furthermore, the approach seems to underestimate the difficultly of determining what a "conciliatory act" looks like to both parties. Finally, the Harvard Negotiation approach offered by Fischer and Brown recommends focusing on how individuals can modify their own conduct to increase the likelihood of well-founded trust. While they offer good advice—be predictable, speak and make commitments carefully, avoid deception, keep secrets and so on— and suggest there is a golden mean when it comes to trusting others, their account does not offer any guidelines for determining when we are trusting too much or too little. As we can see from this gloss, Govier is charitable in her analysis of these various solutions, but careful to note their limitations.

[57] Krishnamurthy sees her view as contrary to Patti T. Lenard (2015), Shayla Nunnlay (2005) and Sandra Susan Smith (2010), who argue that distrust is a threat to democracy. She also cites *Democracy and Trust,* a volume edited by Mark Warren (1999). His own contribution to the volume discusses the value of trust in deliberative forms of democracy. Warren (2017) builds on the same ideas by making a helpful distinction between second and first order political trust, which allows for trust in the democratic process to support deliberation on contentious matters where distrust is warranted.

(Krishnamurthy 2015, 392). Distrust has democratic value, she argues, because it can protect marginalized and oppressed groups from experiencing further injustice and assist them in resisting it.

Krishnamurthy argues that distrust of fellow citizens and political institutions during the civil rights movement was instrumental in the creation of forms of political expression—sit-ins, boycotts, peaceful protest etc.—which "were essential to securing greater racial justice and a more genuine form of democracy" (Krishnamurthy 2015, 392). She specifically cites King's distrust of white moderates which, she argues, was not rooted in an expectation of ill will, but rather stemmed from an expectation of inaction based on empirical evidence drawn from his lived experience (Krishnamurthy 2015, 394-395). By indexing her analysis to a specific historical moment, Krishnamurthy accomplishes what I take Govier to be calling for: an "on the ground" account of distrust that provides insight into how it functions.

Krishnamurthy's account of distrust distinguishes between two forms of distrust—horizontal and vertical. Horizontal distrust describes distrust between individuals as fellow citizens qua citizen, whereas vertical distrust describes distrust citizens have toward other citizens as instruments of political institutions. For instance, horizontal distrusting would be when you distrust your white neighbor because you fear he is a white supremacist, and vertical distrust would be when you distrust your neighbor because he is a cop (i.e. instrument of law enforcement) and you fear that white supremacy is imbedded in institutions of law enforcement. In this example, assuming Krishnamurthy's chosen context of the South during the 1960s Civil Rights Movement in the United States, this vertical trust was certainly warranted, and, depending on one's geographic location, racial identity, social position, and past experiences, the horizontal distrust likely may have been as well. What this distinction brings out is that an individual's past

experiences as well as the historical, social, economic, and cultural context in which they are imbedded can provide additional reasons why distrust and trust may or may not be warranted that are independent of an assessment of that person's trustworthiness of other relevant character traits. Both are important to understanding the consequences and value of distrust and provide helpful terminology for describing different forms of trust. Being able to describe different forms of trust in this spatial way frees us from relying solely on speaking of trust in institutions, governments, and other groups in terms of levels, scope, or object. In this way, Krishnamurthy's work strikes me as offering essential conceptual tools for thinking about how trust might operate, as Baier suggests, as a web, network, or climate.

Warranted distrust can protect oppressed and marginalized groups from further harm and can be useful for mobilizing groups to act in ways that garner support from complacent parties. Her work also suggests new ways we can think about how distrust functions as a survival or coping mechanism at the individual level. An individual who either repeatedly experiences failures of trust or a traumatic event that causes them to lose their "trust in the world", has good explanation about why they distrust or avoid trusting others, though they may not be able to articulate their explanation in the form of reasons that are rational to other, non-traumatized individuals The context dependency of when trust and distrust are valuable has led some philosophers to examine trustworthiness rather than trust itself—the idea being that trust and distrust are appropriate only when the one-(dis)trusted is (un)trustworthy, either in the relevant domain or generally. However, when we look at trust and distrust in practice, it becomes clear that more than just than (un)trustworthiness of the entity one is (dis)trusting is relevant when it

comes to deciding whether extending trust is appropriate. One's past experiences as well as features of one's environment are also pertinent factors. [58]

Before I move on to the second requirement, I want to summarize the thoughts in this section. An account of trust should also be able to account for distrust; the concepts inform each other. As we saw in Hawley's work, a more precise understanding of what trust and distrust are and are not clarifies the boundaries of the conceptual space they occupy. From D'Cruz work, we saw that it is advantageous for an account of trust to be able to account for distrust because the epistemic and ethical harms of distrust bring the multiple dimensions of trust's value and its specific kind of vulnerability into focus. Finally, from Govier's and Krishnamurthy's discussions it is clear that understanding how distrust functions "on the ground" is necessary if we want our account to be capable of saying anything about how to foster, maintain, and repair trust.

Requirement #2: Making a Meaningful Distinction Between Trust and Reliance

Many philosophers agree with Baier that what makes trust philosophically interesting is that it is meaningfully different from reliance. Attempts to explain the difference between the two concepts often imply that other moral agents are the only appropriate objects of trust. Therefore, it is commonly assumed that the entities we are able to trust are limited to human beings and groups or institutions comprised of them. However, some philosophers have questioned this restriction.[59] Furthermore, in the case of groups and institutions, there has been some debate over whether or not it makes sense to insist on maintaining a sharp distinction between these two concepts. In this section, I review several ways this distinction has been made and intervene in

[58] It is unclear to me how thinking about trustworthiness or how framing trust as primarily a matter of belief can explain when trust, distrust—or something in between—is appropriate, warranted, or justified. I will say more about why I believe a developmental account of trust is capable of offering, if not specific criteria, at least guidelines about when (dis)trust is healthy in chapters four and five.

[59] Notably, C. Thi Nguyen who argues that we can trust objects. I say more about his account shortly.

the debates that have been sparked by the numerous answers to this question. I argue that an understanding of the multiple way trust contributes to human flourishing is key to making a meaningful distinction between trust and reliance. An account of trust must be able to make a meaningful distinction between trust and reliance for three reasons: 1) the distinction reveals what entities are and are not appropriate objects of trust; 2) the distinction brings us closer to understanding the nature of the vulnerability trust exposes us to; and 3) the distinction gives us insight into why trust is something we value.

The concepts of trust and reliance largely overlap. They both involve future-oriented expectations, play an important role in day-to-day and long-term planning, and make a variety of activities possible. If we could not trust or rely, we would have to do everything ourselves and, simply put, we would not have time for much beyond procuring nourishment and finding shelter. Baier initially frames the distinction between trust and reliance as one that hinges on the presence of goodwill. Recall her first definition of trust in "Trust and Antitrust": "Trust, then, on this first approximation is accepted vulnerability to another's possible, but not expected ill will" (Baier 1994, 99). In her earliest account of trust, Karen Jones follows suit by adding that trust is "optimism about the goodwill and *competence* of another" (Jones 1996, 7 emphasis in original). She specifies that this competence can be technical in some cases (i.e. trusting a plumber with the pipes) but that in other cases, such as friendship, what we expect "is a kind of moral competence" (Jones 1996, 7). Carolyn McLeod offers another variation of Baier's initial distinction. She argues that the one-trusting expects the one-trusted to demonstrate something like moral integrity, rather than goodwill and competence (McLeod 2002, 6). However, many

philosophers have offered counterexamples that raise doubts about whether the expectation of goodwill or a morally "good" motive is necessary for trust.[60]

Richard Holton's 1994 essay "Deciding to Trust, Coming to Believe" takes Baier's observation about the significant difference between the emotional and psychological responses we have to betrayals of trust and the way we respond to failures of reliance as the starting point for his account of trust. While he does not think that betrayal is the only emotional response one could have to a failure of trust, he believes attending to it puts us on the right track. Betrayal is one example of what P.F. Strawson calls a reactive attitude. Following Strawson, Holton claims that we have a readiness toward reactive attitudes like betrayal when we trust, because trust requires taking up the participant stance (what Strawson calls the participant attitude) toward whomever one is trusting (Holton 1994, 66).[61] Readiness to reactive attitudes is only "partially constitutive" of the participant stance according to Holton. Fully understanding it requires engaging with a network of other attitudes, actions, and beliefs. On Holton's reactive attitude account, trust is distinct from reliance because of the stance one takes toward the trusted: "in

[60] Some examples that have been offered to prove this point include the confidence trickster, Amy Mullin's (2015) chess partner and civil adversary examples, Matthew Bennett's (2021) fading friendship case, and Hawley's (2014) political opponent example, but there are others. I have reservations about these examples, which appear to set the bar for what constitutes 'goodwill' too high. We should be reluctant to let go of the connection between trust and the expectation of goodwill so easily, as it is possible to define goodwill minimally, as Aristotle does in the two books on friendship in the *Nicomachean Ethics* (see Book IX, chapter v in particular). In his forthcoming essay, Jason D'Cruz acknowledges something akin to this thought, noting that the presence or absence of goodwill *does* make a difference, especially in our most intimate, interpersonal trusting relationships. I agree with his point, but I am willing to take a slightly stronger stance and say that an expectation of minimal goodwill underlies the vast majority of our trust relationships, including those where trust is restricted to a fairly narrow domain (i.e. trusting the plumber to fix the sink) and the fleeting encounters associated with "social trust" (i.e. trusting other passengers on the bus not to attack you). Without goodwill, trust in the everyday sense is too easily reduced to the form of trust seen in a contract or promise, which, as we saw in the last chapter, are better understood as limit cases than as the paradigm. I expand on this thought in chapters three and four where I will argue against the view that goodwill should be dropped from our account of trust and suggest that a latent expectation of goodwill characterizes *healthy* climates of trust.

[61] It is important to note that Holton is aware he is taking some liberties with Strawson's actual view here. For instance, on Strawson's account reactive *attitudes* are appropriate when the other party is responsible for their actions. The feeling of resentment, or in this case betrayal, stems from feeling as though one has been disrespected and not treated with the proper regard for the kind of they agent that they are (Strawson 1974).

short, you take a stance of trust toward the person on whom you rely" (Holton 1994, 67). A consequence of distinguishing between trust and reliance in this way is that the appropriate objects of trust are restricted to those we can justifiably have reactive attitudes towards, i.e. other moral agents or human beings.

Jones, in both her early and late accounts of trust, affirms this restriction. In her earlier account of trust, she argues that trust requires the expectation of goodwill. Therefore, we can only merely rely on entities such as machines, which do not have wills of their own, but we can trust other human beings because they do have their own will and, therefore, are capable of bearing us good or ill will (Jones 1996, 14). In her later work, when she abandons the goodwill requirement, Jones expresses a similar idea in terms of trust responsiveness. Trust and reliance in her later account are distinct because trusting others is necessarily a way of responding to interpersonal dependency, whereas reliance is not (Jones 2012, 62). This version makes it plausible to think about trust in groups and institutions, but still maintains that trust and reliance are distinct because trusting is restricted to agents (individual or collective) that have the capacity to respond to the dependency signaled by the one extending trust.

What I find compelling about the accounts offered by Holton and Jones is that both acknowledge that beliefs alone cannot explain why trust and reliance are distinct phenomena, or why the ability to make a distinction between them matters. This is not the same thing as saying that beliefs are irrelevant when it comes to trust, but rather that one's beliefs are not the only factor that should be considered. By recognizing the affective components of trust, these accounts allow us to see another way the distinction between trust and reliance can be made. We can have reasons to extend trust that don't apply in cases of mere reliance. For instance, one might hesitantly extend trust for the sake of improving a relationship or the atmosphere of trust

within a group, but the same is not true in cases of mere reliance (Holton 1994, 69). To illustrate this point Holton offers the example of a rock climber who, instead of relying on an inanimate anchor point, chooses to trust their new rock-climbing partner to belay them for the sake of developing a trusting relationship. By locating the difference between trust and reliance in terms of how they affect our relationships with others, we can see why trust is essential to cooperation and interpersonal relationships in general.

The way affective accounts of trust distinguish between trust and reliance and the resulting assumption that we can only trust other people has an intuitive appeal but, the position does have its critics. As we saw in the previous section, Hawley (2014), believes that while reactive attitudes and Holton's idea about the participant stance can help explain the difference between trust and reliance, the account of trust he offers is incomplete because the readiness to reactive attitudes alone cannot entirely explain why trust and reliance are distinct in a meaningful way. In a later essay, Hawley (2017) claims that the reactive attitudes that characterize trust in interpersonal relationships do not easily map on to the way trust operates in groups and organizations. Unlike individuals in interpersonal relationships, what makes a collective entity trustworthy can often be reduced to its reliability. Furthermore, she claims that when we interact with a member or representative of a group, it is unclear if our attitude of trust is directed toward the individual, the group they are associated with, or some mix of the two (Hawley 2017, 23). While Hawley's contention pushes us to think more about how to make sense of the difference between trust and reliance, empirical research shows that individuals do report feelings like resentment and betrayal toward groups and institutions and that they are inclined to hold such entities morally responsible in the same way they do in interpersonal relations (Pouryousefi and Tallant 2022). Pouryousefi and Tallant argue that these findings appear to undermine

commitment-based accounts of trust like Hawley's and to support reactive attitude accounts. Furthermore, their analysis notes that the survey questions did not ask about trust in these groups and institutions in terms of a three-place predicate, reaffirming that we often trust without specifying a definite candidate for "C". There is more to investigate when it comes to trust and reliance in the context of groups and institutions—too much for me to explore here in detail here. However, whether it makes sense to insist on the distinction between trust and reliance in such contexts seems to depend on how one answers questions about whether groups or institutions can be said to have character, agency, or a will.

A second criticism of Holton's view comes from C. Thi Nguyen. He agrees that the distinction between trust and reliance can be made based on the presence of reactive attitudes but challenges the implied conclusion that trust relations are restricted to other humans or human-comprised entities. Rejecting the assumption that the entity one trusts must also be a moral agent, Nguyen argues that the stance one takes toward certain inanimate objects can also reasonably be described as one of trust. On his account, trust is an unquestioning attitude about whether the trusted entity will properly perform its function: "to trust something is to rely on it, without pausing to think about whether it will actually come through for you" (Nguyen 2022, 2). Nguyen is not arguing that we can trust just any old object, but rather that the relation between ourselves and the objects we have integrated into our agency is more aptly described by trust than by mere reliance. He has in mind particular forms of technology, such as cars and smart phones, but believes his account of trust could theoretically apply to any object that has become functionally integrated into how we move about the world. Objects that we do not integrate into our agency in this functional way—e.g. the shelf we rely on to support the vase— still fall into the category of mere reliance.

Nguyen supports his claim by appealing to the trust we have in parts of our own bodies to function properly— e.g. our hands or our memory. When our memory fails us or when our uncontrollably shaky hands cause an error, he argues, we experience the feeling of betrayal—the same kind of feeling we have when another person fails to meet our trust (Nguyen 2022, 7-8). This feeling still characterizes betrayals of trust and is what makes trust distinct from reliance on his account. Nguyen is clear that the form of trust he is discussing here—trust as a mechanism for expanding our agency by integrating other people and objects into it—is only one among many forms of trust (Nguyen 2022, 5). However, he contends that this form captures how the term is most frequently used.

I'm sympathetic to Nguyen's idea that the way we relate to some objects does seem to succeed in picking out cases of trust. In particular, I think it appropriately describes what self-trust is typically like. Furthermore, I find that his articulation of trust as an unquestioning attitude picks up on a feature of trust Baier identifies that typically does not receive much attention—the tendency for many forms of trust to go unnoticed until they fail. While I would set the bar for trusting objects higher than Nguyen (i.e. I am skeptical that "trust" is the right way to describe the average person's stance toward their smartphone), I do think that the standard of an "unquestioning attitude" is a good one for those designing certain technologies to adopt as a regulative ideal. In particular, it seems laudable to aim at designing technologies that can assist medical researchers or doctors in diagnosing and treating illnesses in such a way that they can be trusted in this unquestioning register. The same goes for devices that assist those with illness and disabilities—we should want those who use prosthetic limbs and glucose monitors, for example, to be able to adopt an unquestioning attitude toward such devices because doing so seems like a way to design environments that enable their ability to flourish, exercise autonomy, etc.

However, I do have concerns regarding the assumptions about trust that are reinforced and highlighted on such an account, particularly those that pertain to the kind of risk and vulnerability that characterize trust.[62]

On Nguyen's account, the risk associated with trust is tied to the expanded agency that trust enables. Nguyen argues that his account of trust helps us "reunite our discussion of trust with concepts of intimacy", but he describes intimacy in terms of agential integration, as letting something in so that it can become unified with oneself (Nguyen 2022, 38). While unification is one way of describing intimacy, it is not the only way, nor does it seem appropriate for describing the intimacy we desire in many of our important interpersonal relationships. For example, in the intimacy we see in friendships, unification is not the goal, nor does friendship derive its value from the expansion of our agency.[63]

---

[62] Two other objections are worth considering here. For one, it seems like *some* people may experience or associate such feelings with objects, but it also seems possible that some people do not have these feelings at all. Second, one might object to the idea that we should design technologies to be *trustworthy* as operating in this register could erode our epistemic capacities. I'm not sure what to say about the first problem—it does not seem like there is anything *wrong* with people who do not trust objects even when they are integrated into their agency, but it also does not seem like much hangs on this in Nguyen's account so long as *some* people do report feeling this way. I am sympathetic the concern that is motivating the second objection; however, I think it misses the point. Hardwig (1991) explains that knowledge always requires trust, especially in contexts where one is working with a team that includes experts from many different disciplines. In a medical context, this might mean that a primary care doctor, who is unsure about what is causing her patient's stomach pain, trusts gastroenterologist to give an accurate diagnosis. If we replace the gastroenterologist with a machine that has the same (or better) rate of accurately diagnosing the problem, it seems reasonable for the primary care doctor to *trust* the special diagnosis machine the way she trusts the gastroenterologist. The kind of trust I describe above as laudable does not entail that the doctor must trust the diagnosing machine beyond the way she trusts the specialist, which means she should check twice if she has hunch that the machine is malfunctioning or reason to believe the sample it tested has been tampered with just as she should if she believed the specialist made a mistake or contaminated the sample. Such a stance would be more like "blind faith" than trust and it would be epistemically irresponsible for the doctor to adopt it toward either the gastroenterologist or the diagnosing machine. Well-founded trust does not lead to the erosion of our epistemic capacities, though poor judgement about who or what is trustworthy may.

[63] To illustrate, we can compare Nguyen's characterization of intimacy as unification to the intimacy described in philosophical accounts of friendship. When we do, we see that Nguyen's account seems like it could apply to the friendships of utility Aristotle describes but would be less applicable to Aristotle's account of virtuous friendships. More recent accounts of friendship, such as the one Daniela Dover (2022) sees as ideal for "interpersonal inquiry" implies that the intimacy that is possible in our friendships is essential to the formation of our own sense of identity. Cathy Mason's (2021) Murdochian-inspired account of friendship as inherently knowledge-involving suggests that the intimacy in friendship is a matter of having deep knowledge about our friends, which comes from seeing them as they truly are. Diane Jeske's extensive work on friendship suggests that friendship offers us a different kind of reasons, ones that reveal the inadequacy of standard categories of reason seen in moral philosophy. She argues that

Exposure (in the vulnerability sense, not the frequency sense) rather than unification, offers a more appropriate image of the kind of intimacy that characterizes trust in close interpersonal relationships. Failures of trust do seem to differ from failures of reliance on the basis of the affective responses that each illicit. However, this is only part of the story. Another part seems to be the degree of unpredictability associated with each stance. When we rely on something, we tend to know in advance (or could at least predict if asked) how the entity we are relying on might let us down. When we trust, our ability to make such predictions is less precise because there is more uncertainty. This uncertainty seems to stem from the fact that the entities we trust are capable of acting of their own accord, so the number of ways in which our trust may (or may not) be met is too great to plan for or imagine. More importantly though, there is also a sense in which trusting, especially in intimate interpersonal relationships, precludes such contingency planning from the outset. An anxious or skeptical person might do so regardless, but if the one-trusted were to discover that the person trusting them had a back-up plan all along they would (rightly) feel as though they were not trusted and perhaps even distrusted. Such feelings can weaken the strength of a relationship and raise doubts about whether there ever was trust in the first place. D'Cruz states this idea with conviction: "avoiding reliance and vulnerability by hatching elaborate contingency plans is *incompatible* with trust (D'Cruz 2015, 473, emphasis my own).

reasons of friendship or reasons of intimacy are distinct from subjective Humean reasons and objective consequentialist reasons because they are found in the relationship itself (Jeske 2001,2002). She calls this the "common sense" view of friendship and argues that, as result of stemming from our relationships themselves, the grounds of our reasons related to friendship can compete with moral reasons in a way that dissolves the puzzle of how morality could permit and even demand that we give special consideration to our friends (Jeske 2018). While these different accounts of friendship are inspired by disparate methodological approaches and meta-ethical commitments, they share the idea that the intimacy involved in friendships play a role in deepening one's knowledge about oneself. This runs contrary to the conception of intimacy as unification, as the ability to distinguish between oneself and one's friends is necessary for development of self-knowledge.

When we focus on the intimacy associated with trust with an eye toward how it makes meaningful relationships possible rather than toward how it enables the expansion of our agency, the moral rather than the epistemic dimension of trust are emphasized. D'Cruz, following Holton, argues that in interpersonal relationships trust serves as more than a mechanism for facilitating cooperation—it also functions as a means of signaling respect and acknowledging the moral standing of another (D'Cruz 2015, 469). This insight suggests that the kind of vulnerability we're exposed to when we trust is what makes trust distinct from reliance.

In the contemporary trust literature, the value of trust has been discussed primarily in terms of its ability to expand our agency. It is true that trust makes cooperation more efficient, and that cooperation allows us to do more than we could do without the assistance of others. This value is highlighted when we look at trust's epistemic dimensions. However, Baier's claim that "some degree of trust in the social world is the starting point and very basis of morality" suggests that there are more dimensions to trust and additional ways in which we can understand its value (Baier 2009, 179).

What other dimensions of trust's value are there? We see one in D'Cruz's work: trust enables the intimacy that makes meaningful relationships with others possible (D'Cruz, forthcoming). Another stems from Johnny Brennan's (2021) argument that there is a connection between basic forms of trust and recognition—a concept often understood to undergird the possibility of morality and seen as necessary for moral respect.[64] Finally, while both of these are sources of value on their own, combining the two observations reveals another potential source of value in trust: the role it plays in the formation of identity. These additional dimensions of

[64] In "Two Kinds of Respect", Darwall convincingly argues that recognition respect is the kind of respect we owe to other persons. He suggests that this is the kind of respect both Rawls and Kant appeal to in their work (Darwall 1977, 36). Axel Honneth's *Struggle for Recognition* is another influential account that Jay Bernstein cites in his account of trust, which we will examine in chapters four and five.

trust's value are obscured when the distinction between trust and reliance is described only in terms of affective dispositions towards and beliefs about others. Unlike reliance, trust requires us to adopt a different stance toward the kind of uncertainty, risk, and vulnerability that characterizes it.[65]

Requirement #3: Attending to the Way Trust Functions in Relationships with Asymmetries of Power

Despite the attention Baier gives to forms of trust in relationships with stark differences in power in her germinal essay, it has not remained among the top concerns for most trust theorists. Most accounts of trust have continued to assume relative equality of power between the one-trusting and the one-trusted. The fact of the matter though is that most relationships, both in general and in those where trust is present, do have asymmetries of power—some significant and some more subtle. These asymmetries are not only prevalent but are also desirable in many of our most important relationships. Thus, trying to eliminate or mitigate these asymmetries is not a proper goal for an account of trust. Given that asymmetries of power are both common and not always undesirable in relationships where trust is found, we already have substantial reason to demand that an account of trust should attend to how trust functions in such cases. In this section, I will first draw from philosophers who have attended to the power dynamics involved in trust. I will examine several examples of relationships that benefit from trust and involve an obvious asymmetry of power. The purpose of this latter part is to indicate why the third requirement is essential for an adequate account of trust, particularly if it is to be useful in applied contexts.

[65] In chapters four and five, I will argue that appreciating these other dimensions of trust's value is necessary if we are to understand the relationship between trust and trauma. Furthermore, doing so will also be helpful for intervening in debates about whether the trust and reliance distinction is applicable in cases of group, institutional, and political trust.

Nancy Nyquist Potter follows Baier in noting that trust inherently alters positions of power. She attributes this to two facts about trust 1) that trust is inherently relational and 2) that interpersonal relationships are always already infused with power dynamics, regardless of whether they are "thick" or "thin" (Potter 2020, 243-244).[66] In other words, Potter argues that trust involves vulnerability and therefore involves power dynamics by default. By its very nature trust involves an asymmetry in power because the one-trusting gives up power to the one-trusted. However, the difference in power that results from extending trust is often an addition to other power asymmetries that already exist between the two parties. These differences in power, coupled with the specific risks that accompany various instances of trust explain why it matters to trust wisely (Potter 2002, 20). Interpersonal trust, Potter argues, is always imbedded in a social context, which means it is impacted by the "social imaginary" that circulates representations which shape our perceptions of others. She also describes this context as a "soup" of social, political, and economic structures and notes that these factors also shape our identities (Potter 2020, 250; 2002, 18-19). Potter, like Baier, notes that interpersonal trust is often a matter of degree, that it has affective as well as cognitive, embodied, and material dimensions, and that it is mediated by conversational and other implicit norms (Potter 2020, 244-248). Since trust and the ability to trust can be damaged when one is betrayed, ruptures of trust often call for repair and forgiveness; therefore, maintaining trust requires moral effort (Potter 2020, 2002). Potter's understanding of interpersonal trust attends to the complexity of trust in a way that allows us to

[66] By "thick" she is referring to intimate, long-term relationships and by "thin" she has in mind momentary encounters between strangers or acquaintances Niker and Specker Sullivan (2018), who I will discuss shortly, follow suit.

see that while there may not be a moral requirement to trust, creating the conditions under which one can live a flourishing life involves doing one's part to cultivate trust when it is appropriate.[67]

Potter is also skeptical about the possibility of finding universal rules that can guide our trusting because she believes that genuine trust is fundamentally particular. She offers the example of trusting one's doctor and argues that we want the goodwill of our doctor to be directed at me as their patient, and not to be of a more universal sort (Potter 2002, 8).[68] Without attention to particularities, trust lacks the connection to loyalty and identity that makes it ethically valuable. Therefore, she suggests that we eschew the principle of impartiality when it comes to trust. To become trustworthy, Potter argues, we must develop a sensitivity to the particularities of others. Invoking the work of Iris Murdoch, she adds that moral effort and imagination are key to attaining the clear vision that makes it possible to attend to others in this way (Potter 2002, 27-28).

While Potter focuses primarily on interpersonal trust, her concerns about asymmetries of power extend beyond dyadic relationships and into social and political realms:

> A society that aims to be genuinely and deeply democratic must take on questions of trust and distrust in a way that transforms structural and symbolic relations of power to ones of nondomination and nonexploitation. (Potter 2002, 20)

Potter is one of the few philosophers who attends closely to the asymmetries of power in trust and the insights her work offers—namely the observation that trust and identity are interrelated—are not be possible unless we attend to the ways that interpersonal trust and broader forms of social trust affect one another. Thus, Potter's work demonstrates that attention to

[67] In her 2002 book *How Can I Be Trusted: A Virtue Theory of Trustworthiness*, Potter offers an account of trust modelled on a 4-part, rather than 3-part relation: A *trusts* B *to be* x *sort of person with regard to* y (Potter 2002, 17). She argues that we can accomplish the task of cultivating appropriate trust by developing a trustworthy character.
[68] One could argue that this desire to have the doctor's goodwill directed at me as a particular individual person could be met by generic beneficence. I think there is a strong case to be made for this point; however, there are other reasons to value trust in a doctor-patient relationship. I describe some of them later in this chapter.

asymmetries of power is an essential component for all accounts of trust, including social and political forms, and not just the interpersonal ones. Attending to the complexity and interaction of variables that lead to power asymmetries reveals that our environmental and material conditions are another factor that we need to consider when evaluating when trust and distrust are appropriate.

Fay Niker and Laura Specker Sullivan (2018) also focus on asymmetries of power in interpersonal forms of trust in their essay "Trusting Relationships and the Ethics of Interpersonal Action". They argue that in our thick trust relationships, trust is better understood as an irreducible quality of the relational environment than as a stance one party adopts toward another (Niker and Specker Sullivan 2018, 5). Niker and Specker Sullivan, who explicitly describe themselves as building on Baier's view that trust operates via webs and networks, argue that conceptualizing trust as a property of the relationship, rather than as a stance one person adopts toward another more accurately captures what trust is like in our enduring, intimate relationships. An additional advantage of conceptualizing trust as a property of a relationship is that it helps us avoid characterizing trust as something that is dependent on the mental states of both parties in the dyad: "Just as we would not characterize a 'good relationship' by referring to the attitudes that either party has for the other, so should we refrain from describing trusting relationships in terms of the trust that one party has in another" (Niker and Specker Sullivan 2018, 2). They argue that their account of thick trusting relationships allows us to see how knowledge and care are intertwined in trust, as well as why asking for permission or consent to take care of the person or things that pertain to the relationship might undermine or weaken trust. This undermining happens because making what's implicitly entrusted explicit can be interpreted as

evidence that the mutual understanding that they believed to already characterize their relationship is not there or is not as mutually understood as they had previously expected.

The shift in how Niker and Specker Sullivan conceptualize trust illuminates several of its important features.[69] One key feature is that power asymmetries are liable to fluctuate and shift over time, particularly in relationships that involve care and intimacy. An account of trust needs to be able to accommodate the fact that trust is dynamic if it is going to adequately explain how trust operates in many crucial relationships. Thinking about interdependency can lead us to a deeper appreciation of the other dimensions of trust's value—those connected to intimacy and identity. Because the asymmetries of power and direction of trust in our relationships fluctuate over time, we need to attend to the particular histories of changes in trust that shape our intimate, long-term relationships.

Before I conclude this chapter, I want to consider a few types of relationships where the presence of trust is beneficial and where asymmetries in power are obvious: parent-child, teacher-student, and doctor-patient. I've selected these relationships because they are ones that almost everyone experiences at least one side of during their lifetime. While not every person will become a parent, everyone has been a child. Similarly, while not everyone will have experience with being a doctor or teacher, at some point in our lives it is likely that we will find ourselves in the position of a patient or student. Furthermore, most of us have (or will have) the experience of caring for someone who is ill and of teaching another at some point in our lives, albeit in a less formal capacity. I argue that trust not only plays a faciliatory role in these

[69] I am inclined to believe that atmospheres or climates of trust are most helpful for thinking about how interpersonal trust and social/political trust are interrelated, which while not necessarily incompatible with their account, would require additional philosophical work to sort out. I also think that while they do a better job than most of giving Baier's work the attention it is due, they overlook the extent to which she is concerned with how different forms of trust are interrelated, especially in her later work.

relationships, but also that the presence of trust enhances these relationships in a way that brings it closer to regulative ideal of what these relationships should be like. In addition, these three relationships each bring into focus the ways trust contributes to human flourishing that I've been pointing to throughout this chapter.

Parent-child relationships have the starkest difference when it comes to relative power. While the gap closes in some ways as the child ages, parent-child relationships show us most clearly that one reason we value trust is because it makes connecting with others possible.[70] Power asymmetries in parent-child relationships are more clearly seen when we focus on the multitude of ways in which the child is dependent on the parent. This includes basic needs for survival—food, shelter, —but there are also more subtle, crucial forms of dependency. Children depend on their parent(s) to learn social norms and practices, including language, which are essential for a flourishing life. Children who grow up without a stable trusting relationship with their primary caregivers often struggle to trust in their other relationships later in life (van der Kolk 2014, 162-163). Aside from the ability to facilitate interpersonal connection, a trusting relationship between a child and their parent is also crucial to the development of the child's agency. Self-trust, a topic which we will explore more in-depth in chapter five, seems to develop alongside one's sense of agency.[71] The capacity to act as an agent and form meaningful

[70] I want to be clear that what I am saying is *not* about biological parent-child relationships. In what follows, anyone who occupies the parental role could potentially be the source of the trusting relationship, though the degree to which the child recognizes them to be filling that role complicates the matter in varying degrees in particular cases. For instance, compare the following: the low difficulty that a child of a closed adoption raised solely by her adoptive mothers has in recognizing her mothers as filling the "parent" role, despite full awareness that she is not biologically related to either of them, relative to that of a child who, lives on and off with his biological parents and in the foster care system, but has received most of his physical care and social interaction from an aunt and uncle who live nearby. The point is, although I am taking the 'child-parent' case as an example of a relationship that most people have, a lot hangs on what a society considers a deviation from the definitions of "parent" and "child" and what the experience of the child is actually like. We must remain open to diverse and expansive definitions of these terms for other, free-standing reasons.

[71] Victoria McGeer's (2008) essay "Trust, Hope and Empowerment" is often taken as the starting point for thinking about self-trust and the role the therapeutic trust plays in its development. It is telling that the example frequently

connections with others are essential to the formation of one's identity or sense of self. Bessel van der Kolk captures this idea eloquently:

> In order to know who we are—to have an identity—we must know (or at least feel that we know) what is and what was 'real'. We must observe what we see around us and label it correctly; we must also be able to trust our memories and be able to tell them apart from our imagination (van der Kolk 2014, 136).

While these abilities can be affected by a number of experiences and relationships, parent-child relationships have the greatest potential for lasting ramifications on one's sense of agency because these relationships accompany the physical, mental, social, and emotional development of the child. A healthy trust relationship between parent and child thus often coincides with healthy social and psychological development from infancy into adulthood.

Given that parent-child relationships have a lasting impact on the development of a capacity to trust, we can hypothesize the following: remedying social and political practices that inhibit a parent's ability to properly care for their child should a top priority if our goal is to cultivate a healthy climate of trust. Children who experience a healthy trust relationship with their parent(s) will not only have developed the ability to trust others, but also themselves. Furthermore, they will have a frame of reference when it comes to judging when a trust relationship is healthy. Without this foundation, expanding, maintaining, and repairing broken webs of trust would be difficult, if not outright impossible.

Teachers play a role in the budding agency of their students, especially when it comes to developing their student's epistemic abilities and confidence in themselves as knowers. Like parents, teachers (especially those who work in early childhood education) also play a role in initiating their students into social norms and developing self-trust. However, we must also attend to the differences between teacher-student and parent-child relationships.

---

cited from this essay is of a parent who trusts (in the therapeutic sense) their teenage daughter to stay home alone for the weekend.

One important difference is the setting—parent-child relationships occur in the home whereas teacher-student relationships are established in the classroom, often the one of the first places where we are socialized outside of our families. Therefore, the way classrooms are organized is crucial to the development of one's social identity—both how one sees oneself in relation to the world and in one's ability to imagine how others see them. Lots of work has been done to show that the classroom, educational institutions, and the pedagogical practices of teachers can be structured in ways that promote trust.[72] What this research speaks to is the importance of looking at how structural features of an environment can promote or discourage trust among the individuals within it. Fostering trust and respect in the classroom thus also has the power to impact trust on social and political scales.

Education research has noted the importance of trust both in the classroom broadly and in dyadic teacher-student relationships. It appears that mutual trust between the teacher and student is conducive to optimal learning. Even as the age and knowledge gap between teacher and student converges, trust continues to play an important role in the collaborative efforts that lead to new ideas and knowledge. Hardwig argues that epistemologists have not realized the extent to which a climate or environment of trust supports knowledge and rejects the belief that "trusting and knowing are deeply antithetical" (Hardwig 1991, 693).

Part of the power asymmetry between a teacher and student is the extent of their knowledge (typically, the teacher has read more, practiced more, or has more experience than their student), but another substantial asymmetry exists because the teacher is in control of what

[72] For more on how institutional arrangements can make it possible for schools to be sites for deliberative communication see Thomas Englund (2011) in "The Potential of Education for Creating Mutual Trust: Schools as Sites for Deliberation". Eve L. Ewing's *Ghosts in the Schoolyard* (2018) offers an extensive case study of how education policy and school closings affect trust among Black Americans living in Chicago's south side. bell hooks (1994) describes how community can be built within the classroom through transformative pedagogy in *Teaching to Transgress,* see pages 40-41, 83-85, 177-179.

the student learns and largely influences how the student learns it. This is a desirable asymmetry of power when certain conditions are met: there are good reasons that teachers make syllabuses for students. This is not to say that teachers don't also learn from their students (it has become cliché at this point for teachers to say this), but rather to recognize that, given their role in facilitating the process of learning, there is more than just a deeper well of knowledge that contributes to the power asymmetry between teacher and student.

Doctors, like teachers, have expertise and knowledge that contribute to the asymmetry of power between them and their patients. Trust in doctor-patient relationships is vital because patients depend on their doctors for explanations of their ailments and potential methods for treating it. As a result, doctors have to take on the role of translator—they need to be able to convey what they know to their patients in a way that their patients can understand. Trust helps facilitate this process: if we don't trust the doctor's translation, we will be less confident about the information they provide us with. Similar to teacher-student relationships, this discrepancy in knowledge is not the only source of the power asymmetry in this relationship. In the current United States' (and many other) healthcare systems, doctors also control the access that patients have to medications and other forms of treatment. While there is some debate about whether restricting patient access to pharmaceuticals is good, the asymmetry here has some intuitive appeal—most patients want someone who understands the human body and knows how it will respond to different medications and treatments to be the one prescribing them.[73] Like any discretionary power, the ability to grant or restrict access to a medication or treatment option can

[73] For the skeptical view see Jessica Flanigan's *Pharmaceutical Freedom: Why Patients Have a Right to Self-Medicine* (2017). While individual opinions may differ. I do think that patients who *want* a doctor's recommendation will also desire a trusting relationship with their doctor.

be improperly used. When it is, room for distrust not only of the specific doctor, but of the medical system as a whole, opens.[74]

What the doctor-patient relationship brings out is how interpersonal instances of trust are often mediated by the role or position one occupies in society. It is helpful to recall Krishnamurthy's distinction between horizontal and vertical trust here if only to show how, in doctor-patient relationships, the line between the two can be fuzzy. To be clear, I think the distinction is a useful one, but in practice, these two forms of trust are often experienced simultaneously and can be difficult to pull apart. Consider the following: a patient seeing a new primary care physician for the first time might go into their appointment and distrust their new doctor because they distrust the medical system as a whole and view the doctor as an agent or extension of that system.[75] However, after a helpful first visit, or perhaps over the course of several appointments, the patient may come to trust the doctor. This seems particularly likely if the doctor has 1) suggested treatments that have benefited the patient (perhaps they prescribe an effective migraine medication) and 2) if the doctor has been attentive and responsive when interacting with the patient. Once trust between the doctor and patient is established, the patient might modify the degree of their distrust in the medical system as a whole. For instance, the patient might also be more inclined to trust a specialist that the (now-trusted) doctor refers them to. If they have a similar experience with that specialist and other healthcare professionals, we can imagine that the accumulation of these experiences might mitigate the patient's distrust of

[74] The lasting impacts of the inhumane Tuskegee Syphilis Study, for instance, has been linked to the relatively low levels of trust in the U.S. healthcare system among Black Americans. The effects are particularly pronounced when looking at their willingness to participate in medical research studies, which in turn exacerbates existing inequities in healthcare. For more on this point see Ralph Katz's team's work on the Tuskegee Legacy Project (2008, 2009). For possible routes to dismantling this distrust and remedying existing inequities see Jaisawal (2019).

[75] See Lanphier and Lomotey-Nakon (2023) for an account of why mistrust in the medical system in the context of birth is reasonable. Lanphier and Lomotey-Nakon recommend trauma-informed remedies that can be used in such cases.

the healthcare system or transform their relation to it from distrust to trust. The interpersonal and often intimate nature of providing medical care, combined with the degree to which the medical system is imbedded within other social and political institutions makes doctor-patient relationships interesting cases for examining the 'webs' and 'networks' of trust Baier thinks we ought to pay more attention to. Through them we can see how historical and cultural factors that impact trust coalesce with how trust functions in interpersonal interactions.

I have three observations about these kinds of relationships as a group. First, the types of relationships I've just discussed are ones in which there is dominant social paradigm, but lots of variation with respect to history and culture. In addition, exemplary forms of each of these relationships require not just trust, but forms of trust where both parties expect goodwill from one another. I believe it is likely that relationships with more subtle or fluctuating asymmetries of power (e.g. friendships and romantic relationships) also involve an expectation of goodwill. One of the tendencies of doxastic accounts of trust is that the role of goodwill gets pushed out to the periphery. In the next chapter, I will argue that this tendency has led to a problematic line of argument that favors "demoralizing" trust.

Second, the importance of trust in these relationships has understandably led many philosophers to be concerned with the cases where trust goes wrong. The dangers of misplaced trust in these and other relationships has increasingly directed philosophers working on trust to think about trustworthiness. They are guided by the assumption that the appropriateness of trust depends on the whether the one-trusted is believed to be trustworthy. As we already saw, D'Cruz and Holton suggest there are reasons to extend trust in the face of uncertainty about trustworthiness. McGeer and others who work on "therapeutic trust" also show that trustworthiness per say is not the only indicator of when trust is appropriate or morally good.

These strong exceptions to the idea that we should trust only the trustworthy indicate that an adequate account of trust must come before we can adequately conceptualize trustworthiness. I will say more about why I think the turn toward trustworthiness by affective and doxastic accounts is premature in the next chapter.

Finally, recall the example with which we began. The recovering addict navigates all of these relationships (as well as others) simultaneously throughout her period of recovery. The relationship between her and her parents fluctuates between trust and distrust. In the process of learning how to manage her addiction she may alternate between trusting and distrusting her teachers and any doctors or therapists she is seeing. Furthermore, each of these relationships will impact the way she sees herself and understands her identity. If the recovering addict continues to be distrusted by those around her, it is likely that being treated this way may cause her to doubt—or lose trust in—herself, which could prompt feelings of helplessness or even a relapse. What this example draws out is that these various forms of trust each contribute to the overall climate of trust she inhabits. The quality of that climate has an impact on her recovery and vice versa.

Conclusion:

What I have offered in this chapter are three requirements for an adequate account of trust. In the process of doing so, I explored the various dimensions of trust's value and suggested that the dimensions related to intimacy and identity can help explain why failures of trust can lead to deep wounds. I argued that the ability to account for distrust is a necessary feature of an account of trust because the two concepts inform each other, because an understanding of distrust illuminates the advantages and risks of trust, and because both trust and distrust have important practical applications. The impacts of distrust can extend beyond interpersonal

interactions and into community, as well as to larger social and political scales. Looking at distrust shows us why trust matters, but it also provides us with a new angle from which to view the connections between trust, identity, and vulnerability.

In section three, I suggested that trust makes us vulnerable in a way that is different in kind from the vulnerability we're exposed to when we merely rely and inferred that this difference might have something to do with the various ways in which trust contributes to a flourishing life. Trust's ability to expand our agency by facilitating cooperation and providing security is certainly one of these ways. However, trust also enables intimacy, thus making meaningful relationships possible. A fundamental form of trust also seems to underlie recognition. In combination, these other dimensions of value seem to confirm my hypothesis that trust plays a role in the formation of our identities, a point I will explore further in the remaining chapters.

In the final section, I described why an account of trust should be able to explain trust in relationships with asymmetries of power. I considered reasons offered in the literature, namely that the relational nature of trust and its role in facilitating the development and maintenance of our most important relationships makes it nearly impossible to avoid doing so. This requirement is particularly significant for those who are interested in an account of trust that is useful in applied contexts. A challenge for meeting this requirement is that the account of trust must be able to account for the complexity of trust (i.e. think of it as something that is cognitive, conative, affective, etc.) as well as the "looping effects" that reinforce our biases about who is and who is not to be trusted.

Chapter Three: Shortcomings & Pitfalls of Doxastic and Affective Accounts

"The whole enterprise of justifying trust in this way is a bit like trying to decide when its ***useful*** *to fall in love."*

–Olli Lagerspetz,

So far, I have revisited Annette Baier's work and identified three minimal requirements for an account of trust. These requirements are as follows: 1) it must be able to account for distrust; 2) it must make a meaningful distinction between trust and reliance; and 3) it must be able to attend to how trust operates in relationships with asymmetries of power. In this chapter, I demonstrate that two of the prominent kinds of accounts of trust—doxastic and affective—fall short of these requirements. The first two sections of this chapter analyze each kind of account with respect to each of the requirements. In the third section of this chapter, I show that, in addition to not meeting all three requirements, these accounts fall prey to two pitfalls—the push to demoralize trust and the turn toward trustworthiness, in part because they fail to attend to power asymmetries and external circumstances. I close by suggesting that a hybrid doxastic-affective account would meet the same fate, as ultimately neither doxastic nor affective accounts have the resources to help us fully appreciate the role trust plays in the attainment of ethical goods.[76]

[76] "Ethical goods" is a term I am borrowing from Jennifer Morton. In her book *Moving Up Without Losing Your Way* (2019) she argues that a clear-eyed narrative of upward mobility requires accounting for the costs of the ethical

Doxastic Accounts

The broad category of doxastic accounts of trust encompasses the following subtypes: commitment-based, obligation-ascription, and cognitivist accounts. What these accounts have in common is that they view trust as a phenomenon that primarily involves a belief the one-trusting has about the one-trusted. Explicitly or implicitly, each of these accounts also adopts the entrusting framework as their model for interpersonal trust. As a result of this assumption, when they offer rules or guidelines about how to determine when trust is warranted or appropriate, they tend to focus on norms regarding the object or domain entrusted.

The most common (and most recently emerged) subtype among the doxastic accounts of trust are commitment-based accounts. Hawley (2014) was the first to introduce a commitment-based account of trust on which "to trust someone to do something is to believe that she has a commitment to doing it and to rely upon her to meet that commitment" (Hawley 2014, 10). Hawley acknowledges that there are two senses of commitment: the psychological (i.e. having a determined intention) and acquired (i.e. the kind we accumulate from making explicit or implicit promises). She sees the latter as most relevant for thinking about trust and distrust. Amy Mullin (2015) suggests that a narrower kind of commitment—a commitment to a shared social norm—grounds our trusting beliefs (Mullin 2015, 320).[77] Unlike Hawley, Matthew Bennett opts to

goods lost as well as the financial costs. On her definition, ethical goods include things like community and a sense of identity, both of which she takes to contribute to a flourishing life.

[77] Mullin states early on in her essay that she shares Jones' view that there is "a significant affective component to trust" (Mullin 2015, 317-318). However, she introduces her social norm commitment-based account of trust as an alternative to Jones' affective account and does not further describe what this affective component is like.

prioritize psychological commitments.[78] On his account, trust amounts to a belief that the one-trusted has a psychological commitment that will motivate them to act in a way that meets my trust (Bennett 2021, 259).

Commitment-based accounts and obligation-ascription accounts of trust share many of the same features. For instance, on both types of accounts, the thing securing trust—i.e. the commitment or obligation—can be implicit or explicit. Both accounts also contend that it is the breaking of the trust-securing mechanism that instigates a reactive attitude response. The difference between them is that commitments are something we can choose to take on, whereas obligations are less voluntarily chosen. Nickel (2007), who popularized the obligation-ascription account, proposes that trust is secured by the one-trusting's belief that the one-trusted has an obligation that *ought* to be met. Though he conceives of trust as involving a moral attitude, the moral attitude is tied directly to the ascription of the obligation, which is what in turn justifies the betrayal response.

The cognitivist accounts proposed by Hardin (1993) and Hieronymi (2008) construe the beliefs involve in trusting as a matter of responding to reasons. On Hardin's encapsulated interest account, trust is a cognitive state where the one-trusting has "adequate reason to believe it will be in that person's interest to be trustworthy in the relevant ways at the relevant time" (Hardin 1993, 505). What grounds trust is believing that the interests of the one-trusted can be relied upon to motivate them to act as the one-trusting expects them to. Despite also tying trust to a reason-giving belief, Hieronymi disagrees with Hardin, arguing that "to whatever degree you need to supplement your confidence in the person with reasons that show the trusting response is

[78] Bennett carves up the territory of senses of commitment in a slightly different way. He distinguishes between "normative" commitments, which are like the acquired commitments Hawley discusses, and "psychological" commitments, which are the kind that are relevant for his account.

important, to that extent you fall short of (fully) trusting the person in question" (Hieronymi 2008, 215). In other words, on her account there is an inverse relationship between the number of reasons supporting the one-trusting's confidence that the one-trusted will act in the way the one-trusting expects and the degree of trust you have in that person. The belief that a person is trustworthy puts a stop to the search for further reasons to trust.[79]

As we can see, there is a lot of diversity among accounts of trust that fall into the doxastic category. However, their shared commitment to defining trust as a matter of our beliefs results in similar outcomes when evaluated through our three requirements.

Doxastic Accounts and Accounting for Distrust

Most doxastic accounts have little, if anything, to say about distrust. This is, in part, due to the trajectory of the trust literature, as accounting for distrust has only become standard since Hawley's 2014 essay. For example, while Nickel (2007) does not discuss distrust at all, Bennett (2021) acknowledges that the psychological-commitment account he offers in "Demoralizing Trust" is incomplete for this very reason. Mullin's discussion of distrust begins and ends with a brief engagement with Jones's observation that it is difficult to judge when trust and distrust are appropriate. Overall, doxastic accounts struggle to explain how trust and distrust are related.

Among those who do mention distrust few attempt to give a complete account of it. Hardin (rightly) suggests that all things considered, trust is preferable to distrust because distrust causes us to forego opportunities (Hardin 1993, 508). He observes that one's past experiences

[79] Hieronymi reaches the conclusion that there are degrees of trust. Full-fledged trust is secured *only* by the belief that the one-trusted is trustworthy within the relevant domain. However, if you begin adding other reasons to the trusting-belief, you start inching closer to reliance. I find it compelling that Hieronymi's account leaves room for forms of trust that fall somewhere between trust and reliance, though I disagree with her chosen terminology. Calling forms of trust that rely solely on a trusting belief "full-fledged" implies that other forms of trust are insincere or less genuine. I think what Hieronymi calls full-fledged trust resembles a basic form of trust—the two-place kind we see in intimate interpersonal relationships, which is closer to the paradigm of trust than the forms of trust typically offered by doxastic accounts.

may impact one's propensity to trust or distrust but does not provide an explanation of how the two concepts inform each other. Hieronymi uses the fact that one might feel resentful upon discovering that the reason they were trusted was something other than a belief about their trustworthiness to support her argument that full-fledged trust does not appeal to other reasons (Hieronymi 2008, 230-231). So, it seems like cognitivist accounts acknowledge that being distrusted is something that we typically take offense to and that being trusted (or at least being believed to be trustworthy) is something we generally desire. While each of these accounts provides a piece, even when taken together, they fail to complete the puzzle. Recall that accounting for distrust involves at least three components: an account of distrust must be able to explain how trust and distrust are related, understand distrust's epistemic and ethical risks, and attend to how distrust emerges in practice.

What distrust might look like on these doxastic accounts is an inversion of the belief that is taken to secure trust—when we distrust we do *not* believe that the person will act in ways that encapsulate our interests, we do *not* believe that they are trustworthy, we do *not* believe that they have a sufficiently motivating normative commitment or that they share a commitment to the relevant social norm, and so on.[80] The problem with construing distrust in this way is that there are often cases where we cannot pinpoint a specific belief about a person that supports our distrust. Distrust can be rooted in a feeling of fear or anxiety and may flow directly from how we perceive our situation without necessarily entailing a particular belief.[81] Furthermore, when the

[80] By inverting these accounts of trust to imagine what their account of distrust might be like, I am following Hawley, the one exception to my overall assessment that doxastic accounts do not meet the first requirement. As we saw in chapter two, when Hawley introduces her commitment-based account of trust, she asserts that an adequate account of trust should be able to account for distrust and proceeds to offer such an account. I do not consider her account at length in this chapter because, as I showed in chapter two, her account is unable to accommodate the financier case D'Cruz proposes.

[81] By feeling here, I am referring to a sensation one can locate in their body. I discuss what I mean by "feeling" briefly in the preface.

source of our distrust is what we might colloquially call a "gut feeling", attempting to articulate a belief to explain our distrust retrospectively would result in post-hoc confabulation.[82] Our tendency to do this in real-life (particularly when a person does something that *confirms* that our distrust was warranted) can mask the fact that our feelings were what was *actually* driving the distrust. In emphasizing the fact that distrust often stems from our feelings, I am not arguing that distrust is *never* informed by our beliefs. Often our beliefs do inform our judgements about whether someone can be trusted. For instance, if we believe that someone is a liar or a thief, distrusting them on the basis of that belief (assuming it is well-founded) is a rational thing to do.

Another reason that doxastic accounts struggle to account for distrust is because of the relationship between beliefs and evidence. In theory, if evidence that contradicts our belief is strong enough, it should lead to a revision of our belief. However, as D'Cruz (2020) and Jones (2019) have pointed out, both trust and distrust are recalcitrant to contrary evidence. This characteristic of distrust and trust is hard to explain on doxastic accounts and becomes much easier once we acknowledge that our trust and distrust are a direct response to our (affectively loaded) perception of our situation.[83] If we accept that we are sometimes inclined to distrust simply because of the way a person or situation makes us feel, then we must also admit that beliefs alone cannot explain how the concepts of trust and distrust are related.

Finally, doxastic accounts only gesture to the ethical and epistemic harms of distrust. As a result, they have a difficult time explaining how distrust emerges in applied contexts. While the content of a belief that grounds distrust may explain why being distrusted feels like an insult,

[82] See Katheen Murphy-Hollies's essay "Self-Regulation and Political Confabulation" (2022) for a helpful account of why we confabulate and what its consequences might be.
[83] I describe the way I am using perception in the preface.

there are many cases where distrust feels like an insult not because of what the one-distrusting believes to be true about us, but rather because of what they do *not* believe to be true about us.

Let's return to the recovering addict case to illustrate. When the recovering addict experiences the distrust of her friends and family years into her recovery, the issue is that her friends and family do not perceive her in a way that accords with the evidence that should inform their beliefs about her. Her loved ones may continue to see her as the person she was when she was abusing drugs, even when they believe (and have adequate evidence to support the belief) that she has been staying sober, has become more responsible and considerate, and that she cares about proving herself trustworthy to her friends and family. The point is a change in our beliefs about a person may not always be enough to sway whether we respond to them with trust or distrust. The recovering addict might describe this experience of being distrusted as a form of misrecognition, as a failure on the part of her friends and family to perceive how her sobriety has changed her as a person. Continuing to be seen in this way by people who have all the evidence they should need to know that she has in fact changed can lead to alienation and isolation. Doxastic accounts fail to acknowledge the ways that a mismatch between one's beliefs and one's distrust constitute a unique sort of harm. They also lack the resources to explain why, in the real world, the distrust of the recovering addict's loved ones might persist despite ample evidence of her trustworthiness.

Doxastic Accounts and Making a Meaningful Distinction Between Trust and Reliance

Doxastic accounts distinguish between trust and reliance by locating the difference between the two in the presence of a particular kind of belief about the one-trusted that provides evidence to support their trust. As a result, proponents of the doxastic view tend to agree that our ability to trust extends only to other human beings, or at most, groups comprised of and

institutions created by them.[84] However, given that the only difference is the *kind* of belief, there is reason to doubt that this way of making a distinction between trust and reliance is meaningful. In order to do so, we need to identify a difference between the values that characterize trust and reliance. Observing this tension, Keren (2020) argues that trust's claim to distinctive value falls prey to a version of the Meno problem. Like the difficulty of weighing the respective value of knowledge and true opinion, we must identify something intrinsically valuable about trust if we are to evade this challenge. Keren suggests that this problem is particularly difficult for doxastic accounts because they must explain how their mechanism for distinguishing between trust and reliance—the addition of a particular belief—is more valuable than *pretending* to have that belief (Keren 2020, 118).

Ross F. Patrizio (2023) takes up Keren's challenge, offers a more thorough analysis of the layers of trust's Meno problem (TMP), and defends doxastic accounts against Keren's claim that they lack the resources to resolve it. He breaks the challenge down into primary, secondary, tertiary, and swamping versions:

> **Primary version of TMP:** we need to explain why *actually* trusting someone is better than merely acting as if one does.
> **Secondary version of TMP:** we need to explain not only why trust is better than feigned trust, but also identify *what* makes this difference without reducing it to a value-conferring property of trust. (If it is only a value-conferring property of trust that makes the difference, what is intrinsically valuable is *not* trust itself, but rather the value-conferring property).
> **Tertiary version of TMP:** we must show that the difference between trust and reliance is a difference in *kind* rather than degree in order to show that trust itself has distinct and indispensable value.

Solving the tertiary problem resolves the primary and secondary versions of TMP. This is good news. The bad news is that we are still left with what Patrizio calls the swamping problem:

> **Swamping problem in TMP:** no additional value can be conferred upon trust by the difference-making component. (If the value of the goods we get from trusting

[84] Hawley (2017) keeps that range rather narrow. She suggests reliance is more appropriate for talking about the way we relate to institutions and the government.

> are the same as the value of goods we get from pretending to trust when we are really just relying, the value of the final goods obtained can "swamp out" any value that is added by the difference-making component).

Patrizio illustrates this version of the problem by making an analogy to two cups of coffee. If both cups possess all the same properties (quality of taste, amount, temperature, caffeine dosage, and aroma) and the only difference between them is that one was made using an expensive machine and the other was not, Patrizio argues, the value that is added to the cup of coffee made using the expensive machine gets "swamped" by the value of the final good. The worry is that the same is true of trusting relationships.[85]

Patrizio attempts to show that doxastic accounts have the resources to resolve both the tertiary (and therefore primary and secondary) and swamping versions of TMP by arguing that trust has non-instrumental final value. By this he means that trust is valuable for its own sake because of its relational properties rather than for its intrinsic properties (Patrizio 2023, 12-13). He finds promise for this idea in Hawley's observation that trusting another can amount to paying them a compliment which says, "something positive about [the one-trusted] qua individual" and does not necessarily entail any additional positive downstream effects (Patrizio 2023, 13-15).[86] Thus, if an account of trust can offer a plausible story about how a compliment is communicated by trusting, it will also be able to "differentiate trust from that which falls short of trust in terms of a difference in *kind* rather than merely degree" (Patrizio 2023, 15). The idea

[85]For the sake of faithfully reconstructing Patrizio's argument, I take this point at face-value. However, I think there is a disanalogy between these two examples. A cup of coffee does not carry its association with the less-fancy coffee machine with it. The same is not true of relationships—part of why we value persons is because of their continuity over time and their connections to past versions of themselves. See Wolf (1986) for a more developed version of an argument about the value of persons.

[86] Patrizio acknowledges that Hawley is not the only philosopher to make this observation. Both Onora O'Neill (2012) and Jason D'Cruz (2019,2020) make similar observations about the link between trust and respect and distrust and disrespect. Furthermore, though Patrizio is trying to show that doxastic accounts can resolve TMP, he does not claim that non-doxastic accounts are unable to. The point of his essay is to show that TMP does not give us a reason to abandon purely doxastic views as Keren (2020) supposed. Keren's own account of trust is not purely doxastic.

here is that an attitude that expresses a compliment is different from one that does not and that dialing up one's degree of reliance cannot do the same. So, any less-than-trusting attitude would fail to secure the full value of trust. According to Patrizio, this difference in *kind* resolves both the tertiary and swamping versions of TMP. So, any doxastic account that can explain how trusting another pays them a compliment, Patrizio argues, is indeed able to make a meaningful distinction between trust and reliance.

I agree with Patrizio that trusting another pays them a compliment that merely relying on them while pretending to trust does not. I also agree with the general idea that a meaningful distinction between trust and reliance can be explained by appealing to a different in the *kind* of value trust has. However, as Hieronymi observes, the "trusting belief" that pays the compliment is not the kind that can be supported by reasons beyond believing that the person is trustworthy Not only should we note that is this feature not unique to doxastic accounts, but we should also note that Patrizio himself describes it as an *attitude.*

D'Cruz's account of trust's compliment-paying feature offers a more plausible explanation. For D'Cruz, the compliment that accompanies being trusted is not found in what the one-trusting believes, but rather in how the one-trusting acts: "Trusting another is a way of *treating* them as persons and of thereby signaling respect" (D'Cruz 2015, 469, emphasis my own). D'Cruz's statement explains why being trusted *feels* different than merely being relied on. Since we are not privy to the beliefs of others (at least not without asking) but can infer what others' beliefs are from how they interact with us, it feels odd to say that it is the trusting belief, rather than the way we are treated, that expresses the compliment. Such an explanation also cannot explain what D'Cruz calls a trusting overture—an act that presumes the trustworthiness of another for the sake of forming a trusting relationship with them. Trusting overtures do not

require the one-trusting to have any beliefs about how trustworthy the one-trusted is and may even be extended *in spite* of evidence about their untrustworthiness. Locating the compliment-paying feature in an attitude that is expressed by how the one-trusting treats the one-trusted provides us with a more plausible explanation of the relationship between trust and the formation of intimate, interpersonal relationships. If we assume that the trusting belief is prior to the attitude, it is unclear how such relationships can get off the ground. We can perceive another's attitude toward us and infer from that attitude what their beliefs about us might be, but not the other way around. On a belief-first account, it appears that any evidence of untrustworthiness would automatically disqualify someone as a candidate for a trusting relationship. We need the affective component to overcome this challenge. A serious consequence of accepting a purely doxastic view is that it diminishes the possibility of repairing or restoring trust after a betrayal.

The ability to make a meaningful distinction between trust and reliance requires introducing a role for our attitudes, which would bring any doxastic account into the territory of the affective accounts I consider later in this chapter. So, what Patrizio has actually demonstrated is that *hybrid* doxastic-affective accounts have the resources to solve TMP (and thereby make a meaningful distinction between trust and reliance).

Doxastic Accounts and Attending to Trust in Relationships with Asymmetries of Power

Doxastic Accounts struggle to attend to how trust operates in relationships where there are significant asymmetries of power for three reasons. First, the beliefs these accounts assume are involved in trusting are often over-intellectualized or too cognitively demanding to attribute to all entities with the capacity to trust and be trusted. Second, the forms of trust seen in many relationships with an asymmetry of power are often latent. The beliefs that doxastic accounts take to ground trust often only become the object of our conscious awareness after the trust has

been broken. Finally, doxastic accounts do not consider a significant variable—the external circumstances that shape the power asymmetry in each trust relationship. I will consider each of these points in turn.

Doxastic accounts preclude the possibility of trust forming in many cases where it would seem fitting because they define trust in terms of a specific and often complex belief. Infants and young children are the first group that comes to mind. Children have beliefs, but they can only have beliefs about what others believe when they reach a certain level of social maturity. Even when we start to have intuitions on this matter, our ability to articulate these thoughts as a coherent belief to ourselves and others is limited during our early stages of life. Children are not the only group that pose a challenge for doxastic accounts. Those who are in cognitive decline or dying can also exhibit trust without having coherent beliefs about others' commitments, interests, or obligations. A similar challenge arises if we want an account that allows for trusting relationships between humans and non-human animals. In other words, if trust is to be possible in all or most of our relationships, it cannot be reduced to a matter of beliefs.

Aside from these three groups that constitute a challenge to the view that trust is primarily a matter of beliefs, the latent forms of trust seen in our thick trusting relationships are also not captured by the doxastic accounts' emphasis on beliefs. Take, for example, two adults who became friends during their childhood. The mutual trust between them builds up over time and is not supported by some belief they both share about what commitments, obligations, or interests will lead them to act in a trustworthy way, but rather by their shared history and understanding of who the other is. In enduring relationships such as childhood friendships, trust is typically mutual and remains unspoken. It also may be difficult to pinpoint the exact moment when it "started". In other words, doxastic accounts do not properly calibrate how our past

experiences with others factor into our future expectations about how they will treat us. At best, they provide a formula that allows us to retroactively describe why we had thought we could trust them after our trust is betrayed. When we reflect on why we trusted after trust fails, we no longer have the feeling of trusting, and so will appeal to other evidence. We may point to a commitment, obligation, or interest that we (mistakenly) believed the other person had to explain why we had trusted someone when prompted. However, these thoughts are often not present in our minds when we are trusting. Though we may appeal to them as the source that prompted our trust, they are often confabulated reasons that we use to make sense of the error in our judgement.

Accurately assessing a person's character takes time, in part because in order to know which of a person's traits are stable we must see that person across a variety of circumstances. Our reasons for trusting another person are often rooted in the fact that we have past experiences with them where they have *not* proved untrustworthy. When we invoke the two-person trust relation what we are essentially offering is testimony that we trust that other person has a trustworthy character. This can be narrowed to a specific domain (i.e. their competency with respect to a specific task), but it can also be a much broader statement that speaks to our trust of another's judgement. The qualitative experience of knowing who another person is thus what often grounds the trust in our intimate, interpersonal relationships. [87]

Doxastic accounts thus give us an inaccurate descriptive picture of how we arrive at trust. However, one might argue that they provide a better description of when we should and should not trust. Rather than relying "vibes" (i.e. our unreflective feelings about a person) we *should*

[87] While a doxastic account may be modified to incorporate value-laden beliefs concerning other people and our relationships to or history with them, the qualitative experience I am referring to here involves the parts of our memory that can't be reduced to beliefs. It strikes me as extremely difficult to try to articulate the entire history of a relationship as a belief.

consider what beliefs would justify trusting that person and what evidence can justify such a belief. Such a practice, the defender of doxastic accounts might argue, will lead to better outcomes because our trust would be rooted in something other than our (likely flawed) impressions of what others are like. However, such an explanation leaves out the fact that other people matter to us, at least in part, because of the history of our relationship with them.[88]

Accounts of trust that portray trust as mostly a matter of beliefs are also insufficient for explaining the kinds of trust seen in the three characteristically asymmetrical relationships I considered at the end of chapter two: doctor-patient, teacher-student, and parent-child relationships. In such relationships, it is desirable for the trust to be rooted in the history of the relationship rather than in some belief about what abstract commitments motivate either party. We want our doctors, teachers, and parents to trust us *and* be trustworthy for us because of who we are and not because of what they believe to be true about our interests, commitments, or obligations. This is because trust signals an investment in the relationship that is different in kind from the reliance that characterizes more purely transactional relationships. Healthcare and education (and higher education in particular) have increasingly trended toward a consumerist model and such shifts have been criticized on the grounds that they alter the relationships between doctor-patient and teacher-student in undesirable ways. In healthcare, for example, the "informative" model of the doctor-patient relationship frames the patient's role in maintaining their health and well-being as a kind of shopping for services. Such a framework makes it difficult to see patient health or well-being as a shared project. The same goes for education: when the relationship between teacher and student is reduced to a mere transaction, seeing

[88] See Susan Wolf's "Self-Interest and Interest in Selves) (1986) and "Happiness and Meaning: Two Aspects of the Good Life" (1997).

education as a collaborative production of knowledge becomes difficult because the transmission of knowledge from teacher to student is the only thing that is treated as valuable.

Being trusted because of *who* one is by their doctor, teacher, or parent, is integral to the development of self-trust in the patient, student, or child. This formation of such self-trust is in turn beneficial for the shared projects that are the proper objects or outcomes of such relationships. Tying the formation of self-trust to a belief or set of beliefs that the patient has about the doctor (and vice versa) rather than to the shared history of the relationship cheapens the significance of such relationships, ignores the experience of care, and the labor involved in providing it. It suggests that the particular doctor, teacher, or parent we trust could be exchanged for any other doctor, teacher, or parent, so long as the patient, student, or child believes that they have interests, commitments, or obligations that are sufficiently similar. This implication is troubling because it undermines the idea that the persons in the relationship are intrinsically valuable.

Finally, our judgements about who to trust, when and with what are impacted by factors that have nothing to do with our beliefs about the one-trusted. Our circumstances can also justify or warrant trust even when our beliefs about the one-trusted provide evidence to the contrary. Recall Baier's story about the unmarried, junior professor from chapter one.[89] In this case, the woman's circumstances played a significant role in her judgement about trusting her new boss and colleagues. It would be naïve, especially given the historical-social context provided in the example, to suppose that the junior professor trusted the men in her department to treat her as an equal because she believed they had a commitment to dismantling the sexist norms in the discipline. In actuality, she probably believed it was likely that her male colleagues held sexist

[89] This appears on page 19.

views and that, as an unmarried woman in a discipline dominated by men, she anticipated prior to the starting the job that she would encounter challenges to her credibility and unwanted sexual attention. Despite her awareness of these possibilities, she still extended trust to them because her circumstances were such that *not* trusting would have come with significant costs—forgoing the opportunity of making new friends, of belonging to the intellectual community in the department, and missing opportunities that might advance her career. More importantly she trusted the men in her department (and the chair in particular) and the university, despite having beliefs that, on most doxastic accounts, would be indicative of distrust. Attending to how trust functions in relationships with asymmetries of power requires reckoning with circumstances, which include both the social and historical context, as well as the past experiences of the particular individuals involved.

To summarize: doxastic accounts fail to adequately account for distrust; their ability to make a meaningful distinction between trust and reliance requires the introduction of an affective component; and they face several challenges when it comes to attending to how trust operates in relationships with asymmetries of power—challenges that cannot be overcome by appealing to beliefs that the one-trusting holds about the one-trusted. Doxastic accounts fail to attend to the way trust operates in relationships with asymmetries of power because they cannot explain why we trust in spite of evidence-based beliefs that count against trusting.

Affective Accounts

Unlike doxastic accounts, affective accounts do not sort themselves into further subcategories. The shared starting point of affective accounts is the observation that trusting, unlike merely relying, makes us susceptible to betrayal. This difference in our emotional response is typically used to justify a restriction on the entities we can trust (i.e. other moral

agents) and is appealed to by these accounts as reasons why trust cannot be reduced to belief. Instead of focusing on beliefs, these accounts see the emotional or attitudinal component of trust as most basic. However, I find them to be divided by their stances on whether trust can be voluntarily willed.

Holton (1994) is typically identified as the first to offer an affective account. He defines trust as adopting the participant stance toward someone when we rely on them. He argues that the desire to improve or establish a relationship with someone can motivate us to *decide* to trust and takes this to demonstrate that trust can be voluntary. Jones (1996) initially argues that trust involves optimism about the goodwill and competency of the one-trusted. Though her account does not permit voluntary trust in the way Holton's does, she suggests that our tendencies to trust and distrust trust can be gradually shaped by individual's intentional efforts to direct their attention. In her revisions to her earlier account, she drops the requirement of expecting goodwill and replaces it with the idea of "trust-responsiveness". On the trust-responsiveness view, in trusting you to φ, I also trust that you will take the fact that I am trusting you as a reason to meet my trust. McLeod (2002) offers a different revision—trust is an optimistic expectation, not about the one-trusted's goodwill, but rather about their moral integrity. Kirton (2020) suggests that trust is best understood through the concept of attachment security. On his account, what is of concern to us when we trust is whether we feel like we matter to the person or group we are trusting.

Though each of these accounts leaves room for trust to be influenced by our beliefs, they do not think that trust can be reduced to them.[90] Another view they share is that the emotion or attitude involved in trusting is what makes it distinct from reliance. As a result of prioritizing the

[90] Someone who thinks that our emotions are at least partially constituted by beliefs will not find this move convincing. See Martha Nussbaum's 2012 book *Upheavals of Thought* for a compelling version of such an account.

emotional component, affective accounts fare better when it comes to accounting for distrust and making a meaningful distinction between trust and reliance. Affective accounts are also more amenable to explaining how trust works in relationships with asymmetries of power, but I will argue that they still fall short because of their inattention to external circumstances and how they affect the voluntariness of trust.

Affective Accounts and Accounting for Distrust

Most affective accounts do a decent job of accounting for distrust. Typically, these accounts understand distrust to be an inversion of the affect or emotion associated with trust. For example, if trust involves the one-trusting's optimism about the goodwill and competence of the one-trusted, the one-distrusting is described as having pessimistic expectations—they doubt either that the one-distrusted bears them goodwill or that the one-distrusted has the relative competence (or both). If being trusted is described as feeling affirmed in our status as a moral agent or member of the moral community, being distrusted undermines this feeling. The inversion of each account captures both what distrusting and being distrusted feel like and, thus have the resources for explaining how the concepts of trust and distrust inform each other.

Accounts like Jones's and McLeod's also show us how distrust and trust can emerge from multiple avenues. Because they do not require the formation of a specific belief, affective accounts offer a more intuitive explanation of why trust and distrust are often immediate reactions. Furthermore, they also can account for the self-reinforcing and self-fulfilling tendencies of trust and distrust. For instance, in her 2019 essay, Jones writes that "non-affective or 'cold' beliefs do not evidence-tamper in the way that trust and distrust do" (Jones 2019, 960). The idea here is that the affective charge of trust and distrust cannot be disregarded or reduced to beliefs. Furthermore, Jones argues that when coupled with an allied emotional state, the prior

emotional state of distrust can be reinforced. This creates a looping effect that explains the self-confirming nature of distrust. Jones offers the example of Donald Trump's manufacturing of fear and contempt for undocumented migrants to illustrate this point. As Jones puts it, fear and contempt, are trust-phobic and distrust-phillic. When one is inundated with a crime schema about a group, like undocumented migrants, the fear and contempt that accompany that schema drive out trust, leaving it to be filled by distrust. Distrust, because it can be fueled by fear and contempt, then continues to grow as the criminal schema or narrative is spread. The same is true of trust—when our trust is met with openness it is reinforced and grows. To illustrate we can imagine a case of someone who nervously comes out as transgender for the first time to a close friend and is met with acceptance. We can imagine that this positive experience would reinforce the interpersonal trust between the two and that it might also make the idea of coming out to other friends and family less intimidating.

Affective accounts also appreciate the epistemic and ethical risks of distrust without losing sight of its occasional advantages. The ethical and epistemic risks of distrust can be hard to separate because they are often incurred in pairs by both parties. For instance, the one-distrusting incurs the epistemic risk of forgoing knowledge or cooperation from the one-distrusted as well as the ethical cost of cutting off the potential for a trusting relationship. At the same time, the one-distrusted experiences this distrust as an insult, thus incurring ethical harm. Depending on the case, the one-distrusted may feel gaslit or come to doubt their abilities, thus experiencing the epistemic harm of being made to second-guess themselves without good reason. The fact that these two sorts of risks are closely linked speaks to the distinctive value of trust. A key advantage of affective accounts is that they can explain why being distrusted feels like an insult. For instance, on Kirton's (2020) account distrust is described as a feeling of tension that

stems from an underlying concern about whether one belongs to a group or not. When one feels as though they do not belong, they are less likely to feel as though their contributions matter and therefore may refrain from making contributions entirely. Epistemically speaking, this is a risk because we miss out on new ideas (or counterpoints to existing ideas). Ethically is a risk because it may further entrench existing inequities or exclusionary practices.

Let's consider how well the feeling of tension described by Kirton captures the way distrust operates in practice. Imagine the social structure of a typical research lab—a principal investigator (PI), his research staff, graduate students, research assistants, etc. are all working on a project that aims to identify which protein should be targeted in the cancer drug they are developing. Now, suppose the PI, a Harvard man himself, distrusts the reasoning and technical lab skills of any scientist in his lab that did not receive their undergraduate degree from an Ivy-League institution. He does not say this out loud (nor does he necessarily need to be conscious of the fact that he has this bias), but the other scientists in the lab notice that he interrogates the reasoning of non-Ivy-League graduate students' suggestions more harshly, that he is more critical of their proposals for new experiments, etc. As the scientists without an Ivy-League pedigree in the scenario feel increasingly distrusted by the PI, they stop contributing their ideas at lab meetings. As a result, the lab loses out on an insights that might have challenged or transformed the way they thought about the experiment. Eventually, some of the distrusted members of the lab leave. Disillusioned by their experience, they may leave behind their interest in scientific research entirely and seek employment in a new field. Affective accounts allow us to see why such a decision is reasonable.

As the reputation of this lab starts to spread, fewer and fewer candidates from non-Ivy-League backgrounds apply for the open positions and eventually the lab is comprised entirely of

Ivy-League graduates. If we substitute "non-Ivy-League graduates" with women, disabled, or any marginalized racial or ethnic group, it becomes even clearer that the distrust has ethical as well as epistemic risks. In this example, the self-confirming nature of trust and distrust helps explain why, from the perspective of the PI, his distrust of the non-Ivy-League graduates might appear rational. It also helps explain how the initial tension Kirton describes can grow over time. In sum, affective accounts can account for distrust because they acknowledge the emotional roots that lead us to distrust others. They also appreciate the ethical significance of the emotional response that follows from being distrusted.

Affective Accounts and Making a Meaningful Distinction Between Trust and Reliance

One unifying feature of affective accounts is that they distinguish between trust and reliance by appealing to the differences in the emotional responses that follow from their failures. Though betrayal is often considered the standard emotional response, proponents of affective accounts agree that there is no reason to think this emotion is the only appropriate response. For example, we might also feel angry, sad, or even dumbfounded. On its own, this approach to distinguishing between trust and reliance does tell us more about the kinds of entities that we are able to trust (i.e. entities for which reactive attitudes are appropriate). However, it does not fully explain how trust and reliance are meaningfully distinct because, as Hawley (2014) notes, the presence of reactive attitudes are not unique to trust.

As we saw in the last section, the ability to solve TMP can be used as a proxy for determining whether an account of trust makes a meaningful distinction between trust and reliance, as it invites us to consider the difference in their respective values. To recap: if *pretending* to trust and *feigning* the emotional response of betrayal when our trust is not met yields the same benefits as *actually* trusting and *really* feeling betrayed, trust is no more valuable

than reliance. After describing the entrenchment of the TMP, Patrizio proposes that we can find a solution in the idea that trusting someone amounts to paying a compliment. In response to Patrizio's analysis, I suggested that this compliment-paying feature of trust cannot be described without appealing to an affective component. In this section I will demonstrate why an affective component is necessary for understanding how trusting someone amounts to paying them a compliment.

Compliments come in many different forms. They (typically)[91] involve a positive evaluative remark about an object a person owns (e.g. I like your hat), an aspect of person's appearance (e.g. he has beautiful hair), their personality (e.g. you are so tactful) or an assessment of their actions (e.g. that was a kind gesture). Each of these examples, though, could also simply be a descriptive report. What makes them compliments is not merely their content, so we must look to the context and purpose of such statements as well as the intention behind them. For such statements to count as compliments they must be stated with the intention of conveying a positive emotional response and within the context of some larger relationship. In other words, to pay you a compliment, I must also 1) intend to bear you goodwill 2) implicitly acknowledge that you are the sort of entity with which I can have a meaningful relationship. Complimenting someone is thus not a matter of what beliefs we have about the other person, but rather about the way in which we interact with them. This is where an affective component is necessary. For a positive evaluative remark to count as a compliment and not simply a descriptive statement that uses value-laden words, I must intend to convey a positive emotion or attitude toward you.

[91] I am setting aside "back-handed" compliments because I take them to be a derivative form of compliment which is not actually a compliment because it involves ill will. A person may pay someone a back-handed compliment without the recipient realizing or experiencing it as a gesture of ill will. Whether the recipient of the back-handed compliment is intended or expected to realize that the "compliment" is back-handed varies.

The compliment that trusting pays another most closely resembles the type of compliment that we may give someone about their personality, however what is unique about it is that it can be conveyed through actions and does not necessarily require verbal expression. The act of trusting another person thus conveys, at least implicitly, that I take you to be trustworthy. Trusting someone amounts to paying them a compliment because it implies that I respect your judgement, which requires me to acknowledge your freedom to choose how to best meet my trust (and implicitly the possibility that you may choose not to).[92] In other words, by trusting someone you acknowledge their personhood and status as part of a moral community. In this way, trusting someone is, at a basic level, a way of showing respect.[93] Most of us are inclined to hold other beings like us responsible for their actions. Respecting another's personhood is, for most of us, the default, so it is easy to forget that respecting another is bound up with trusting them.

Part of the difference between trust and reliance, according to affective accounts, is their relationship to evidence. The idea that it is the *attitude* or affectively loaded way of interacting with another that pays the compliment to the one-trusted has led proponents of affective accounts of trust to argue that part of what makes trust different from reliance is trust's relationship to evidence. This observation leads Jones to argue that trusting is more similar to *hoping* than it is to *predicting* insofar as both trust and hope do not always require evidential justification (Jones 1996,15). Though I agree with Jones's argument, we should also add that trusting is more like predicting than hoping when it comes to the degree of confidence or certainty that accompanies our judgement. I take it that predicting is what we do when we rely. The observation that trust

[92] For this reason, relationships that are built on forms of oppression and domination cannot be described as relationships of trust, at least not on my account. The structure precludes the possibility of embracing vulnerability. What appears to be trust in such relationships is really more like reliance.

[93] In a forthcoming article, "Cooperation and Climates of Trust", I explain the relationship between trust, recognition, and respect. I argue that, at their most basic levels, trust and respect are two sides of the same coin by employing Darwall's notion of recognition respect and Brennan's concept of recognition trust.

shares some qualities with both hope and reliance is key to understanding *why trust involves risk* and *how trust often goes unnoticed*.

All three of these actions—hoping, trusting, predicting (or relying)—involve orientations to the future. Making a prediction involves identifying confounding variables and anticipating how they will affect the outcome of a given situation. At the very least, predicting involves an implicit acknowledgement that things could go otherwise coupled with a judgement that, despite this possibility, our prediction is the most likely of the possible outcomes. When we hope, we are also aware of other possibilities and of the fact that these other possibilities are often more likely than the outcome we are hoping for. In this way, the expectations in both hoping and predicting are inherently comparative. The nature of the expectations involved in trusting are meaningfully distinct from the kind involved in predicting and from the kind involved in hoping. When we trust, our expectation is that that the entity we are trusting will meet our trust. In some cases, what is required for meeting our trust will be fairly specific, making it appear more like predicting or relying (e.g. I trust my partner to take the trash cans out to the curb on Monday nights) because there is a more limited range of actions that could meet my trust. In other cases, the range of actions that could meet one's trust is broader. In simply saying I trust my partner, for instance, the actions that could meet such trust are numerous. The point is, while we may not have specific expectations about what the one-trusted will do to meet our trust, we do *not* consider the possibility that things could go otherwise (i.e. that our trust will not be met). This is partly because the judgement we make when we trust is about *who* the person we trust is. When we trust someone our affectively loaded way of seeing the person restricts some possibilities from factoring into our judgement.[94] We take the person we are trusting as fixed variable that—

[94] How one understands the relationship between our emotions and our perception is a question worthy of the attention it has received. I'm still unsure of what my own definition of an emotion is, but fortunately for the

barring exceptional circumstances—secures the outcome we trust will happen. However, the fact that other possibilities do not enter our minds (because of the way you perceive the person you are trusting) does not mean that the possibility does not exist. This is why trust necessarily entails risk. Even though we do not consider it, other people always in principle have the freedom to act otherwise. This is also why trusting others often goes unnoticed—it causes a perceptual constraint, not unlike a "blind spot".

When we rely on, rather than trust, someone, our expectations about how they will act more closely resemble the kind of comparative judgements seen in hoping, Facts about others' qualities, habits, and/or character traits are factored into our prediction the same way other external variables are. When we trust who we judge the other to be, more than any other variable, shapes our expectations. This also lends additional support to explaining why being distrusted can wound us and make us feel disrespected—when we are unwarrantedly distrusted; we feel as though the person distrusting us is failing to see us for who we really are. In sum, affective accounts make a meaningful distinction between trust and reliance because they acknowledge the moral significance of both our relationships and our emotions.

Affective Accounts and Attending to Trust in Relationships with Asymmetries of Power

Because affective accounts prioritize the emotional component of trust over the cognitive component, they have less difficulty explaining how those with limited rational abilities can trust and be trusted. For the same reason they also have the conceptual resources to explain the more latent forms of trust that we are often only able to see in retrospect, after trust has been betrayed. Presupposing less about the rational abilities required and the (correct) presumption that our

purposes of this argument, I only need the reader to seriously entertain the possibility that our emotions have some effect on how we perceive a given situation and the people within it. Consider, for instance, how being angry with a colleague might affect your perception of their behavior during a meeting.

relationships with others cannot be reduced to some nexus of beliefs about commitments and interests gives affective accounts an advantage over doxastic accounts. However, although affective accounts can attend to a wider range of trust in relationships with asymmetries of power, they still fall short of meeting this requirement because they run into a dilemma. Affective accounts of trust must choose between having the ability to accommodate voluntary forms of trust and being able to accommodate how external circumstances affect the power relations between the two parties in the trusting relationship.

Holton (1993) is an example of an affective account that emphasizes the voluntariness of some forms of trust. On his account, deciding to trust seems to only require that one adopt a participant stance toward the person they are trusting—that is to choose to hold them accountable for their actions. He initially uses the example of a trust fall to demonstrate that trusting can, in fact, be a voluntary decision (Holton 1994, 64).[95] Holton goes on to argue that the voluntariness of such forms of trust demonstrates that trust does not require evidential justification. The fact that we appear to be able to *choose* to trust, in the same way that we could choose to adopt an objective stance toward those we interact with, implies that we can trust even in the face of evidence to the contrary. To illustrate he offers another example—a store owner and his newly hired, ex-convict employee. In the example, the employer is faced with the decision of either

[95] I concur with Jones and others that while this example picks up on an interesting quality of trust, it makes a poor paradigm. I would argue that the trust fall exercise, especially in the acting class context that Holton specifies, is much closer to reliance. It involves the kind of expectations I attributed to predicting in the previous section—one could have a generally low level of trust that strangers will bear them goodwill while still believing that their classmates will have other reasons to participate in the exercise: respect for the instructors' authority, social pressure to conform to the group, etc. In other words, there are a variety of reason for "trusting" that you'll be caught in the trust fall exercise, but only some of them involve a desire for a trusting relationship. Holton implies in the example that this desire is the only explanation—the student in the class *wants* to develop a trusting relationship with their classmates—while allowing that this desire may stem from other goals .(Perhaps the person believes having a trusting relationship with their classmates will make for a more authentic performance or perhaps they are simply taking the class because they want to make new friends). It seems likely that a mix of these reasons would go into the decision to trust in this example. His rock-climbing partner example does a better job of making this point clear.

installing a security camera to monitor his new employee or foregoing surveillance. Holton argues that one reason the employer might choose to do the latter is to establish a trusting relationship with the new employee. We can imagine the employer is motivated by further reasons. Perhaps, the employer believes that an employee who feels trusted is more likely to act trustworthy. Perhaps, the employer intentionally hires ex-convicts because he feels as though it is his moral duty to help mitigate the harms that flow from the cyclical relationship between poverty and crime.

Though the case does seem to provide an example of a voluntary extension of trust, it does not fully attend to the asymmetry of power in the relationship. Regardless of the employer's reasons for foregoing surveillance (and the degree of moral praise those motivations may deserve), the employer still has a substantial degree of power over the employee. If the employee were caught stealing, or even acting suspiciously, the employer could fire them, leaving them without a source of income and in danger of experiencing food or housing insecurity. The degree of power the employer has is amplified with the added detail that the employee is an ex-convict: public opinion and the law (imagine, if you will, that being employed is a condition of parole) are also powerful forces on the employer's side. Though the employer may not be consciously relying on his ability to fire or report his employee, the power dynamic still shapes the structure of the relationship.[96] It may look like the employer is taking on a substantial risk when, in reality, the risk is minimized to the kind taken when we rely by the norms that support the hierarchical relationship between employer and employee.

[96] It is also worth noting that the employee is likely more vividly aware of the power asymmetry and is thus more likely to allow it to guide their actions. The employer may feel as though he is trusting the employe, but the employee may interpret the relationship as one based on reliance, not trust.

Affective accounts, like Jones's, take the inattention to structural power dynamics seen in Holton's account as a point in favor of their view that is *not* voluntary. Jones's account leaves room for consideration of the ways our social and historical context affects our tendencies to trust, but vastly reduces the extent to which we can voluntarily alter such tendencies. Instead, she suggests that we can gradually and intentionally shape our tendencies to trust by making use of our reflective abilities: "If trust and distrust are partly constituted by our patterns of attention lines of inquiry, and tendencies of interpretation, it should be possible to cultivate them by controlling our patterns of attention, our lines of inquiry, and our interpretations" (Jones 1996, 22). This description of the kind and degree of control we have over our judgements about trust and distrust initially strikes me as correct for a few reasons. First, it offers a plausible explanation as to why, much of the time, we trust in ways that are "unconscious". The patterns, lines, and tendencies she refers to exert influence over our perception. Second, they appropriately capture the variables that factor into our judgements regarding trust or distrust. Our patterns of attention, lines of inquiry, and tendencies of interpretation are not themselves beliefs—they are rather necessary conditions for the possibility of belief. Third, the imaginative "inner" work that is involved in the controlling of these habits, and the implication that changing these habits takes time and concentrated effort are resourceful: it encourages us to consider what can be done in practice to mitigate the harms that our unconscious bias about who is and who isn't trustworthy can cause.[97] However, Jones's account does not consider that our external circumstances must be stable and secure in order for us to engage in this reflective work. In dangerous conditions characterized by extreme scarcity our ability to reconfigure these patterns is limited. For

[97] Notable ideas it has prompted include Jason D'Cruz's notion of "humble trust", which I discussed in the previous chapter.

someone in survival mode, the reflective inner work she recommends is a luxury. Indulging in it can have serious consequences.

To solve this dilemma, affective accounts must make a trade-off, and the trade-off is one that I don't want to make. In an account like Holton's, where trust is considered voluntary (and thus emphasizes the individual degree of control we have), we often lose sight of the social and historical structures and practices that shape power asymmetries, often in subtle (and therefore easily overlooked) ways. On the other hand, accounts like Jones's, which deny that trust is voluntary or suggest that voluntariness may only apply in peripheral cases, are sensitive to the subtle ways in which our lived experience is shaped by social and historical structures but tend to minimize the impact that an individual's voluntary action can have. Giving up the possibility of individuals' actions having such an impact becomes more troubling when we start to think about its implications for addressing the practical problem of distrust. The worry is that Jones's minimization of an individual's ability voluntarily extend trust is a bit pessimistic, and that Holton's picture is too insensitive to the very real challenges that our external environment can impose on our ability to choose to trust those around us. Having to give up either voluntary forms of trust or the ability to see how social and historical factors shape power asymmetry is too high of a cost. For this reason, affective accounts do not meet the third requirement.

Making this tradeoff is undesirable for several reasons. First, I believe that there are cases where trust is involuntary and cases where trust is voluntary. Second, while it is important to resist downplaying the extent to which our judgements about trust and distrust are shaped by social and historical features of our environment, doing so should also not deny the power that voluntary acts of trust can have on altering a given climate of trust. When we deny that trust can ever be voluntary, we preclude the possibility of solidarity. I follow D'Cruz in maintaining that

trusting overtures—voluntary extensions of trust that exemplify bravery and aim to inspire trustworthiness in the one-trusted—can have powerful effects on an individual's outlook on life, the relationship, and even on the surrounding community. However, even someone who denies that any individual action can have resounding effects must admit that, if acted upon in concert, trusting overtures could have structure-altering implications.

It is clear that affective accounts fare better than doxastic accounts when it comes to meeting the third requirement: attending to trust in relationships with asymmetries of power. Trust despite asymmetries in rational abilities can be easily explained. However, depending on their commitments with respect to voluntariness about trust they are forced to either give up the voluntary forms of trust or the ability to explain how our social and historical context affects our tendencies to trust and distrust. Accounts that do the former run the risk of neglecting the power we have to alter our material conditions. Accounts that do the latter often end up missing key details about asymmetries of power that operate in any given trust relationship. Giving up either end is troubling for an account of trust that aims to provide guidance on how to address the practical problem of distrust. Furthermore, the assumption that trust must be either voluntary or involuntary is itself suspect.

Trusting relationships are built over time through interactions where needs are expressed and met. In a genuine instance of trust within a relationship there will always be a power asymmetry present, though it may be subtle or stark. Attention to *this* detail is necessary if we are to understand how our interpersonal trust relationships affect local and global climates of trust.

Pitfalls of Doxastic and Affective Accounts

In this section, I look at two paths in the literature that have emerged out of debates between affective and doxastic accounts and offers reasons for characterizing them as pitfalls. I believe doxastic and affective accounts that pursue either route take us further away from an adequate account of trust. Though not *all* of the accounts I've discussed make these moves both affective and doxastic accounts have been used by those who do. In this section, I draw attention to these deeper problems with the intention of demonstrating that a conjunctive or hybrid doxastic-affective account cannot save the necessary and sufficient condition approach that has largely dominated the literature.

The Push to Demoralize Trust

The first pitfall is the push to "demoralize" or "avoid over-moralizing" trust. The general argument in favor of this push stems from the observations that trust is not always good; that we trust others to do morally questionable (and even morally reprehensible) things; or that trust is only good insofar as it is placed in the trustworthy or supported by evidence. The observation is not new, but the renewed attention it has received has led some philosophers to adopt the troubling position that we ought to strip trust of its moral content.[98] It is troubling because it yields an account of trust that is overly concerned with the epistemology of trust.

What I want to disambiguate is the claim that trust is *always* good from the claim that trust is conducive to human flourishing.[99] Though I readily accept the fact that there are cases where trust facilitates morally problematic practices and relationships, I do not think it follows

[98] Baier acknowledges that the presence of trust can enable exploitation and conspiracy (i.e. things we ought to discourage, rather than encourage) just as easily as it enables fellowship and justice in the second paragraph of "Trust and Antitrust".

[99] Another way of framing this that I considered, but felt would involve too many tangents, is that the problem here is about a confusion of definitions, or difference in methodology. Bernard Williams' distinction between "moral" and "ethical" in *Ethics and the Limits of Philosophy* (2011) as well as Vida Yao's (2023) reply to D'Arms and Jacobson's "The Moralistic Fallacy: On the 'Appropriateness' of Emotions" (2000) come to mind as examples of arguments with a similar arc. Williams' is about the philosophical study of morality, whereas Yao is speaking more narrowly about the moral significance of the emotions that is natural when one adopts a virtue ethics approach.

that trust's moral significance and value have been overemphasized. Those pushing for the demoralization of trust seem to either hold that the latter follows from the former; or they fail to appreciate the significance of the difference between the two claims. I will demonstrate that if we demoralize trust, we will lose sight of the full range of ethical goods that trust makes possible.

Though several prominent philosophers of trust have indicated they hold this position, Matthew Bennett (2021) is the first to explicitly argue that trust ought to be demoralized.[100] In his essay "Demoralizing Trust" he never explicitly defines what he means by "demoralize", but given the argument he provides, I take him to be suggesting that trust should be evaluated primarily through an epistemic lens on the grounds that it only contingently produces good effects. Rather than attempt to reconstruct a definition of what "demoralizing trust" means to Bennett, I will instead consider the three reasons he offers in his attempt to push us in this direction. Then, I will demonstrate that his demoralized account of trust fundamentally misconstrues the role that trust plays in a relationship of perennial philosophical interest: friendship.[101]

The following are reasons, offered by Bennett, that support the move to demoralize trust:[102]

1. Trust can be betrayed or disappointed without moral culpability (i.e. that is possible for our trust to go unfulfilled despite the trusted meeting all relevant moral expectations)

[100] Amy Mullin (2015) writes, "a good theory of interpersonal trust should not moralize trust. It must be understood that trusting another and being trustworthy are not always morally good, and that distrusting another and betraying someone's trust can sometimes be morally good" (Mullin 2015, 316). Onora O'Neill (2020) expresses her doubts about the value of trust when it is placed in those who are not trustworthy and questions those who suggest we need to "restore trust". Though Karen Jones is more optimistic about trust's value, she departs from her earlier account of trust by offering a non-moralized account of trust in her later work and argues that trustworthiness is not a virtue, only not for the reasons typically supposed (Jones 2012, 78-84). Katherine Hawley is yet again a notable exception. In her 2014 essay, she suggests that trust is fundamentally a moral concept when she argues that the distinction between trust and reliance matters because being trustworthy is a category of moral assessment, whereas being reliable is not. Olli Lagerspetz (1998) and Thomas W. Simpson (2023) are two others that resist this push and I will discuss in their views next chapter.

[101] The demonstration I will offer is a summary of a more thoroughgoing analysis of the undesirable consequences a demoralized account of trust has on our understanding of its role in friendship that I offer in an essay currently under review.

[102] Note that each reason offered by Bennett tracks the ideas that have led other philosophers to express that they share this view, even if they do not defend it outright.

2. We can trust others to do more than morality demands.
3. The expectation of goodwill or the attribution of some other moral quality is neither a necessary nor a sufficient condition for trust.

The implication is that these three premises should lead us to conclude: trust ought to be demoralized.

Bennett spends much of the essay building out an argument that purports to demonstrate the third reason, so my responses to each of these reasons will be proportional. It is worth noting from the outset that I only disagree with the second reason. Though I agree with the first and third reasons, I do not think it follows from them that we ought to demoralize trust, especially once we consider what we lose in doing so.

Let's start with the premise that I disagree with. Bennett argues that we often trust others to do things that go beyond what is morally required of them. However, Bennett does not sufficiently spell out any procedure for how to determine what a person is morally required to do. He also does not consider the possibility that what is morally required of a person partly depends on their particular relationships.[103] What is morally required of you in some relationships may be inappropriate or supererogatory in others. Similarly, virtue ethicists tend to appreciate how our specific contexts (including the social roles we occupy) also shape what is morally required of us. Bennett does not discuss how our relationships might shape what is morally mandated, but he does adopt the language of obligations when discussing this point, implying that there are some universal moral obligations as well as some that are specific to those who incur them voluntarily.

The example Bennett relies on to make this point shows that something is amiss. To illustrate the idea that we can trust people with more than morality demands—which on his account amounts to trusting them to do more than what they are obligated to do—Bennett asks us

[103] I take this to be one of the foundational tenets of care ethics.

to imagine we have a colleague that is organizing a conference. Bennet considers trusting a colleague to organize the conference to be trust that does not demand more than morality requires because the obligation is incurred by the fact that they have this job. However, he argues that one ventures into the territory of trusting their colleague with *more* than what morality demands if they trust them to do a *good* job organizing the conference (Bennett 2021, 519). He argues that in such cases, our confidence in the other is secured by nonmoral features of the one-trusted, such as a belief that your colleague sees the value in running a good conference or is committed to making the most of the project.

Are such beliefs about our colleague really "nonmoral" features? Though the fact that your colleague possesses the virtues of ambition and dedication is independent from the fact that your colleague has certain obligations, I see no reason to assume that morality cannot involve considerations of both our duties and our character traits. Furthermore, let's say that your colleague finds out that you "trust" them to organize the conference, but then hears that by this you do not mean that you trust them to do a *good* job of organizing the conference. Your colleague, in addition to being rightly offended, could argue that if you aren't expecting them to do a *good* job, you are merely relying on them to organize the conference—not *trusting* that they will. I'll return to the importance of optimistic expectations in my reply to Bennett's third reason for demoralizing trust.

There is also the unsettled issue of what makes a "good" conference different from one that simply happens. While organizers can certainly do better or worse jobs coordinating the logistical elements of the conference, if such logistical tasks are done in a "good enough" way, the conference *will* happen. The content and delivery of the presentations at the conference will arguably have a greater impact on its quality. Though the organizer can try to pick the best

possible presentations, the quality of the presentations is largely a matter beyond their control and, therefore, not the kind of thing for which it would be appropriate to blame your colleague.[104]

To summarize: the argument that we sometimes trust others to do more than what morality demands is only true if we ignore the variety and nuance of duties that accompany our relationships and deny the moral significance of character. Such a statement also offers a threadbare account of obligations: the quality or way in which our obligations are met are of little to no importance on Bennett's account. In other words, the statement that we often trust others with more than morality demands only holds if morality isn't all that demanding. Whether or not an obligation is met is *an* important criterion for evaluating the moral worth of an individual's action; but to suppose that it is the *only* criterion leaves us with an impoverished view of the complexities of moral life and with little guidance about what to do when our obligations conflict.

Let's return to the first premise. Bennett (correctly) argues that there are cases where trust can go unfulfilled without either party being legitimately blameworthy. However, it appears to me that these cases can easily be accounted for by appealing to the Aristotelian idea of a "mixed case" (NE: 1110a-1110b). In cases where a betrayal of trust is not culpable, it is either because the external circumstances and conflicts that arose out of them were such that neither the one extending the trust nor the one who intended to meet it could account for them or because the expectation was unfair to begin with. Understanding the conditions under which betrayals of trust are excusable requires us to look at what was entrusted, the circumstances that befell the

[104] Supposing, of course, that they did their due diligence and chose the best of the abstracts or proposals submitted to the conference; they could be blameworthy if, say, they chose presenters based on how wealthy or close of friends they are.

one-trusted, and the choice they made instead. It is not hard to find examples that illustrate this point. Recall the moving example from the last chapter: your friend, who had promised to help you move a few weeks ago is nowhere to be found when moving day comes. You get a call from him several hours later, and he explains that one his way over, he found out his sister had been in a car accident, and so rushed to the hospital to be with her. Anyone not willing to excuse their friend in this case would themselves be blameworthy for expecting their trust to be met in these extenuating circumstances. Note that this failure to meet one's trust is excused because of the sibling relationship. If he chose to visit a stranger's sister who had also been in a car accident that he had no direct relation to, you might rightly blame him for this choice. Note also, that had your friend showed up and revealed to you that he had chosen to be here to help you move instead of being with his sister in the hospital, he would be blameworthy *for* meeting your trust.

Cases where trust goes unmet because of a misjudgment about the competence or motivations of the one-trusted on the part of the one-trusting can fall into the "mixed case" category if the cause of the misjudgment can also be chalked up to some relevant feature of the circumstances. However, in cases where the one-trusting could have avoided their misjudgment or should have known better, there is blame—only in the opposite direction. The one-trusting is blameworthy. This parallels the distinction Aristotle makes between acting *from* ignorance and acting *in* ignorance. Furthermore, when external circumstances are the reason for the betrayal of trust, any expectations about the goodwill and competence of the one who failed to meet the trust are not lost. Bennett accepts that circumstances can sometimes excuse us for what would otherwise be a blameworthy betrayal of trust. However, since there are in principle excusing conditions for *all* actions, and not just those that pertain to meeting trust, it is hard to see how the fact that trust can be betrayed without moral culpability lends support to the conclusion that trust

ought to be demoralized. To circumvent this, Bennett either must apply the same logic to all other actions (thus leading to the demoralization of other practices that we might want to keep "moralized") or prove that the possibility of exceptions to culpability is unique to trust (which seems insurmountably difficult).

Now to address the third reason. Bennett dedicates more space in his essay to proving that what he calls the moral motivation principle (MMP) is neither a necessary nor a sufficient condition for trust than he does to proving the first two of the reasons. This suggests that he believes this point carries substantial weight in favor of his conclusion and that undermining this premise will show that the argument in favor of demoralizing trust is weak. He defines the MMP as follows: *when X trusts Y to Φ, X's willingness to rely on Y's future Φ-ing is secured by X's optimism that Y has moral motivations that will lead Y to Φ*" (Bennett 2021, 518). He argues that nearly all prominent accounts of trust are guilty of making this (purportedly) faulty assumption.[105] He illustrates this point with two examples (originally offered by Mullin), which are designed to show that trust can exist without an expectation of goodwill or optimism/confidence in the moral qualities of the one trusted.

The first example involves chess partners. Bennett argues that one can trust their chess-playing partner to play the game competitively, to abide by the rules of the game, and to show up for scheduled matches on time all without attributing a moral quality or obligation to them or expecting that their chess partner bears them goodwill. The second example is the civil adversary case, which asks us to imagine someone with whom we are in an adversarial relationship, but whom we nonetheless trust to be polite and civil towards us in the presence of others. Bennett and Mullin both claim that these two examples show that trust does not require a belief about the

[105] Bennett classifies all the following as accounts that rely on this principle: Jones (1996) McLeod (2002), Holton (1994), Hieronymi (2009), Kriton (2020) and Nickel (2007).

moral qualities of the one-trusted (i.e. that the trust in these cases does not rest on MMP). Instead, they ground the trust seen in these cases in the one-trusting's beliefs about the one-trusted's commitments. In these cases, the relevant commitments are to playing a good game of chess and to the social norm of civility. Mullin acknowledges that the trust in these cases does seem to involve the attribution of certain virtues to the one-trusted, but adds that the possession of a virtue does not preclude it from being put at the service of evil (Mullin 2015, 322). Bennett argues that both cases show the MMP is not at play. I find Mullin's and Bennett's analysis of these examples misleading.

Let's take Mullin's analysis of the civil adversary example first. A person who possesses a virtue (or even several virtues) certainly can do evil things, but this is not the same as using a virtue in the service of evil. The notion of what it means to act virtuously rules out this possibility, as doing a virtuous act requires the agent to have a virtuous disposition and a virtuous intention. The civil adversary who acts politely and disguises his animosity toward you when you find yourselves in the same company may do so with the hopes of catching you off-guard later or with the aim of annoying you. However, he may also simply wish to avoid making a scene.[106] Furthermore, and contrary to Bennett's analysis, the attitude adopted toward the adversary seems more like distrust than trust. It is possible to expect to expect that someone will do something without trusting that they will. This case implies that, given the history of our relationship, we have reason to be suspect of our adversary's behavior. In short, it might be more apt to say that, in the civil adversary case, we rely on our adversary to be civil in public, even

[106] Actively sustaining his adversarial relationship with you may be less important to him than keeping face or maintaining his reputation. Furthermore, what is immoral or vicious about his behavior is the deception involved. Politely interacting with you around others is not in and of itself an immoral or evil action. What might make it "evil" has nothing to do with the display of civility but would rather consist in the vice of dishonesty or cruelty that lurks behind the overt actions. In the scenario where the intention behind your adversary's civility is to maintain social harmony, one could argue that your adversary is in fact demonstrating practical wisdom by prioritizing the happiness of the group over his desire to be adversarial toward you.

though we otherwise distrust him. Even if one wishes to maintain that this is an instance of trust, it is at best an example of narrowly confined trust as it is restricted both in the scope of the action (being civil) and to a particular context (when we find ourselves in the same company). Note that we can only see this complexity when we attend to the particular relationship between A and B and the specific context in which the trust takes place.

In the chess partners case, the circumstances are similarly restricted to the domain of chess-playing. One obstacle lies in how the relationship in the example is presented. Our encounters with others are sometimes limited to specific settings, but they are rarely atomized in the way this example suggests.[107] To truly confine our interactions with our chess-playing partner solely to the domain of chess-playing, she would have to appear at the location where the match is held, speak only about the matters directly related to the match, and leave immediately after the game ended. While it is possible to imagine someone who sits down for a chess match without a word, plays in silence, and then immediately leaves, interacting with a person like this would be awkward and does not represent what most of our encounters with others are like. Nonetheless our trust in this odd character still entails that we attribute the epithet of "being a good sport" to her or believe that she possesses the virtue of sportsmanship.

I propose that at some level, describing someone as a good sport or taking someone to possess the virtue of sportsmanship implies that one expects at least a minimal degree of goodwill. In this case, the presence of goodwill would make the difference between a chess

[107] One might argue this has been changing, particularly for younger generations, and especially since the Covid-19 lockdowns. This suggests that the alienated and depersonalized modes of interaction that we are becoming more accustomed to may be an obstacle to fostering a climate of trust. If the relationship between the chess partners resembles a game played against a computer, it would become a relationship of reliance. Chess is an especially apt (and difficult) example to consider because it is a strategic game that involves predicting what one's opponent will do next.

partner whose aim is to win and the chess player whose aim is to make sure that you lose.[108] The example offered is slightly more plausible if we only encounter this chess partner in a single match, but it is hard to imagine that this narrowly confined expectation of goodwill would not spill over into a more general sort if we played chess regularly with this person, especially in the absence of evidence that she should not be trusted in other ordinary ways in which we usually trust strangers.

Bennett also mentions an example, originally offered by Baier, that many philosophers have taken as conclusive evidence that goodwill is neither a necessary nor a sufficient condition for trust: raising a white flag in battle. The thought is that enemies on opposing sides of a war could not possibly bear one another goodwill, but they can still trust that the other side will recognize and respond appropriately to signal of surrender. The idea of enemies fighting in a war bearing each other goodwill does sound counter-intuitive on the surface. However, I am inclined to interpret the symbolic gesture of raising a white flag as akin to making a plea for or appeal to one's enemies' goodwill. Though this expectation of goodwill is narrowly confined (I may trust only that they will stop firing), its presence is notable, especially when we consider the alternative. As Baier observes: "For some kinds of enemy (perhaps class enemies?) one will not trust even with one's bodily safety as one raises a white flag, but one will find it 'safer' to fight to the death. With some sorts of enemies, a contract may be too intimate a relation" (Baier 1986, 259). My point is: the expectation of goodwill is not as demanding as Bennett lets on because what an expectation of goodwill demands is indexed to the type of relationship we have with that person.

[108] For those who do may not immediately see a difference, consider the following: a person whose sole aim to ensure that you lose at all costs would be more willing to cheat at the game than someone who was aiming to win (and therefore cares more about being disqualified).

To me, it seems much more intuitive to see the chess partner and white flag cases as expectation of goodwill that are narrowly confined, perhaps even as narrowly confined as the specific domain that pertains to the entrusted object or action. While I believe it may be possible to defend the position that trust always involves goodwill, one does not need to accept my argument to see that the claim offers dubious support for the idea that we ought to demoralize trust. We can agree with Bennett that the MMP is not a necessary or sufficient condition for trust and still see that this way of framing the question obscures the fact that the presence of goodwill *is* an expectation that characterizes *most* instances of trust and, more importantly, that it features prominently in instances of trust that could be described as paradigmatic, such as in the mutual trust between friends. Therefore, even if the statement is true, it still does not provide us with sufficient reason to demoralize trust. It does, however, provide us with reason to reconsider the necessary and sufficient conditions approach to conceptualizing trust.

I've now shown that each of the three reasons Bennett puts forth as points in favor of demoralizing trust depends on a narrow conception of what counts as "moral" and provides insufficient reasons for accepting his conclusion. To conclude this section, I will take a closer look at Bennett's example of a fading friendship to show that the demoralized account of trust he offers in the essay has the undesirable consequence of misconstruing the role of trust in friendship. Bennett's account of trust amounts to the following: *When X trusts Y to Φ, X's willingness to rely on Y's future Φ-ing is secured by X's belief that Y has (psychological) commitments that will lead Y to Φ (*Bennet 2021, 531). The example is intended to show again that we can trust others with more than morality demands and that trust can fail without any moral culpability. I've quoted it at length below to capture its rich detail:

> Two close friends—I will call them Ben and Andrea—grew up together in the same town, where they still live, and share many of the same interests and

> values. They confide in one another with very personal and sensitive matters and they trust each other to support them through difficulty and join them in celebrating good news. One day, Andrea leaves to spend a year abroad. Ben stays home. The two stay in contact in an effort to sustain their friendship over distance. But when Andrea returns, she finds that Ben's character has changed, not so much as to make Andrea suspect something unusual has happened, but enough for her to think he is a different person that he was before she left. His tastes have changed, he no longer finds Andrea's jokes funny, and he has lost interest in the hobbies they used to share. Ben has also made new friends; Andrea thinks they are nice enough, but she does not have a lot in common with them. In the months that follow Ben is less inclined to accept Andrea's invitations. His shared interests with Andrea are vanishing, and his warm feelings for her as a friend have cooled. Andrea finds that conversations by phone, text, or email are increasingly prompted and led by her. Eventually Ben turns down all of Andrea's invitations to spend time together, a long time passes without any communication between them, and the friendship is effectively over. (Bennet 2021, 520).

A few sentences later, Bennett adds that, if we think that Ben is let off the hook too easily, we can imagine a variation of the example where Ben tries to rebuild the friendship, and when that fails, that he takes the time to explain to Andrea (in a caring way) that his feelings about continuing to be friends have changed (Bennett 2021, 521).

To begin, given that Ben and Andrea are portrayed as having been close friends from childhood onwards, I think Ben only escapes moral culpability in Bennett's slight modification to the initial example where Ben explains his changing feelings about their friendship to Andrea.[109] Bennett construes the trust involved in Ben and Andrea's friendship as those commonly seen in similar intimate, interpersonal relationships. Yet because of how psychological commitments fit into the entrusting framework that his account is built on, one is forced to describe the trust one has in their friend as a matter of what one expects their friends to do for them. Bennett attempts to offer a representative list of what these expectations are, including things like: "trusting relationships with loved ones involve trusting that the other will

[109] This is not to say that ghosting Andrea would make Ben an irreparably wicked person, but it would make him a pretty lousy friend.

prioritize us enough to come to our aid", "to take pleasure in our happiness", "to keep our confidence, act with sensitivity regarding whatever it is that pains us", "to be charitable with us", and to be "less quick to judge negatively what we say and do" (Bennett 2021, 520). While the items on this list are applicable to many (good) friendships, it does not reflect the experience of what trusting our friends is actually like. When we say, "I trust her, she's my friend", we are not necessarily tying our trust to a specific domain, object, or action. Typically, we intend to suggest something about *who* our friend is and not *what* they can be relied upon to do. The relation between trusting a friend in this broader sense is related to the ways we trust our friends in these more specific ways, but the relation between the two is more complex than Bennett lets on. Furthermore, if the trust we have in our friends is genuine, it seems to prohibit in advance the degree of specificity seen in the items on Bennett's list. [110] When we tie the trust we have in our friends solely to the things we entrust them with, we undermine the reasons we value the trust in our friendships. Even if we could create a list of Φ's that completely capture everything that we might trust a friends with, it also seems to be the case that we trust different friends with different things. So, it is implausible to assume that any list could capture what trust looks like in

[110] References to "genuine" or "true" friendships imply that it is possible to have "fake" or "untrue" friendships, as well as friendships that fall in between these two ends of the spectrum. In Books VIII and IX of the *Nicomachean Ethics*, Aristotle offers us one way of distinguishing between friendships of these various kinds—our "true" friendships on his account are those of virtue. Given his discussion, friendships of utility and pleasure seem to end up not really being friendships at all. While these categories are helpful in some ways, they do not map easily onto what friendships are like in reality, as many of our friendships seem to be mixture of these categories. Furthermore, the requirements for a friendship of virtue are rather demanding and may leave us wondering if anyone has ever had a true or genuine friendship. While I think that a distinction between "true" and "fake" friendships exists, I am still uncertain about how to articulate what makes the difference between the two. Part of the difficulty lies in the fact that there are many kinds of friends whom we trust with different kinds of things. A larger part of the difficulty though lies in that fact that the future of our relationships, including friendships, play out in unpredictable ways. The problem is analogous to one Aristotle encounters early in the *Nicomachean Ethics*, namely, that it seems impossible to determine whether someone is virtuous while they are still alive. Friendships, like people and their characters, change over time, so it may be the case that we can only know for certain whether a friendship was "true" or "genuine" in retrospect.

every friendship. In thinking too much about what we trust our friends with we thereby treat friendship as instrumentally, rather than intrinsically valuable.

Bennett's description of the trust seen in friendships is that it exemplifies the very concern Baier expresses about the three-part framework when she introduces it: "taking trust to be a three-place predicate (A trusts B with valued thing C) will involve some distortion and regimentation of some cases, where we may have to strain to discern any definite candidate for C" (Baier 1986, 236). Baier's concern points to the fact that while it may be true that a more general form of trust supports the narrow instances of entrusting that occur in friendships, what this more general form amounts to is nebulous and difficult to pin down precisely. What matters for us is that in friendship a single commitment is unlikely to be up to the task.

On Bennett's account, the general sort of trust seen in Ben and Andrea's friendship would be described as follows: *When Andrea trusts Ben to continue being her friend, Andrea's willingness to rely on Ben being her friend in the future is secured by Andrea's belief that Ben has a (psychological) commitment to their friendship that will lead Ben to continue being her friend.* The set-up here is proleptic—trust in the continuation of their friendship is secured by a belief that Ben is committed to their friendship in ways that will motivate him to act as a friend. Furthermore, Bennett claims that Ben disappoints Andrea's trust by ending their friendship and implies that he fails to act out of a commitment to their friendship. Yet, at no point does this appear to be the case.

Let's look at Ben's actions in the example. Ben turns down a few of Andrea's invitations, and he leaves it to her to prompt conversations. Eventually he tells her that he no longer wants to continue their friendship. The first two actions do not seem like they could be described as a failure of this general trust or as a failure to meet his commitment—friends will understand that

their friend has other commitments that may make it difficult for them to accept all their invitations and that there may be periods where their schedule does not make it easy for them to initiate conversations. A friend will also not over-commit themselves when making plans, as doing so would indicate a lack of respect for their friend's time. Therefore, neither of these actions on their own seems to amount to disappointing Andrea's trust or failing to act in the way that is expected given that she believes he has a commitment to the friendship. The final action—telling Andrea about his change of feelings regarding their friendship—is the only remaining candidate. However, in explaining his change of heart, Ben has continued to act as Andrea would trust her friend to act—with concern for her feelings, respect for her time and for the effort she has put into sustaining their friendship—so it does not follow that he betrayed, disappointed, or failed in any way to act as Andrea trusted him to (i.e. out of a commitment to their friendship).

Trust is clearly important in friendships, but identifying an object or domain of the trust we see in friendships is extremely difficult to do. When we try to do so, we are left with a picture of friendship that instrumentalizes the value of the trust found within them. The value gets cashed out in terms of what *I* expect my friend to do *for me*. Bennett's account suggests that what is entrusted is the same as the belief that secures the trust in this case, thus proving Baier's observation that we may have to "strain" to discern a definite candidate for "C" or "Φ" correct. Rather than describe the trust we have in our friends through our beliefs about their commitments, we should look for an alternative explanation of how trust in friendship operates and consider what else could ground the general trust we experience in our friendships. We can start by considering how the fading of a friendship changes what we owe to our friends and what (if anything) we can reasonably still trust them with.

While there are certain things that Andrea should not trust Ben to do given the fading of their friendship (e.g. attend her dinner party), it seems reasonable for her to trust that Ben will keep secret the things that she told him in confidence back when they were still friends. If he were to expose an embarrassing secret or private information about her, Andrea would rightly feel as though Ben betrayed her trust, even if she no longer considered him a friend. So, it is clear that we owe our former friends something, but to what extent does this bear on the degree of trust we can reasonably continue to have in them? Aristotle ponders this same question and provides some guidance:

> Surely he should keep a remembrance of their former intimacy, and as we think we ought to oblige friends rather than strangers, so to those who have been our friends we ought to make some allowances for our former friendship, when the breach has not been due to an excess of wickedness. (NE IX: 1156b32-37)

Bennett states from the outset that the faded friendship between Ben and Andrea is a case where there is no moral culpability; although Ben's character has changed, neither he nor Andrea have become wicked. Therefore, they still owe each other more than what they owe strangers (at least, according to Aristotle, who is admittedly vague in this passage). Aristotle states that we ought to make "some allowances" for former friends that we wouldn't make for strangers, but he does not tell us what these allowances are, nor where they begin or end. Given that Aristotle implores us to not demand more precision than our subject matter admits of a complete list of the allowances we should make for our former friends is not something his ethical theory will offer. However, insight into what these allowances might include can be found in the connection he establishes between friendship, trust, and goodwill.

Aristotle argues that we can bear goodwill to those we have never met, so long as we judge them to be good or useful (NE VIII:115b35-36). This implies that goodwill can be rather minimal with respect to what it demands. If we can bear goodwill to those we have never met,

then it seems reasonable to say that we can also bear goodwill to acquaintances and to those we have fleeting encounters with, provided that we judge them to be good or useful. Recognition of *mutual* goodwill is, for Aristotle, a prerequisite to friendship. We can thus infer that while we do not owe our former friends the same degree of intimacy as we previously did, what will count as bearing goodwill to them requires more than it will for us to bear goodwill to strangers. We can say that, on Aristotle's account, goodwill is a necessary, but not sufficient condition for friendship. This interpretation is further substantiated by Aristotle's descriptions of goodwill as "a beginning of friendship" and "inactive friendship" (NE IX:1167a3-11). Goodwill is a necessary condition for friendship, but it is not the only one Aristotle identifies. The other conditions include that each party in the friendship must be "lovable and trusted by the other" (NE VIII:1156b25-30). At the point when prolonged mutual goodwill reaches the point of intimacy Aristotle argues that it "becomes friendship", and importantly, not the instrumentally valuable kind rooted in pleasure or utility, but the intrinsically valuable friendship of virtue.

Let's return to the version of Bennett's example where Ben considerately explains his change of heart to Andrea. After this conversation, Andrea should certainly no longer trust Ben to do things like make plans to spend time with her or attend her dinner party. Yet, it still seems reasonable for Andrea, in light of their many past years of friendship and the absence of wickedness in both their characters, to trust that she and Ben will continue to mutually bear each other goodwill. What this suggests is that Andrea's trust in and friendship with Ben is not secured by her belief that Ben has a commitment to sustaining their friendship, but rather by an expectation that Ben bears her goodwill, which is at no point disappointed by Ben.

Insisting on a connection between the trust we see in friendships and the expectation of goodwill yields a more accurate picture of what trust in friendships is like. Furthermore, it does a

better job of explaining the general sort of trust that supports the more specific things we trust our various kinds of friends with. An additional advantage of accounting for the trust seen in friendships in terms of an expectation of goodwill is that we don't lose sight of the uncertainty that accompanies the future-oriented nature of trust. Instead, the uncertainty that will continue to characterize Ben and Andrea's relationship after their friendship has faded remains at the forefront. I will illustrate this point by adding to Bennett's example.

Suppose Andrea, though saddened by the loss of her friendship with Ben, understands Ben's reasons and thanks him for his honesty. She decides to leave her hometown and travel abroad again. She makes new friends, discovers new interests and hobbies, and makes new commitments. A few years go by, and Andrea accepts a job in a city far away from the hometown where she and Ben grew up. On her first day, she settles into her cubicle and, to her surprise, discovers that it happens to be right beside Ben, who accepted a job at the same company a few months earlier. In this scenario, it is reasonable for Andrea to expect, at a minimum, more goodwill from Ben than from her other co-workers. While their former level of friendship may never be restored, it is also not impossible that it might. Perhaps some of Andrea's new interests overlap with Ben's now that time has passed. Maybe they both get along with the same co-workers and find themselves going out in the same groups on weekends. Perhaps they find themselves reminiscing about old times during their lunch breaks. Their revived friendship may only last as long as they are working together, or as long as they are in the same city. However, we can also imagine a scenario where their reminiscing restores their friendship, which goes on to last into their twilight years (it may even lead to romance and the two ending up "happily ever after"). The thought of the (comparatively brief) time they spent as "not friends" might become something they look back and laugh at. Even if their friendship is

never restored, they should still trust that they bear each other mutual goodwill. Each of these scenarios is at least as plausible as the one Bennett offers in his original example, yet they all become impossible if we suppose that by "ending" the friendship Andrea can no longer expect Ben to bear her goodwill.

What I have demonstrated here is that the fading of friendship is (normally) not a sufficient reason to completely cease trusting a former friend, and that the fading of the friendship in Bennett's example does not really entail a failure of trust. What this suggests is that the trust we have in our friends is not secured by a belief about our friends' commitment to our friendships, but rather by an expectation of reciprocal goodwill. The entrusting framework Bennett and others assume in their accounts of trust leaves us with an incomplete picture of what trust is like and why we value it in our friendships, as well as in our relationships more generally. While the entrusting model highlights the instrumental value of trust, it does so at the expense of neglecting the other ways in which trust is valuable. In particular, it ignores the role that trust plays in the formation and sustainment of our intimate interpersonal relationships and instead focuses on its epistemic dimension. The emphasis that is placed on knowing the object or domain of our trust overlooks the fact that by trusting someone we forfeit our right to know how they will meet that trust. As a result, the push to demoralize trust obscures the ways in which these trusting relationships shape our self-identity and contribute to a flourishing life.

The Turn Toward Trustworthiness

The second trap that often befalls doxastic and affective accounts is the turn toward trustworthiness. Some philosophers of trust turn to trustworthiness to explain how trust functions.[111] In recent years, many have turned toward theorizing trust through trustworthiness

[111] Karen Jones's (2012) trust-responsiveness account is the most explicit example of an account of trust that makes this move.

under the assumption that being able to assess trustworthiness accurately can help us address the practical problem of distrust. Those that do tend to focus on trust's epistemic value.[112] Attending to our assessments of trustworthiness does bring into focus the ways in which trust and distrust are subject to looping effects. However, there are two reasons why this turn, on affective and doxastic accounts, fails to explain how trust operates. First, it implies that moving from distrust to trust only requires making the would-be trusted more trustworthy. Second, it tends to repeat the mistake of neglecting how variables like our external circumstances, our social and historical contexts, as well as our individual past experiences affect our tendencies to trust and distrust.

Judging or believing someone to be trustworthy can ground or provide sufficient evidence for trusting them; however, it is not the only trait or virtue that moves us in this way. Depending on the context, we may trust someone because they are compassionate or creative or charming without believing that they are trustworthy. Trustworthiness is a virtue that is typically associated with our assessment of another's testimony. When we describe someone as trustworthy, what we often mean is that they speak honestly or truthfully. One might argue that this issue can be resolved easily if we reject this interpretation of trustworthiness and instead adopt a definition of the virtue that has a broader scope. I am sympathetic to this approach. Whether or not someone is trustworthy seems to depend not just on the fact that their testimony is characterized by honesty or truthfulness, but also on whether they follow through on their promises, meet their implicit and explicit commitments, etc. Though I am willing to concede that there is a behavioral as well as testimonial component to trustworthiness, expanding the definition in this way makes does not make the rule of *only trust the trustworthy* less tautological or circular. In other words, it does not yield principles or rules that are helpful in practice.

[112] See Almassi (2022), Bennett (2022), Branch (2022), Branch and Origgi (2022), Brennan (2020), and Camporesi, Vaccarella and Davis (2017).

Understanding what trustworthiness is or what qualifies an entity as trustworthy to greater or lesser degrees will—at best—give us half of the story.[113] The assumption that affective and doxastic accounts that turn to trustworthiness appear to share is that a trusting relationship is valuable or morally good if both parties are trustworthy. However, trust is a relational phenomenon, involving at a minimum two parties and the health of a given trust relationship depends on *both* parties having the right attitudes, habits, dispositions, and virtues.

Trustworthiness is only one half of the pair of virtues both parties in the relationship must possess. In a healthy trust relationship, there will not only be the virtue of trustworthiness, but also the virtue of trustfulness. Sungwoo Um, one of the few contemporary philosophers to appreciate the significance of this virtue, describes the trustful person as "one who keeps the proper balance between distrustfulness and gullibility" (Um 2024, 323). A trustful person makes good judgements about who to trust when, and with what. Moreover, they do not merely possess the relevant epistemic abilities that allow them to make accurate predictions about how others will behave; their judgements about who to trust issue from a distinct attitude. This attitude, Um argues, involves seeing the one-trusted as a "free agent". Without it, a person is more accurately described as "prudently reliant" than as trustful. This point is related to the argument that the compliment-paying feature of trust cannot be explained without appealing to some affective component involved in trusting. I will discuss the significance of the virtue of trustfulness and its relationship to self-trust in the next chapter. The point I want to make for now is that in turning to trustworthiness to find answers to questions about how trust functions (and particularly how we move between distrust and trust) when combined with the unidirectional nature of the

[113] Kelp and Simion (2023) offer a bi-focal account of trustworthiness—where trustworthiness simpliciter is distinguished from trustworthiness to Φ—which provides a helpful lay of the land with respect to defining trustworthiness. However, their account runs into challenges that parallel existing questions about how interpersonal, social, and political forms of trust are related.

entrusting framework seen in affective and doxastic accounts leads to a roadblock. When we only examine trust from the perspective of the one-trusting, it is easy to overlook the number of variables that affect our tendencies to trust and distrust.

Another issue with the turn toward trustworthiness is that it assumes that the things that will improve or signal trustworthiness are stable. In reality, these practices are contingent and depend heavily on context, an individual's past experiences, and the shared history of a given relationship. A climate of distrust can make otherwise accepted ways of signaling trustworthiness appear suspect or performative to the intended recipients of that signal. Signaling trustworthiness in such situations might backfire—deepening the distrust or further eroding any remaining trust instead of restoring it. For example, a willingness to make promises and attempts to be transparent are often effective mechanisms for establishing trust in a relationship where neither trust nor distrust exists. However, if the relationship in question is already characterized by distrust, these sorts of gestures might be (incorrectly, but understandably) interpreted as an attempt at further deception. We do not need to look far for real-world examples that demonstrate this point. If a particular community distrusts the government, or even just a particular subset of the government, because of some harm that it inflicted upon the community in the past, the members of the community have good reason to be wary of accepting a new promise and taking what is said at face value. If there is a long history of wrongdoing or deception, the suspicion that the transparency being displayed is merely a ruse is an even more reasonable response.

Turning toward trustworthiness yields some helpful insights about some reasons that may lead us to trust others. However, an accurate interpretation of what trustworthiness entails often amounts to an incomplete or inaccurate understanding of what fostering, maintaining, and

repairing trust demands. Furthermore, those who have turned toward trustworthiness have said little about the self-trust that is required for being a good judge of who is trustworthy. I return to the role of self-trust in chapter five.

Conclusion:

The findings of this chapter are summarized in the table below. While it may initially appear that a hybrid doxastic-affective account could meet all three of the requirements I've outlined, the propensity of falling into one or both pitfalls identified in this chapter suggests that even a hybrid account would also fall short of providing an adequate account of trustworthiness. The limitations of both kinds of accounts invite us to consider what might be wrong with the kind of approach these accounts adopt. Rather than trying to define trust using a set of necessary and sufficient conditions, I adopt a stance of pluralism about trust in the developmental account I put forth in the next chapter and focus on paradigm instances of trust.

	Requirement #1	Requirement #2	Requirement #3	Avoids Pitfall #1	Avoids Pitfall #2
Doxastic Accounts	Yes	No	No	No	No
Affective Accounts	Yes	Yes	No	No	No
(Hypothetical) Hybrid Account	Maybe	Maybe	Maybe	No	No

Chapter Four: The Developmental Account

"Trust is the foundation of intimacy. When lies erode trust, genuine connection cannot take place."

-bell hooks

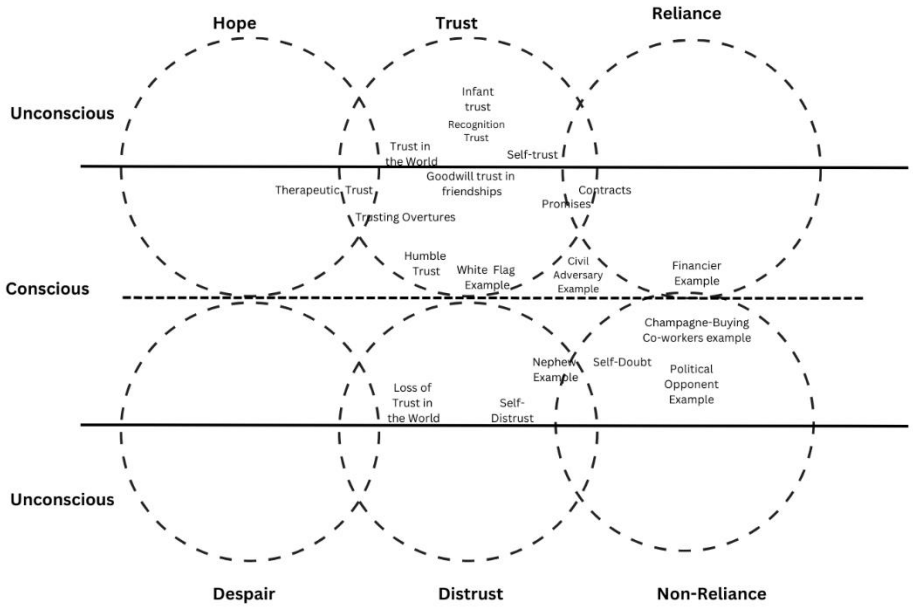
In this chapter I have two aims: 1) to further explain my account of trust, which includes showing that it meets all three requirements as well as how it differs from other accounts that adopt a stance of pluralism about forms of trust and 2) to demonstrate that trust enables three things that are necessary for human flourishing. The first of them is cooperation, which makes possible the proliferation of human activities, thereby allowing us to develop our personal interests and talents.[114] The second is the creation and sustainment of our intimate relationships. The third is a stable personal identity.

On my developmental account, trust is defined as a response to one's situation that embraces vulnerability. I argue that this embrace of vulnerability is the shared element that appears in all forms of trust.[115] I also maintain that trust can occur at varying levels of conscious

[114] While it is *possible* for people to work together without trust, sustaining such activity without trust requires other enforcement mechanisms. People may appear to "cooperate" on the surface, but once the enforcement mechanism is removed the coordinated activity falls apart. Enforcement mechanisms can range from the literal use of force (i.e. the threat of violence) to more subtle forms of coercion. In either case, the foundation supporting coordinated activity is brittle.

[115] While I realize that "response" is a bit of squishy word here, I am aiming for a definition that does not reduce trust to either a belief or an emotion. Relatedly, I use "embrace" instead of "accept" because of its association with a physical act. If we stretch the metaphor, we can imagine how the different kinds of embraces might correspond to trust in relationships with varying degrees of voluntariness. I elaborate on this further in the preface.

awareness and with varying degrees of voluntariness. Trust develops out of interactions where our needs are met. We develop the capacity to trust when our biological needs are met by others. As we develop our needs and the possible ways they can be met proliferate. Trust is one of many ways a person can respond to their situation. On my account, it falls between hope and reliance. Peripheral forms of trust may possess some of the features associated with them. Below you will find a diagram of how my developmental account conceptualizes the relationship between hope, trust, reliance, non-reliance, distrust, and despair where I've mapped the examples and kinds of trust discussed in this dissertation.



In proposing that an embrace of vulnerability is a quality shared by all forms of trust, my aim is *not* to add a sufficient condition for trust to the necessary conditions I outlined in chapter two. Instead, I propose that we conceptualize trust using the mutual trust between friends as the paradigm of interpersonal trust. There are several reasons that the trust seen in friendships makes a good paradigm. First, it acknowledges that the expectation of goodwill is a feature shared by

most forms of trust. Second, it draws our attention to the non-instrumental value of trust that comes from its constitutive role in the creation and sustainment of intimate interpersonal relationships and its enabling role in the formation of our personal identity. Furthermore, although we can step back and reflect on the reasons why we trust our friends, our trust in them often goes unnoticed when it is active—it acts as the backdrop against which our interactions with them play out.[116] Reciprocal, goodwill trust between friends also has the advantage of being a type of relationship that is well-represented in the human experience across cultures and throughout history. Finally, because friendships themselves come in many varieties, some may admit of a wide range of power asymmetries, and some may approximate perfect symmetry. In either case, the fact that friendships endure over time makes it evitable that there will be some fluctuation over the course of a lifetime. Unlike the entrusting framework, which masks the ways power asymmetries and material conditions affect our tendencies to trust and distrust, the paradigm of trust between friends reminds us that trust always occurs within a relationship where such variables are present.

Now that we have a definition and paradigm case in view, I will explain how the developmental account meets each of the three requirements and avoids the pitfalls discussed in the last two chapters. In the process of doing so, I will make evident the affinities my account has with two other philosophers who reject the necessary and sufficient conditions approach to conceptualizing trust and adopt an ordinary language account instead: Olli Lagerspetz and Thomas W. Simpson. Despite sharing fundamental commitments, I ultimately show that that what makes my account distinct from their respective proposals is its appreciation of the three distinct ways in which trust contributes to a flourishing life.

[116] We can also engage in this reflective activity to develop an understanding of what we can't trust our friends with, under what conditions they should be distrusted, and what limits there are to our trust.

The Developmental Account and Accounting for Distrust

The developmental account uses the metaphor of a climate to help explain the relationship between trust and distrust. I will show how the use of this metaphor helps illuminate the ethical and epistemic harms of distrust, as well as distrust's advantages, while maintaining that, all things considered, trust is preferrable to distrust. As a result, we will be able to see that trust is conducive to flourishing, even though it is sometimes present in morally questionable relationships or an enabler of immoral actions. Because the climate metaphor allows us to consider the full range of factors that impact our tendencies to trust and distrust—particular relationships and past experiences, more localized social norms, cultural practices and historical patterns of oppression and marginalization—it can offer a more concrete explanation of why we distrust. Furthermore, it yields insights about how we can move between trust and distrust in the real world.

Recall from chapter one that climates exhibit both patterns and are affected by several variables which themselves can interact in ways that alter those patterns. This image thus supports the idea that there are many different forms of trust and suggests that cases of overlap are possible. Local climates can fluctuate drastically in relatively short periods of time, but on a global scale significant changes are gradual and happen over a longer duration.

The developmental account characterizes trust as a response to one's situation that lies between hope and reliance and allows for overlap as we move further away from the paradigm. It follows then, that on my account distrust is a response to one's situation that lies between non-reliance and despair. Distrust eschews vulnerability rather than embracing it, and thus functions

as a barrier to the creation and sustainment of intimate relationships rather than as a foundation for them. When we think of the conceptual space between trust and distrust through this metaphor, which allows for fluctuation but admits some consistent patterns, we get an intuitive characterization of what the overlapping cases look like. In cases where distrust and non-reliance overlap, for instance, the response typically stems from a judgement of incompetence.[117] Whether this response is felt as an insult or slight to the distrusted party depends on a few factors: 1) the object or domain of the distrust, 2) the distrusted's past experience with respect to that object or domain, and 3) the shared history between the one-distrusting and the one-distrusted. Forms of distrust rooted in fear, on the other hand, overlap with despair, particularly in the ways they prompt us to act.

The room I leave for overlap between non-reliance and distrust and between distrust and despair may leave us wondering how a meaningful distinction between the two can be made. My developmental account suggests that, like the primary difference between hope and trust and between trust and reliance, the distinction boils down to the ways in which their responses avoid vulnerability. Not relying on someone is a response that follows from calculating that it is statistically improbable that a person will act in a certain way. It is to take someone's behavior into account the same sort of way we would take into account the weather forecast. In this way, it preserves the insights about trust seen in some reactive attitude accounts. When we distrust someone, our expectations are not merely that it is unlikely or irrational to count on them to do what we would like them to do, but rather that they cannot be expected to do so because of *who*

[117] Recall my example from chapter two of the couple that distrust their teenage nephew to babysit their infant for the weekend on page 50, footnote 47. The nephew, who judges himself to be competent might experience their refusal as distrust of a more insulting sort if he felt as though his relative's judgement of incompetence emerged from their evaluation of *who* he is as a person and not from a judgement about *what* things it is appropriate to entrust a teenager with.

they are.[118] The assessment of character that is built into our judgement that someone ought to be distrusted explains why being distrusted can create wounds and injure, whereas not being relied on typically does not.

Furthermore, the conceptual space between trust, reliance, non-reliance, and distrust is maintained by the fact that trust and distrust, as well as reliance and non-reliance are mutually exclusive responses to the same situation. This allows for the possibility of extremely narrow, restricted forms of trust co-existing within relationships characterized by broader, generalized distrust. The civil adversary case discussed in the previous chapter illustrates this point. If we know that we can only trust our nemesis to be civil to us in the presence of others, we can rightly be distrustful of them when we run into each other on an empty street. Forms of trust narrowly constrained in this way are too brittle to sustain intimate relationships, but they can clear the ground for a more stable foundation to be laid. If we desired to transform our relationship with the civil adversary, it follows that we ought to begin our attempt to change the relationship in the presence of others, perhaps by inviting a mutually trusted third party to mediate or by making an apology the next time we find ourselves in the same company. Thinking about the relation between trust and distrust through the climate metaphor encourages us to identify the specific variable or pattern that most directly exerts influence over our response. Moreover, it does this without overlooking the challenges or interference that the other variables pose and with the awareness that making lasting changes is a gradual process that requires sustained effort.

Viewing the relationship between trust and distrust through the climate metaphor also allows us to keep general rules about what justifies or warrants trust and distrust in perspective.

[118] In the next section, I explain how trust and reliance are meaningfully distinct by arguing that when we trust someone, we rely on them because of *who* they are. To understand this point, it can be helpful to think of our intimate relationships with others are intrapersonal projects. See Diane Jeske's work on friendships and reasons, especially her 2001 paper "Friendship and Reasons of Intimacy".

Rules such as *don't trust someone if you believe they are lying to you* (or it's alternative formulation, *distrust anyone you believe to be lying to you)* often provide dependable guidance, but they are nonetheless still fallible. Particular relationships and circumstances can be such that you can justifiably trust someone even if you *know* they are lying to you. Consider the following case:

Two FBI agents (Agent D and Agent F) are partners working together to uncover a government conspiracy that they believe involves some of their fellow agents, the military, leaders of other nations, and possibly even their direct supervisor.[119] Despite the pervasive climate of distrust that characterizes the nature of their work, the interpersonal trust (or extremely localized climate) between the two of them has proven strong. Their shared personal history of personal and professional experiences has forged what philosophers would call a "thick" trusting relationship between them. While investigating a case that initially appeared to be unrelated to the conspiracy, Agent F discovers a clue that may be the key to uncovering the identity of the mastermind orchestrating the conspiracy. He calls Agent D and tells her he is on his way back to the office with a "big development in the case" (they have an implicit agreement to avoid talking about such matters over the phone since it may be bugged). When he arrives, eager to tell Agent D about the clue, he finds her talking with Agent X who has conspicuously been assigned to help them with the case at the last minute. When Agent D asks Agent F about the big development he mentioned on the phone, Agent F lies. Though Agent D can tell that Agent F is lying and (correctly) believes he is withholding information from her, she trusts his judgement of their situation (including his decision to lie to her), pretends to believe the lie in a way that is convincing to both Agent F and Agent X, and proceeds to investigate the case using

[119] This is example is inspired by the complex relationship between Dana Scully and Fox Mulder in the TV show *The X-Files*.

Agent F's false information to keep Agent X distracted. In doing so, she cooperates with Agent F, further enhances the micro-climate of trust that characterizes their partnership and makes it possible for Agent F to continue the *real* investigation. She can do all of this *only* because she trusts herself and has a deep understanding of who Agent F is.[120]

While the circumstances in this example are far removed from the everyday experience of the average person, there are mundane cases analogous to it. For instance, we can imagine a couple who has an unspoken agreement to start leaving a party whenever the other mentions that they have to get up early (even though they *know* that neither of them ever gets up early). The ability to successfully communicate in ways that defy our basic communicative norms depends on having intimate knowledge about who the other person is and a prior history with them to appeal to.

In sum, the developmental account accounts for distrust by acknowledging that distrust can stem not only from our beliefs and emotions, but also the circumstances and history we share with particular others. A consequence of being attuned to this wide range of variables is that the account denies that there are fixed, absolute, or universal rules that can help us determine when distrust is warranted or justified. However, still allows for common-sense guidelines that encourage us to consider how well we know the person we are (dis)trusting and how our external environment might factor into the (dis)trusting relationship.

Developmental Account and Making a Meaningful Distinction Between Trust and Reliance

[120] One might object that Agent F is not really lying to Agent D if he did not intend to deceive her, but let's say he *does* intend to deceive her because the clue he discovered reveals information that would be deeply upsetting to Agent D. Furthermore, a lie (uttering of falsehood) is still a lie even if it no one is fooled. Whether Agent F is being dishonest to Agent D in example as described above is a different matter. My point in offering this example is that it is tricky to offer general rules about trust, lying and honesty.

On the developmental account, trust lies between hope and reliance. Positioning trust in this way helps us make sense of therapeutic or proleptic trust.[121] It also incorporates Baier's insight that the narrowly confined forms of trust seen in contracts and promises are more accurately characterized as limit cases between trust and reliance than as the paradigm. The fact that my account of trust allows for non-paradigmatic cases of trust to be mixed with hope and reliance brings the question of how to meaningfully distinguish trust from reliance to the forefront. On my account, the difference between trust and reliance lies in the fact that the vulnerability one is exposed to in each is different in *kind*. As I suggested earlier, this difference is related to the fact that our judgements about trust are informed by *who* the person is more than they are informed by *what* we expect them to do. When we are wrong about our judgements about the former, our relationship with that person is called into question. Depending on the particular relationship, errors in such judgements may also call into question our self-trust. When we merely rely, we do not jeopardize our relationship with the person we rely on or our capacity for self-trust in the same way. What is at stake is *only* the outcome—which can go as we expect or not. Fully appreciating this distinction requires looking at the several ways in which trust is valuable.

[121] Therapeutic trust is a term used to refer to trust that is placed in others with the *hope* or ultimate aim of cultivating the virtue of trustworthiness within them. The commonly used example is of a parent who trusts their teenager to stay home alone for the weekend, even though they realize it is possible (and perhaps even likely) that their teenager will break the house rules or otherwise take advantage of their parent's trust. There has been some debate about whether such cases qualify as trust. On the developmental account, therapeutic trust is a peripheral form of trust that lands on closer on the spectrum to hope. For more on the relation between trust and hope, see Victoria McGeer's "Trust, Hope, and Empowerment" (2008) and chapters two and three of Margaret Urban Walker's *Moral Repair* (2006). Baier also acknowledges that there is something like this relationship between the two. I am inclined to agree with her argument that "what we hope for is more than we trust that we will get" and have thus opted to place hope, trust, reliance, non-reliance, distrust, and despair on a horizontal spectrum rather than a vertical one (Baier 2009. 218). Having a picture of how these concepts are related to one another is useful for thinking through how we can move from one type to the other. Simpson (2023) refers to this phenomenon to as proleptic trust. I see no significant difference between the two, so I am using the terms here interchangeably.

Before I proceed with this examination, I want to note that, on this point, the developmental account I offer shares some affinities with ordinary language accounts of trust, which begin their analysis of trust by looking at how the concept functions in ordinary language and how it is used our everyday practices. Although there are still plenty of disagreements among the philosophers that adopt this method of inquiry, their shared initial commitment to pluralism with respect to forms of trust consistently yields the conclusion that trust admits of multiple dimensions of value and affirms that trust is indeed a moral concept.[122]

In his 1998 book *Trust: The Tacit Demand*, Olli Lagerspetz expresses his frustration with his contemporaries' attempts to define trust using a set of necessary and sufficient conditions, writing that it is "probably neither realistic nor helpful" to do so, and arguing that instead that "the meaning of 'trust' is best elucidated by looking at various situations in which the word is applied" (Lagerspetz 1998, 4). This difference in starting point ultimately leads him to argue trust should be seen as a tool we can use to invoke a perspective on human action. In the last chapter of his book Lagerspetz emphasizes that trust's connection to moral concerns is constitutive of its meaning. He concludes that:

> Trust is not a phenomenon whose existence we establish 'neutrally': we do not simply discover it independently of our own position. To 'discover' it is to see human relations in a light that requires a moral response. Conversely, it can be argued that the fact that we have moral concerns is, in the last analysis, intelligible only because we live in human relations—such as friendship—that involve trust. (Lagerspetz 162)

[122] In addition to the shared commitment and conclusions identified in this paragraph, both Simpson and Lagerspetz observe that there is an intimate tie between trusting and respecting another, suggesting that, at a basic level, trust and respect are two side of the same coin. I explore this idea at length in Cunningham (forthcoming) and will discuss the connection further in the final chapter.

As we can see, for Lagerspetz trust is the very starting point of morality. Trust provides a perspective without which our moral evaluations of human interactions would be incomprehensible.

On the other hand, Simpson argues that trust has two distinct kinds of value—instrumental and interpersonal. The instrumental value of trust refers to the practical benefits that follow from trusting others—it facilitates cooperation and, in doing so, multiplies our agency (Simpson 2023, 3). Trust's interpersonal value comes from the expressive capacity inherent to the attitude of trust. It is because "by trusting another person, I show my respect for them" that the interpersonal value of trust avoids being collapsed back into its instrumental value (Simpson 2023, 8).

Despite the differences in their respective approaches, I find both Lagerspetz's and Simpson's assertions intuitive. Furthermore, unlike doxastic and affective accounts of trust, their accounts resist the push to demoralize trust by bringing trust's moral significance to the forefront. With these other views in place, we can now proceed with the examination of the how the developmental account meets requirement number two.

The first way in which I find value in trust largely parallels Simpson's description of trust's instrumental value. It is also the aspect of trust's value that typically receives the most attention. Trust expands our agency by facilitating cooperation. By facilitating cooperation, trust makes it possible for us to produce knowledge and diversify our activities (which in turn allows for more specialization and division of labor). Another way of putting this is to say that trust allows us to distribute resources and satisfy our basic needs in a more efficient matter, which in turn increases our ability to pursue our personal interests and projects. Writing a dissertation about trust would be impossible if I could not trust that the grocery store nearby will continue to

regularly receive shipments.[123] Trust's role in facilitating cooperation on both small and large scales makes it an extremely valuable social resource. However, it is important to note that even trust in these situations involves the embrace of vulnerability even when the cooperation does not involve intimate interpersonal relationships. Trust's prudential value functions as a regress blocker—it allows us to keep calm and carry on without constantly worrying or wondering if the strangers we depend on will do their part.

The second way in which I find trust to be valuable resembles Simpson's interpersonal value of trust and Lagerspetz's contention that trust is the tacit demand that makes moral relations possible. Trust is a constitutive component of intimate, interpersonal relationships; it both enables and sustains them. As a component of intimacy, rather than an optional lubricant, trust is deeply intertwined with the kinds of relationships that are in themselves intrinsically valuable. One might argue that this is just another item we could add to the list of ways that trust is prudentially or instrumentally valuable. However, trust is "instrumentally" valuable to our intimate relationships in the way that a beating heart is instrumentally valuable to good health. Just as the absence of a vital organ precludes one from being considered "healthy", the absence of trust precludes the possibility of intimacy and secure attachment, and as a result, the kinds of relationships that are necessary for a flourishing life. Describing trust's role in enabling and sustaining intimate interpersonal relationships as "instrumental" neglects the fact that trust is a constitutive component of the relationship, making it intrinsically valuable itself despite its enabling function.

[123] Some may wonder if is possible to rely, rather than trust in this case. My answer is yes, but relying on would be less efficient, in part because there would be some worrying involved and in part because we'd more likely to "stock up". For instance, during the early months of the Covid-19 lockdowns, most people could not trust that their grocery store would receive adequate shipments, but still had to rely on it and this caused grocery stores to run less efficiently. As a grocery store employee during this time, my first-hand experience confirms this.

The final way in which trust is valuable emerges out of the second: trust makes it possible for us to have a stable personal identity. Our sense of self is maintained by a basic, latent form of trust that is engrained into our daily lives and understanding of our place in the world. We become who we are, in a large part, through our interactions with others. Accepting this does not require an endorsement of the view that our identities are relationally constructed. We can remain uncommitted with respect to that question if we accept that our capacity for self-trust is forged through our trusting relationships.[124]

Jay Bernstein's definition of can help us see this point more clearly: "Trust is the attitude of mutual recognition that comes to be through socialization processes in which the child is loved as and in order to become a person" (Bernstein 2011, 410). Humans learn the meaning of their actions through their interactions with others. In the earliest stages of life, this occurs through mimetic responses. Over time, this instills in us both the fact that we are self-determining and the fact that we are vulnerable to others. As we progress through other stages of life, we develop a sense of what we can trust ourselves to do, which in turn shapes our understanding of who we are and who it is possible for us to become. While our ability to independently reflect on and evaluate the extent to which we can trust ourselves becomes more independent as we progress into adulthood, our self-trust continues to be impacted by others throughout our lives. The words and actions of others can inspire us to overcome self-doubt just as easily as they can create it. This explanation of trust's role in the on-going creation and

[124] In her 2022 essay, Daniela Dover argues that we can "allow that others play an important role in self-construction" on both auto-constructivist and relational constructivist accounts of self-identity (Dover 2022, 219). I wish to follow her on this point and affirm the antecedent of the following claim she makes at the start of the same paper: "If what we want is to understand the full range of human experience, our philosophical theorizing should allow for multiple, parallel, non-competing ways of thinking about selfhood and subjectivity" (Dover 2022 195).

maintenance of our identity is, admittedly, still a bit sketchy. I will explore it further in the final chapter by looking at the relationship between self-trust, trauma, and identity.

Spelling out the ways in which trust is valuable allows us to fully appreciate that there is more at stake when we trust than when we rely. Though the extent to which this vulnerability is embraced can vary on the account I've put forward, it is different in kind from the vulnerability seen in cases of mere reliance because, when injured, it directly impacts our ability to lead a flourishing life. People who consistently have a reliance or distrust response to their situation (even if with good reason) miss out on the ethical goods of living in community and a stable sense of control over the formation of their identity.

The Developmental Account and Attending to Relationships with Asymmetries of Power

Because the developmental account of trust looks at the environment and shared history of the parties involved in any given instance of trust it can attend to the numerous factors that shape the power dynamics in each of our relationships. These variables are captured by the first part of the definition: trust is a response *to one's situation* that embraces vulnerability. The existence of other people is a significant and constant part of our shared situation, but so are our material conditions (which I take to include the fact that we share the world with non-human life forms) and our unique personal histories. The latter two components are less stable than the first, but all three of them bear significantly on whether we can trust our needs to be met. Unless, like Henry Bemis,[125] we find ourselves the sole survivor of a major disaster, there will always be some other people that we share the world with. Our material conditions can change at a moments' notice as well as through our own efforts. Similarly, our personal history is always

[125] This is a reference to an episode of The Twilight Zone titled "Time Enough at Last", where the main character, Henry Bemis, happens to be hiding in a large safe when an H-bomb is dropped. He emerges to find that he is the only person left within his hometown and become elated when he realizes that there is no one left to stop him from reading.

evolving. Though the past events in our personal histories themselves do not change, their significance and how we interpret them might. The developmental account argues that evaluating any given instance or climate of trust requires considering how these three variables interact. It will be easiest to see how this is done by appealing to a case where a sudden change in material conditions drastically restructures the power dynamics with a relatively contained network.

Ruben Östlund's film *Triangle of Sadness* (2022) provides such an example. At the end of the second part of the film, a luxury vacation yacht is blown up by pirates and begins to sink. The third part of the film opens with a shot of the motely array of the ship's surviving passengers who have all washed up on what appears to be an abandoned island. Abigail (a middle-aged woman and immigrant who was a cleaning lady on the cruise ship), quickly realizes that the other people on the island—Yaya and Carl (a young, attractive couple who got free tickets for the cruise because they are social media influencers), Dimitry and Jorma (two older male billionaires), Therese (the disabled wife of different billionaire whose husband has presumably perished in the attack),[126] Nelson (a younger, African man who worked in the ship's engine room) and Paula (the leader of the service staff who, while on the ship, indulged every whim of the wealthy guests while wielding her managerial power over the staff)—do not have any survival skills. When the group discovers that the lifeboat Abigail washed up in contains bottles of water and a limited supply of snacks, Paula demands that she distributes them to the others stranded on the island. Abigail is reluctant but concedes by handing over one of the boxes of chips and a tray of water bottles. As the rest of the group quickly chugs their water and ravenously consumes their snack-sized bags of chips, Abigail takes the opportunity to hide what limited rations remain.

[126] This character's disability is the result of stroke which, in addition to paralyzing her legs, left her with a limited vocabulary.

Later that evening, she catches an octopus. As she approaches the group, who are all cheering for her success, she asks if any of them know how to make a fire. They shake their heads. She asks if anyone knows how to clean or cook what she has caught. Their heads shake again. Abigail proceeds to prepare her catch, build a fire, and cook while the other members of the group watch in awe of her abilities. When Abigail finishes making their dinner, Paula attempts to take the final product from her to distribute among the group members. Abigail replies with a shake of her head and a *nun-uh,* and pulls the cooked octopus back toward her before distributing it, keeping one piece for herself for every piece she gives to another member of the group. Paula is outraged by Abigail's unequal distribution: *no no no,* she asks, *why do you get so much food?*

Abigail replies with a list of the labor she contributed to the meal and a defense of her decision: *I made the fire and I cooked. I did all of the work and everybody got something*. Paula tries to put Abigail in her place*: We work on a yacht. You are toilet manager. You don't know how to handle this*— Abigail interrupts her—*What yacht? Where is the yacht? On the yacht, toilet manager. Here, captain. Who am I?* Nelson is the first to submit to her authority. When Abigail repeats the question, he responds quickly with *You're the Captain*, and is rewarded by Abigail with another piece of octopus. The other members of the group follow suit (with varying degrees of enthusiasm) and, as Paula realizes that she is outnumbered, she is forced to publicly hand over authority to Abigail as the rest of the group silently internalizes the new hierarchy.

The rest of this part of the film invites viewers to consider how the power dynamics that result from intersecting factors like age, gender, ability, class can be altered by a change in external circumstances. Attending to the stark and subtle power asymmetries in the scenes that follow this one are telling—Abigail is the first to notice that drastic change in their material

conditions creates an opportunity for the power dynamics within the group to shift. She uses her skills to earn the respect of the other group members and takes further steps to reorient the power dynamics by playing on the existing trust and distrust among the group members. Abigail's superior abilities afford her power in this situation and this power is amplified by the resources she has hoarded (i.e. the snacks she'd prudentially tucked away shortly after they'd washed ashore). After they finish eating the dinner she has prepared for them, she invites the women to join her in the lifeboat (the only enclosed structure the group has) and asks the men if they will stay awake and keep watch over the fire. As the women turn in for the night, Abigail not only leaves the men in fear of what dangers lurk in the dark jungle emitting spooky noises, but also with her backpack which contains a package of pretzel sticks—some of the rations she hid earlier. When the other men fall asleep, Carl and Nelson peer into Abigail's bag and cannot help themselves—they eat the entire package.

The next morning, Abigail wakes up to discover that all of men fell asleep and let the fire burn out, and notices that the pretzels she'd left in her bag are gone. She immediately senses that Nelson and Carl are guilty. Nelson apologizes immediately, Carl attempts to defend himself, to the dismay of Yaya and the rest of the group. Abigail chastises the men for letting the fire go out and punishes Carl for lying by refusing to give him any of that night's dinner. However, later, she invites Carl into her lifeboat and offers him some more of the rations in exchange for sex. As these exchanges continue, they drive a wedge between Carl and Yaya's relationship.[127]

What is important about this example is what we can observe about how individuals' responses to their situation move between hope, trust, reliance, non-reliance, distrust, and

[127] It is relevant to our assessment of Abigail that his punishment of not receiving dinner only lasts for the one night. He continues to sleep with her in exchange for food (a fact he hides from everyone but Yaya). There is some coercion involved, but Carl could stop at any time and in doing so would give up only the extra rations he receives from Abigail, not food altogether.

despair. Abigail was initially distrusting of the group she was stranded with, and with good reason: she had experienced first-hand the messes they made and the waste that they produced. Her self-trust, in both her skills (swimming, catching fish, making fire, cooking) and her judgement of others' characters are clearly essential to her ability to quickly rise to the top of the group hierarchy. Her distrust of them is made evident by the fact that she feels the need to hide the remaining rations. The scene where Abigail ascends into the leadership role of the group is pivotal, and the most complex to classify. Abigail establishes momentary trust with the group through a contractual exchange: if they agree to let her be captain, they can trust her to provide food and fire. Whether or not she trusts or relies on them at this stage is somewhat unclear, in part because of the varying degrees of enthusiasm about her rise to power and in part because we are looking at a group, not a dyad. Since she is greatly out-numbered, occupied the bottom of the hierarchy prior to the yacht sinking, and they are all unsure of how long they will be on the island, her actions in this moment qualify as a limit case of trusting on the developmental account—she embraced some vulnerability, but for primarily prudential reasons. Abigail recognizes that despite the appearance of her position in the group being secured by her actions at dinner, she is nonetheless still in a precarious situation.

Her decision to leave her backpack behind and entrust the men with the fire could be interpreted as a response of trust of the therapeutic sort, as she seems to have left the backpack behind intentionally. Her decision to allow the women to sleep in the safety of the lifeboat is the most puzzling, partly because it is where the audience gets to see that Abigail is also socially savvy. Her action is clever: Abigail uses the patriarchal logic that shaped the social dynamic when they were on the yacht (i.e. women are delicate, must be protected) to secure an advantage for herself. However, the act also intentionally recreates a gender division, and it is unclear if her

reasons for doing so are to ensure the group's survival or if it is part of a larger plot to remain on top of the pyramid. The only way we could know for sure if we knew more about the kind of person that Abigail was prior to this experience—the limited knowledge we have of her personal history is a crucial variable that is unknown to the audience. Nonetheless Abigail's action appears to be guided by a form of trust and there is potential for a more trusting relationship between her and the others to develop by the end of the evening. At this point, the trust relationships begin to diverge again.

Carl and Nelson's betrayal quickly reverts her relationship with them back to one of distrust. Nelson's quick confession and apology appears to earn him some trust back, but the relationship between Nelson and Abigail is not explored much further in the film. As Abigail becomes more corrupt her interactions with Yaya and Carl represent the rotten trust relationships—ones where there is tenuous trust, secured by force or advantage (i.e. the kind that would fail Baier's "Moral Test for Trust").[128] Abigail treats Carl and Yaya like predictable objects rather than as free and autonomous agents. She secludes herself from the rest of the group most of the day, and seems content with the contractual form of trust-bordering-on-reliance arrangement that she has secured: the group's need for food is satisfied, she is respected for her hard work and abilities, and her position of power allows her to enjoy the luxuries of having sex with Carl and of having the lifeboat as her own private lodging. The group, except for Carl, who they all mock for sleeping with Abigail, appears satisfied with the arrangement. [129]

[128] See pages 16-17 in Chapter one.

[129] Some people prefer and enjoy being in a leadership role; some do not. Others still do not enjoy it but will begrudgingly take it on to avoid following someone who would be worse at it. Abigail seems to fall into the first category, and it is easy to imagine the power and respect afforded by a leadership role is appealing to someone like Abigail someone whose labor (which was not only taxing, but often demeaning) had been underappreciated for years.

The other interesting aspect of this example that the developmental account is well-equipped to explain is that although the relationships Abigail has with the group as a whole rests on the borderline of trust and reliance, her ability to ensure the that basic needs of the group are met allows a climate of trust to form. This climate of trust transforms the way the group members interact with one another. After the initial shock of the shipwreck and restructuring of the power hierarchy, the film offers several glimpses that highlight how this change to their material conditions, which levels power in terms of socioeconomic class while leaving other asymmetries (like age, gender, and disability) intact, clears the ground for the formation of new trusting relationships. In one particularly poignant shot, we see Dimitry and Nelson chatting as Nelson shaves the beard Dimitry has started to sprout with a knife. In addition to being an intimate act in the sense of physical proximity, the scene exemplifies how drastically even a small amount of trust can lead to a willingness to make oneself radically vulnerable. Dimitry sits perfectly still, smiling with his head tilted back as Nelson—a man he had accused of being a pirate and threatened to kill just a few days earlier—presses a blade against his throat.[130] In another vignette, Therese, who had been portrayed as burden on the yacht and was initially neglected when they all washed ashore, is seen laughing as she and Nelson play a board game they have created together. It is clear in the scene that the two found a way to converse and that the group has found ways to include Therese in their small community.

The very last scene of the movie confirms that material conditions we find ourselves in have a significant impact on whether we trust or distrust (stop here and skip to the next paragraph if you don't want me to spoil the ending!). Yaya wakes Abigail up early in the

[130] Nelson's accent and the color of his skin, combined with the fact that Dimitry "never saw him on the yacht" have Dimitry convinced at this earlier point in the film that Nelson is one of the pirates. Nelson is only spared from violence because Paula and the others are able to convince Dimitry of the truth: Dimitry hadn't seen Nelson on the yacht because Nelson worked in the engine room and therefore never interacted with the guests.

morning with news that she discovered something. Abigail follows her on a hike—which is difficult for her, but easy for the younger, more fit Yaya—through the jungle and over a mountain to the other shore of the island. To her dismay, she parts through the trees to find that what Yaya has discovered is a luxury spa resort. While Yaya's back is turned, Abigail picks up a large rock and begins to creep toward Yaya as she delivers the final lines of the movie: *Abigail, I can try and help you. I don't know how, but… Abigail, maybe you could come and work for me? You could be my assistant.* Whether Yaya's offer to Abigail, who is at this point standing behind Yaya, ready to smash her head in with a rock, is an example of a trusting overture or a cruel act of revenge depends on your interpretation of the film (which ends without letting us see what Abigail decides to do) and Yaya's character.

My account of trust is called developmental because it conceptualizes trust as dynamic rather than static. By emphasizing that trust always occurs in existing relationships and surrounded by contingent conditions, the developmental account does not let us forget that our past experiences as well as our environment bear on our ability to feel trust in and trusted by those around us. While the capacity to respond to one's situation with trust is something we share with others, the way that capacity is shaped is ongoing and deeply intertwined with our lived experience of our physical environment. Our capacity for self-trust—a capacity that is integral to every aspect of our decision making—depends on us first having experienced trust with another. It begins developing in the moment that we enter the world, but only because the moment we enter the world happens to be the same moment we are initiated into social and moral life.

This brings me to another affinity and difference that my account shares with Lagerspetz's and Simpson's. While both fare better with respect to the third requirement because

they acknowledge that trust admits of multiple kinds of value, their respective accounts each assume limitations for the objects of our trust that the developmental account avoids.

Simpson argues that trust should, under normal circumstances, follow the evidence.[131] He supports his position by investigating a subtle distinction between accepting and believing what another says. Though accepting someone's promise or testimony changes the normative relationship between the two parties, it does not necessarily entail that a belief that the person will act in accordance with the normative change that has occurred within the relationship. We can accept that a person says they will do X without actually believing that they will (Simpson 2018).[132] He attempts to resolve the challenge this distinction creates by appealing to the "folk" version of the problem. Because promising and telling are both commitment instantiating practices, he argues, they yield both practical and theoretical reasons. Thus, when the practices of promising and telling function properly, they create a route from accepting another's word to believing it (Simpson 2018, 456-457). Simpson argues that his analysis of this phenomenon shows that cognitivist, commitment-based accounts can accommodate rational and social intuitions about trust whereas affective accounts cannot.

While I agree with Simpson that the practices of telling and promising rely on the norm of truth-telling and can see why he argues that this reinforces the link between trust and commitments, he overlooks the way that these practices depend on the more basic form of trust that are centered on my developmental account. Promising, as we have seen, is a social practice that, like a contract, is a peripheral form of trust: it depends on the existence of other, more basic

[131] He does not consider himself to offer a doxastic account, in part because he thinks that there are abnormal cases of two-place trust where trust can be justified without evidence.

[132] This can also help us explain what is happening in the recovering addict case. Her friends and family may accept her promise to never drink again without actually believing that she will. This might be seen as rather pessimistic, but we do have to consider the past experiences of the person who holds this view. Perhaps they have had the unfortunate past of seeing many of their other loved ones repeatedly relapse.

forms of trust. Telling involves the basic form of trust that supports our language (broadly construed here to include gestures and other, non-verbal forms of communication).[133] The practice of telling is something that we do all the time as language-using creatures. We are introduced to these norms far before we learn what it means to make or have a commitment. In other words, Simpson's account fails to recognize that language itself depends on a basic form of trust—the expectation that what I say will be intelligible to you and vice versa. We develop this basic form of trust through our interpersonal relationships with others. As I argued in chapter two, the power asymmetry between parents and their children is due in part to the fact that parents are typically the ones who introduce their offspring to social norms and practices. In other words, they have substantial power in shaping their child's judgements about trust, including when to trust testimony, and what kinds of people can be trusted. Given that our ability to trust comes before our ability to testify and assess the testimony of others, I think that Simpson's account overlooks a crucial variable.

In the opening chapter of his book, Lagerspetz describes his investigation of trust as driven by two larger theoretical questions: *what is the philosophical significance of the fact that we share our lives with others?* and *What is the philosophical significance of seeing a difference between human beings and things?* I consider my own project to be driven by the first question and take it that trust is part of the answer. Furthermore, while I am sympathetic to the second question, I believe the way it is framed poses a false binary. Human beings are a class of beings that are (and should be) seen as morally significant in a way that this is distinct from the potential

[133] One question I am still working out is whether how I am using language here accounts for written language. Excluding it might require a defense of an account of human language that excludes its written form, which I don't think is likely to be compelling. However, it is worth noting that writing developed out of spoken language after a considerable amount of time. I think my answer or reply at this point is that written language is always mediated by its form and is (except, possibly in journals or diaries) written for a particular audience. These facts make it such that we are always aware (or at least it is always possible to be aware) of an author's intended audience when we read. Whether or not we default to trusting the author depends on the form as well as the genre.

moral significance of mere objects. However, humans encounter and interact with entities that do not fall neatly into the two categories of "human being" and "mere object". For one, there are morally significant distinctions among various kinds of objects. Objects, particularly those tied to specific places or particular people—can be more than "mere" objects by possessing relational properties that imbue them with value. There are buildings and works of art, as well as certain tools that can all reasonably described as something more than a "mere object". Consider the difference between the *actual* Declaration of Independence and a replica sold in a gift shop, a jersey that *belonged* to a famous quarterback and a jersey with the player's name and number worn by a fan at the game, and so on. By suggesting that the former are more than *mere* objects, I am not elevating their worth to the level of human beings, but rather introducing the point that there is a gradient or spectrum, not a binary. More importantly though, human beings also interact with other *living* beings like plants and animals. It is not uncommon, in ordinary language, to talk about gaining an animal's trust. In some cultures, it is common to talk this way about the trees or the land itself. Determining which other living beings can trust or be trusted is not a part of my argument in this project. However, I do want to keep open the possibility of trust in some non-human animals, trust in organizations and groups as well as "the world". Lagerspetz's question implies that human beings are the only entity capable of being trusted, and as a result he fails to see how the world, other living creatures, or a web of trusting relationships might be a fitting candidate.

Furthermore, contrary to Simpson, Lagerspetz insists that any weighing of variables or evidence automatically disqualifies a response from being trusting. Contrary to me, Lagerspetz also argues that we can't define trust in terms of vulnerability or risk because, from the first-person perspective, the possibility of betrayal never arises (if we are actually trusting). The risk

or vulnerability that is present is only apparent from the perspective of an outsider or third party. By emphasizing *only* the relationships involved, Lagerspetz fails to recognize the impact that our environment has on our judgements about trust. Though my account considers latency to be a paradigmatic feature of trust, it admits of exceptions in the limit cases between hope and trust and between trust and reliance. Lagerspetz's account lacks the dexterity we saw in Baier's account because it is inflexible on this rule which, if followed, would eliminate forms of trust that are important in our daily lives.

Given its emphasis on the embrace of vulnerability in trust, one challenge to meeting this requirement my account faces is explaining how it is possible for there to be trust in relationships of radical dependency. The objection might run like this, if it is true that all forms of trust are built up over time through interactions that meet our needs, then the trust between those who, given their extreme dependency, cannot meet needs but only have their needs met, and those in the position that affords them power to meet those needs can only be unidirectional. To clarify, I'll put this objection in the form of one of the examples of relationships with stark asymmetries of power that I considered in chapter two.

Of the three asymmetrical relationships I consider, the doctor-patient case is the most challenging because the asymmetries of power in it do not fluctuate over time the way that teacher-student and parent-child relationships do. Except in a very rare cases, doctor-patient relationships have a relatively stable asymmetry of power.[134] Teacher and student relationships are such that they can transform into relationships of intellectual peerhood over time. They may

[134] A case I have in mind can help illustrate a possible exception, as well as its rarity—let's say a doctor's patient is so inspired by their doctor's work that they decide to attend medical school, train under their doctor, and then, years later end up providing medical care for the person who was once their doctor. My point is, that unless there is a literal reversal in the positions or the teacher-student or parent-child dynamic overlaps with the doctor-patient relationship, the power asymmetry between a doctor and their patient does not generally fluctuate.

even encounter moments where the usual asymmetry in their level of expertise is reversed. Parent-child relationships have a slightly different developmental arc—as the child develops, the asymmetry of power and degree of vulnerability approach each other, only to diverge again as the parent ages and becomes more dependent. We do not see the same diachronic change in power dynamics in typical doctor-patient relationships. In doctor-patient relationships, patients are vulnerable to their doctors in more ways than one. In addition to the asymmetry in their respective knowledge of the body, medicines, and illnesses, there is also a difference in the degree to which their physical bodies are exposed and made vulnerable to each other. While the patient has good reason to want a doctor she can trust, the reverse does not necessarily follow. Put as a question: *Why would a doctor need to trust their patient and what possible difference could it make?*

When we think of trust only in terms of its prudential value, this objection carries more weight and there is a way to answer this question without leaving this perspective. Patient narratives are a source of evidence that can provide doctors with facts about the patient's illness that cannot be accessed through the doctor's methods.[135] When patients trust their doctors, they are generally more likely to tell their story, to tell it truthfully, and to be willing to provide additional details if requested. So, one could say that even if a doctor doesn't *need* their patient's trust, it is still useful to have and therefore worth cultivating.[136] Though I do not disagree, I think this such a response falls closer to the reliance side of the spectrum of forms of trust.

As soon as we appreciate the way trust forms a foundation for intimacy, the concern about implausibility of doctors trusting their patients expressed in the question begins to fade. Intimate relationships with others, particularly those that involve care labor, can satisfy both our

[135] See Misak (2010) for an argument in favor of counting patient narratives as evidence.
[136] See McNeally et al. (2004, 2009) for an account of why trust is valuable to surgeons.

need for connection and our need for creativity. Held argues that caring labor, "can be transformative rather than merely reproductive and repetitious" (Held 2006, 32). By working to cultivate trusting relationships with their patients, doctors not only increase the chances for improving the quality of the care outcome, but also can also experience the satisfaction that can accompany care work.[137] While this need pales in magnitude to the needs that doctors meet for their patients, there is still a reciprocal meeting of needs that can occur in mutually trusting doctor-patient relationships. The degree of dependency between the two parties does not need to be equal for the actions that meet such needs can be creative and fulfilling (i.e. enjoyable in their own right). Appreciating this fact is necessary for understanding why the trust seen in various forms of care labor requires trust from both parties, and not just the party in the more dependent position.

Avoiding Pitfalls

In the last chapter, I identified two pitfalls that doxastic and affective accounts tend to fall prey to. I suggested that they fall into these traps—the push to demoralize trust and the turn toward trustworthiness—because they rely too heavily on the three-part entrusting model. To conclude this chapter, I will explain how my developmental account of trust helps us avoid making these missteps.

Resisting the Push to Demoralize Trust

As I demonstrated in the last chapter, demoralizing trust leaves us with an incomplete, undesirable, and inaccurate picture of what trust is like in friendships, arguably one of our most important (or at least philosophically interesting) kinds of relationships. In Bennett's example, the trust we see in friendship is supported by the proleptic assumption that the trust we have in

[137] This is not to say that care work is *always* enjoyable or that it must be accompanied by the feeling that one is exercising their creative powers to be done well.

our friends is secured by our belief that our friend has a commitment to our friendship. Bennett is forced to ground the trust we see in friendships in this circular way because his call to demoralize trust requires jettisoning the MMP on the condition that it is neither a necessary nor a sufficient condition for trust. I agree that the MMP is not a sufficient condition, as it is possible to believe that someone bears us goodwill and refrain from trusting them. When this happens, it is usually because we know (or anticipate) something about the circumstances that they do not or because we have judged them to be well-meaning, but incompetent. However, once we abandon the project of searching for a set of necessary and sufficient conditions that characterize every instance of trust and instead ask what its paradigmatic features are, it becomes easier to challenge the argument that the goodwill requirement should be dropped.

It is hard to pinpoint what counts as an expectation of another's goodwill. On a more minimal definition, the expectation of goodwill might simply amount to the absence of an expectation of ill-will. A more robust conception might require bearing goodwill to another to include taking steps to support or increase the likelihood that things will go well for that person. My own position is that the degree of goodwill we owe to strangers requires something more than non-interference, which approximates beneficence. What is required to bear a non-stranger goodwill is indexed to the social role and particular relationship. What it means to bear goodwill to stranger is different from and demands less than what is required to bear goodwill to a close friend.

If we can bear goodwill to someone without ever having interacted with them, the number of cases where we can drop the expectation of goodwill are much less common than those pushing for us to demoralize trust suggest. *Not* expecting goodwill either implies that one cannot rule out the possibility of ill-will, or that one expects to be met by others with what P.F.

Strawson has called the objective attitude. Being regarded through the objective attitude can, like being distrusted, be experienced as a slight or an insult, as it suggests that one is not fit for membership in the moral community. [138] I take this similarity to show both that we are concerned with the quality of another's will in most of our interactions with others and that we cannot have normal interpersonal relationships with someone who always responds to their situation with reliance.

The expectation of goodwill, when defined minimally, seems to capture most cases of trust, except for a few outliers. A closer look at these outliers reveals that they tend to fall into two categories—rotten cases of trust (i.e. trust relationships that would fail Baier's moral test) and instances of supererogatory trust. The former tend to be relationships that are abusive, manipulative, or exploitative. The latter tend to be instances where the risk involved is voluntarily and consciously embraced. As D'Cruz points out these "trusting overtures" bring out the relationship between trust and courage (D'Cruz, forthcoming). One has to be courageous if they are going to extend trust without expecting that the person they are trusting will bear them goodwill.

When we reintroduce goodwill as a paradigmatic feature of trust on the grounds that it only missing in the two kinds of cases mentioned above, it becomes possible to say make the intuitive, if circular, claim that we trust our friends because they are our friends, so long as our friendship emerged out of and is sustained by reciprocal goodwill. When we eliminate goodwill from our understanding of what it means to trust and be trusted, we lose sight of trust's foundational role in the formation of our intimate relationships. When we prioritize particular

[138] I'd like to thank the members of the 2023-24 Ethcetera reading group for the challenging and intriguing conversations we have had on this text and Hieronymi's interpretation of it in *Freedom, Resentment & the Metaphysics of Morals.*

relationships and social and historical context, we get a clearer understanding of how trust and distrust function. The push to demoralize trust takes us further away from attending to trust in this way.

Re-turning to Trustworthiness

Emphasizing the reciprocity and responsiveness inherent to trust also helps us avoid the second pitfall—the turn toward trustworthiness. While the turn toward trustworthiness makes some valuable contributions toward investigating the importance of trust in our relationships, on affective and doxastic accounts it does not make the relationships themselves its focus. If we take our particular relationships to be morally significant, any discussion of trustworthiness must also consider its complementary virtue—trustfulness. Since many of our trust relationships are ones where the roles of one-trusting and one-trusted alternate over time, testing the health of the trust in them requires considering the extent to which each party embodies not just one, but both of these virtues.

Um's account of trustfulness brings us back to Baier's argument that trusting someone entails giving them discretionary power, and in doing so, risking exposure and making ourselves vulnerable. She argues that the vulnerability in trusting has "two apparent dimensions, renunciation of guard or defense and renunciation of intelligence, [which] do really seem to be two neither irreducible to the other" (Baier 1994, 158). Um's account of trustfulness, which describes it as part epistemic skill and part moral attitude, can help us make sense of these two dimensions. When we judge someone to be competent, to have the knowledge necessary to act in a way that would meet our trust, we forfeit our right to know the means by which they will satisfy our trust. This is what we saw exemplified in the Agent F and D case. When we treat someone as a free agent we implicitly recognize and respect their right to make their own

choices. This recognition involves acknowledging (again, at least implicitly) that they could at any moment choose to *not* act in a way that meets your trust, despite the evidence you have that makes you willing to renounce your right to know. It also requires one to let down their guard. In trusting them you are accepting that they *could* wound you if they chose to. When coupled together they describe the distinct vulnerability that trust exposes us to.

Not only is signaling trustworthiness difficult (particularly so in a climate of distrust), but it is not the only trait that can justify trust. Trustworthiness is a virtue that is most relevant in the context of testimony and promise-making—when what matters to us most is if we can rely on the person to speak honestly or truthfully and follow through on the things they say they will do. However, in many cases, how honest a person is does not bear on our judgement about whether or not we trust them. We may not believe the babysitter is being perfectly honest when she says that the kids were perfect angels but will still trust her nonetheless because we take her to be a patient, caring, and competent person. These enduring character traits, not the degree to which she can be relied upon to tell the whole truth and nothing but the truth, are what *really* ground our trust in her. The turn to trustworthiness is in fact a path worth pursuing because in our daily interactions we need to trust more than just what others say. [139] However, it only avoids being a pitfall when we have an account of trust with the conceptual resources to resist the push to strip trust and trustworthiness of their moral significance.

Conclusion:

[139] There is an interesting relationship here that is worth exploring further. We might start to distrust the babysitter if we found out she was regularly dishonest, especially if she started being dishonest about serious matters pertaining to the well-being of the children she is supposed to be caring for. Finding out she did some other egregiously dishonest action will affect our perception of how trustworthy she is and may even spillover and affect our perception of how patient and kind she is. As I showed in an example last chapter, it is possible to trust in spite of being aware that one is being lied to or misled. The relationship between honesty and trust is complex and exploring exactly how complex is a task I plan to take up in my future research.

The value of trust comes from more than its ability to facilitate cooperation. Though this aspect of its value is itself ethically significant, fixating on it alone tends to allow the epistemic aspect of trust to cloud its ethical or moral aspect. Trust, as bell hooks argues, is the foundation of intimacy. As a foundation to something we value intrinsically, the necessity of trust for a flourishing human life is paramount. If we take away the foundation—through gradual erosion or a sudden shattering—our ability to have meaningful relationships with others vanishes along with it. This very basic role that trust plays in our relationships also explains why it often goes unnoticed. We do not typically inspect the soundness of the foundation of our homes unless there is an indication that something has gone wrong with them.[140] Stretching this metaphor forces us to recognize just how necessary trust is for not only the establishment of meaningful relationships, but also the maintenance of them. When the foundation of a house is disturbed, it affects every level of the house built on top of it. When the trust in a relationship is betrayed or eroded, it can have the downstream effect of calling one's own ability to make future judgements about who is trustworthy into question.

The developmental account is able to explain this effect because it appreciates that a climate of trust involves both trusting relationships and the surrounding environment in which such relationships develop. On the developmental account, trust is seen as not only conducive to cooperation and the expanding of our individual agency, which assists us in securing resources and makes it possible for us to achieve collectively more than we ever could alone, but also as a constitutive part of our intimate relationships. Furthermore, it attends to the fact that intimate relationships with others inform and shape our sense of who we are. In this way, my account

[140] I am not saying this is wise—one can avoid a costly and stressful repair if they catch sinking and cracked foundations early on and perform the proper maintenance. The same seems to be true in our relationships—if left unaddressed small betrayals can develop into larger breaks in trust and the accumulation of lies can erode what was once a stable foundation of trust.

builds on the work of care ethicists, including Baier as well as we Held and Cavarero, who all emphasize our interdependencies and their relation to the formation of our identity.[141] My account takes as its starting point that we all begin our life in a state of dependency, and that dependency relations are matters of trust.

Chapter Five: Trusting Yourself and Trusting the World

"Trust is only intelligible against the background of some developmental story."

— J.M. Bernstein

In the last chapter, I argued that part of trust's value comes from the enabling role it plays in the formation of our identities. This enabling role is related to trust's role as the foundation of intimate interpersonal relationships. To further understand how trust is implicated in the formation of our identities, we need to take a closer look at two forms of trust that have been largely set aside by the prominent theories discussed in chapters two and three: self-trust and trust in the world. In this chapter, I will show that my developmental account can make sense of these two phenomena. This ability gives my account an additional (and substantial) advantage over alternative accounts. I argue that this advantage is substantial because, in addition to answering the question I drew out of Mill's work (*Why are failures of trust among the greatest wounds human beings can sustain?*), a deeper understanding of self-trust and trust in the world also illuminates how we can address the practical problem of distrust.

[141] In *The Ethics of Care* Held writes, "the ethics of care values the ties we have with particular other persons and the actual relationships that partly constitute our identity. Although persons often may and should reshape their relations with others—distancing themselves from some persons and groups and developing or strengthening ties with others—the autonomy sought within the ethics of care is a capacity to reshape and cultivate new relations, not to ever more closely resemble the unencumbered abstract rational self of liberal political and moral theories (Held 2006, 14).

I begin this chapter by elaborating on trust's role in the formation of our personal identity. This elaboration requires a description of how self-trust and trust in the world typically develop. On my account, self-trust and trust in the world both arise out of a more basic form of trust—recognition trust, a concept introduced by Johnny Brennan in his 2021 paper of the same name.[142] In the next section, I show that my account of trust can also fruitfully explore atypical developments of self-trust and trust in the world. A closer look and attempt to categorize these atypical developments, I argue, provides us with a helpful framework for remedying distrust. In the concluding section, I describe the paths forward that my account of trust illuminates.

Before I proceed with the plan above, I will elaborate on what I mean by "typical". Something that is typical is "normal" with respect to a particular kind of thing. We often use the word "typical" to comment on the habitual or characteristic features of individuals or institutions we frequently interact with. In this context, I am using "typical" to describe the standard progression of human biological, psychological, and social development. Though "normal" means more or less the same thing in this context, its counterpart, "abnormal", carries with it the stigma that is associated with deviations from the norm. The atypical developments of self-trust and trust in the world are not "wrong" or "bad" in the ways that the term "abnormal" suggests. An atypical development of self-trust or trust in the world can however be maladaptive as they often impinge on our ability to live a flourishing life. Situating self-trust and trust in the world within the context of their typical course of development enables us to avoid talking about atypical cases in a way that would disparage those whose self-trust and trust in the world departs from the typical course of development. It also encourages us to use a richer vocabulary for

[142] Danielle Petherbridge (2021) also explores the relationship between trust, recognition, and vulnerability in the work of Habermas. I found her argument helpful for understanding how trust and language are related.

describing the variety among atypical cases. For instance, instead of saying that a person's self-trust is abnormal, we can describe its development as interrupted, re-routed, or delayed. Focusing on the typical course of development also allows us to talk more clearly about the differences among causes that affect our self-trust and trust in the world. The impact that aging can have on our self-trust and trust in the world, for example, is included in the typical course of development in the way that the experience of surviving a natural disaster, severe injury, or trauma is not. These differences matter a lot once we start thinking about what practices are required to remedy or bolster these forms of trust. From this variety of atypical developments, we can expect the remedies will admit of a similar variation.

(Typical) Development of Self-Trust and Trust in the World

On my account, self-trust and trust in the world both arise out of recognition trust, a basic form of trust. Brennan, who introduces the concept, defines it as follows: "it is normative, it is necessary for living a fully human life, it is interwoven with human development, it is a construal of others, and it is latent" (Brennan 2021, 3808). Brennan suggests that recognition trust is first established through our interactions with our earliest caregivers. Most scholars, Brennan observes, explore the importance of trust in our development by looking at how distrust hampers it. I discuss cases like these in the next section. To continue looking at the typical case, we can turn to Bernstein's description of how trust develops for some clarity:

> My hypothesis is that the standpoint of trust is the incorporation by the child of her being seen through the eyes of love as worthy and valuable, where the anticipation of being so valued becomes her being seen and counting for others as a self-determining, vulnerable being who can and will respond to others as self-determining, vulnerable beings too (Bernstein 2011, 410). [143]

[143] This quote immediately proceeds the one in chapter four where Bernstein argues that the development of a capacity to trust is a requirement for becoming a person.

Bernstein goes on to explain how the formation of our capacity for trust is built up through mimetic and call and response interactions. I agree with Bernstein's reconstruction. If a human survives through infancy, it will be because they were cared for in an environment capable of meeting their basic needs. Such an environment will at least include minimally responsive and minimally competent caregivers. The quality of care the infant experiences and how well the infant's environment is equipped to meet their needs will vary, but the fact remains that the embrace of vulnerability that is at the core of trust is something we typically become accustomed to through our very first interactions with others in the world.

In the moment when we start expecting that others will recognize us, we also become capable of recognizing others. At this stage, it becomes possible for self-trust and trust in the world to sprout. Learning to control and coordinate the movement of one's body (e.g. learning to walk, wave, and smile) is something that it is easier for infants to do when they have the encouragement and instruction of others. Similarly, the development of an infant's memory is aided by those around them with more developed memories. Our memory and our control over our body are two things that become the object of our self-trust and their growth is nurtured by an ability to trust in the world. In a typical case, one's self-trust and trust in the world are provisionally established over the first few years of a person's life. These two forms of trust are also what enable the formation of our identity.

For the purposes of this chapter, it is not necessary for me to review the long history of philosophical work on the topic of personal identity. Instead, I will focus on two concepts that feature prominently in many theories of selfhood and identity—memory and agency—and explain how my developmental account shows that both self-trust and trust in the world are

necessary for the proper functioning of these two capacities.[144] Whether one holds a more constructivist or more relational view of identity is of little consequence at this stage in my argument. So long as one's preferred account of identity acknowledges that our ability to remember and our ability to act factor into the process of identity formation, it can be inserted here.[145]

Our capacity for memory allows us to link our past experiences and actions to our present selves. It makes it possible for us to develop and refine our skills as well as retain knowledge. Without the power of memory, it would be impossible for us to form a stable sense of who we are. Lacking memories of our past actions (or in-actions), particularly if they had significant consequences, is a cause for concern. In the practice of medicine, sustained inability to recall basic details about oneself and recent experiences is a sign that a patient lacks decision-making capacity. Furthermore, our memories are also necessary for enduring relationships with family and friends. The various relationships we have and social roles that we occupy require us to remember not just who we are, but also who others are as well as the history of our relationship with them. Our memories can only aid in the formation of our identities if we are able to place trust in them. Even though our memories are imperfect, most of us retain the ability to trust our memories so long as their imperfections are within a typical range. The inability to form, recall, and draw on our memories can hamper our ability to act. This brings me to agency.

---

[144] Aristotle, Descartes, Locke, and Hume are a few historical figures that immediately come to mind for me.

[145] That being said, I personally am partial to the understandings of identity formation offered by Christine Korsgaard, Ruth Chang, and Adriana Cavarero. Korsgaard's influential book *Self-Constitution* pulls together insights from Plato, Aristotle, and Kant to argue that we constitute ourselves—our practical identities—through our actions the maxims behind them. Chang (2002, 2012, 2017) makes a similar argument: when we choose, we put our agency behind our choice and thereby create ourselves. For Chang, what makes some choices hard is precisely this connection between our actions and our identity. Cavarero's understanding of selfhood is less individualistic than Korsgaard's or Chang's. While it is less focused on our individual actions, it emphasizes the role of our memory and the desire (which she takes to be universal) to hear our life's story narrated (Cavarero 1997). Despite their differences, each of these three understandings of identity and how it is formed converge on the thought that there is an intimate connection between who we are, what we do, and what we remember.

All I mean by agency is the ability to be the authors of our actions. We are free to choose among available options; however, the range of actions available to us is constrained by our physical and cognitive abilities, which are limited by our bodies as well as our contingent historical and social circumstances. Self-trust and trust in the world can help us explain these constraints in a way that appreciates the reality of these limitations without undermining the power of individuals. The formation of our identities requires trust because our degree of self-trust and the extent to which we trust the world bear directly on the range of possible actions available to us and, therefore, also to the choices that we can freely make.

Now that I've explained how our identity is forged by our capacities for memory and agency, I can now further explain how self-trust and trust in the world play an enabling role. Self-trust has a meta or self-referential quality that makes it a puzzling phenomenon. Unlike other trust relations where our judgements about trust refer to someone or something external to us, our judgements about when we can trust our memory and other capacities refer to something that is within or a part of us.[146] Self-trust is a form of trust that is typically unconscious and unchosen. When we recall memories, make judgements, and use our bodies to execute voluntary actions, our self-trust is operative, but latent. When we do not or cannot place trust in our memories, it becomes difficult to complete even routine, everyday activities. We can see this in the lives of those with enduring and chronic memory difficulties, such as those with schizophrenia or Alzheimer's disease, as well as through those who experience temporary memory impairments brought about by injury or induced chemically. Our ability to retain and

[146] I would like to thank my dissertation committee, especially Matthew Congdon, for raising questions that pressed me to clarify this point. The fact that self-trust is different from other forms of trust, particularly interpersonal forms, initially struck me as obvious. However, while it seems obviously different from other kinds of trust, the difficulty of articulating what precisely is distinctive about it has led me to question this assumption. Something I plan to explore in my future work is how we might make sense of what self-trust entails by comparing it to trust in friends that function as "second selves".

recall information is essential for engaging in activities, developing skills, and pursuing interests or projects that further shape who we are.

The other possible objects of our self-trust, such as our eyesight, hearing, and the use of our hands and feet, demonstrate other ways that self-trust is typically latent. The incredible feats of dancers and gymnasts vividly illustrate the upper limits of what strong, but still latent self-trust can help us accomplish. The effortless appearance of a backflip is supported by self-trust in several parts of the self at once—trust in one's balance, trust in one's feet, trust in one's judgement of distance, etc. While learning these moves takes conscious effort at the start, it requires confidence in one's "muscle memory" to pull off. The same is true of other skills that we refine through practice. Strengthening or adjusting our self-trust can sometimes require reflection. The ability to reflect on our self-trust points to the fact that, once established, our self-trust can be bootstrapped. Convincing yourself that you have what it takes to get through, endure, or accomplish a challenging feat is often a crucial first step in making that accomplishment possible. Though we can temporarily suspend our self-trust in our judgements, memory, and physical abilities for the sake of reflection and revision, we cannot remain in this state of suspension and continue to operate as usual. [147]

Trust in the world is the other side of the coin; it refers to the trust we have in our environment (which includes other people and their behavior) to remain predictable in basic ways. Like self-trust, it is typically latent; it involves expectations that many of us take for granted (e.g. that the ground will not suddenly crumble beneath our feet and that we do not need to fear that strangers or acquaintances will intentionally inflict harm upon us). The phrase "trust

[147] To clarify, I do not mean to suggest that the experience of self-doubt or self-distrust is a statistically uncommon one. Self-doubt and self-distrust can be normal or healthy when we are confronted with a new experience, unfamiliar way of life, or a hard choice. However, the suspension of self-trust gets pathologized if it is unable to resolve itself or resolves itself poorly.

in the world" was first coined by Jean Améry in his autobiographical work *At the Mind's Limits: Contemplations by a survivor on Auschwitz and its Realities* to describe the lasting impact of his torture at the hands of Nazi S.S. soldiers. It has since been used by other trauma survivors to describe how their trauma shattered what has been called "social trust", "general trust", "basal trust", and "basic trust" (or some combination of these and similar qualifiers) by scholars working in a variety of disciplines.[148] For the purposes of this chapter, I take them all to refer to roughly the same thing: the trust we have in our surroundings, which includes the existence of others. I find the phrase "trust in the world" more fitting than these other terms for a few reasons. First, the phrase has its origins in the trauma studies literature that I draw on in the next section. Second, it invokes the presence of a reality that is external to the agent doing the trusting. Third, it does not limit that external reality to just the existence of other human beings; it is expansive and therefore conducive to attending to other life forms and the features of our environment. Our trust in the world is susceptible to changes in our environment—if resources become scarce or conditions become dangerous, we may lower our baseline degree of trust in the world and become more skeptical of others.

As we saw in my description of how they begin to form, self-trust and trust in the world are linked in such a way that one often has a reciprocal effect on the other. If the ground we'd previously trusted to support us were to suddenly crumble beneath us, we would not only adjust our trust in our expectations of our surrounding environment, but would also adjust the degree of self-trust we have in our senses and their ability to accurately inform our judgements about what surfaces are and are not able to support our weight. Lacking self-trust or trust in the world hinders our ability to make choices and thus has an adverse impact on our ability to form an

[148] For a helpful overview of how these various categories are defined and used in political science, psychology, and sociology, see chapters one and two of Kevin Vallier's *Trust in a Polarized Age* (2021).

identity. Having self-trust and trust in the world enables our ability to make choices and thus makes it possible for us to form an identity.

Our self-trust and trust in the world are continually shaped by our life experiences and can be refined through self-reflection. Illness, injury, rejection, and betrayal, and consciousness-altering experiences can all weaken our self -trust and trust in the world. When our expectations of our own abilities and of our surrounding environment are stable, self-trust and trust in the world fade into the background, becoming latent like the recognition trust they emerged out of. In the typical course of development, self-trust and trust in the world resemble deeply engrained habits, akin to breathing. Like breathing, trusting ourselves and trusting the world are things we can do both with and without conscious effort. When we make them the object of our conscious attention we can refine and strengthen them, as well as muddy or weaken them. Barring an atypical occurrence in the path of one's development, the conscious attention we give to our self-trust and trust in the world will eventually return back to their latent, unnoticed state.

Just as they were at the start of our lives, our self-trust and trust in the world remain vulnerable to others at all stages of their development. Self-trust can be fostered or improved through therapeutic and educational, as well as mutually loving relationships. Baier's anecdote about learning to drive[149] as an adult and my earlier example of infants learning to walk are both cases where we can see how one person's trust can transfer to another. The trusting relationships we have with others are made possible by our sense of trust in the world and these relationships can in turn boost our self-trust. While there is bound to be variation in the extent to which an individual's self-trust is receptive to the encouragement or discouragement of others (not to mention different levels of receptively at within difference stages of their development), it

[149] See chapter one, page 33, footnote 34.

remains the case that our self-trust and trust in the world are linked by their susceptibility to being helped or harmed, supported or undermined, by the trust others have (or pretend to have) in us. The shared origin and structural similarity of self-trust and trust in the world suggest that the two remain deeply intertwined over the course of our lives.

As we age, most of us will find ourselves revising or reassessing the self-trust we have in our memories and in our ability to control various parts of our bodies. Though some are fortunate enough to reach the end of life with their cognitive abilities intact, all our bodies age and eventually begin to deteriorate in some ways. Part of aging, and indeed a difficult part for many, is coming to terms with the fact that range of things we could once trust ourselves to do independently begins to shrink.[150] For many it takes a serious injury or similarly transformative event for this realization to occur. Many people resist accepting this even after enduring such an event. How we are treated by others during this stage of life can also impact our sense of trust in the world. While interesting in their own way, the gradual narrowing or wearing away of our self-trust is typical and thus should be responded to differently than an atypical erosion or shattering of self-trust that we will explore in the next section.

To summarize, the typical development of self-trust and trust in the world begins with repeated moments of recognition trust, which occur during our very first interactions with those who care for us during the earliest developmental stage of our lives, which is characterized by total dependency. Once established, our continued interactions with others, the habits we

[150] Atul Gawande's book *Being Mortal* offers compelling narratives of his patients, friends, and family members different ways of responding to aging and dying. One particularly poignant story is of Alice, who managed to live independently until her declining memory led to her being victimized by scammers. Two men she had previously paid for a service repeatedly returned to her house and insisted that she had not paid them. She was too embarrassed to admit she couldn't remember, so continued giving them money until one of her family members noticed that a large sum was missing from her bank account. Her story is an illustration of how self-trust and trust in the world can be reciprocally impacted even at this stage, as well as of how something other than a physical injury can prompt one to revise their self-trust.

cultivate, the stability of our environment, and our capacity for reflective thought, all help to expand and render latent our self-trust and trust in the world as we become less dependent. As we age, we return to a state of dependency and the number of things we can trust ourselves to do without assistance begins to contract. It is crucial that we recognize this as the typical trajectory so that we can fully appreciate the impact that significant deviations from it have on our ability to form a personal identity and maintain meaningful interpersonal relationships. In the next section, I draw on findings from the trauma studies literature to identify and classify these atypical paths of development.

Atypical Developments of Self-Trust and Trust in the World

So far, I have shown that my developmental account can make sense of the phenomena of self-trust and trust in the world by explaining how both enable the formation of our identity. When self-trust and trust in the world follow the typical path of development, they become latent and make it possible for us to trust our memory, our senses, and our physical abilities, as well as our surrounding environment and the people within it. Being able to trust yourself and being able to trust the world enables us to make choices that both determine and reflect who we are. Those who lack self-trust and trust in the world will also lack the ability to make choices in this way. Long spells of self-doubt and self-distrust will make it difficult for a person to maintain a stable personal identity. Because we must eventually act, those who lack self-trust will defer to others' judgements about what they do or allow other external forces decide for them.[151] A loss of trust in the world will similarly make it difficult, if not impossible, for individuals to relate to others in the ways that meaningful relationships require. In this way, the absence of either (or both) presents an obstacle to living a flourishing, fully human life.

[151] Chang (2002, 2012) describes people like this as "drifters" in her work on parity and hard choices.

Atypical developments of self-trust and trust in the world thus present a problem that is analogous to the practical problem of distrust at the interpersonal and broader social level. Just as distrust between individuals or groups creates obstacles to communication and effective cooperation, a loss of trust in the world and/or a lack of self-trust poses a challenge to forming a personal identity and being able to engage with others. A deeper understanding of how and why development of self-trust and trust go awry can further illuminate how these problems are related.

One way to identify significant deviations from the typical path of development is to consider their possible causes. Brennan provides one answer: moral injury, which he describes as involving three tiers of harm:

> There is the initial harm of the violence itself; there is the secondary harm of a disabused assumption about one's safety in the world, including the harm of a lost connection to others; and there is the tertiary harm of *a lost sense of self, not in control of one's identity (or at times, of one's very body),* riddled with guilt, shame, and anger for not stopping what one felt incapable of overcoming. Moral injuries degrade the victim's sense of her own value. (Brennan 2021, 3804, emphasis my own)

As we can see from Brennan's breakdown, by damaging recognition trust, moral injuries have a lasting effect both our sense of trust in the world (i.e. the disabused assumption of one's safety and lost connection to others) and our self-trust (i.e. not in control of one's identity and body). The lingering impact of moral injuries thus has effects on one's ability to respond to others and on one's ability to relate to oneself.

Brennan acknowledges that moral injuries admit of a wide variety. He compares moral injuries that occur from sudden, discrete traumatic experiences (such as rape and torture), moral injuries that follow from repeated or accumulated experiences of misrecognition, and moral injuries that are imposed by the structural conditions of one's environment. He also observes that

moral injuries vary in their degree of severity. Recognizing that a loss of trust in the world and a lack of self-trust can follow from multiple kinds of events particular to the individual's personal history should guide the way we approach attempts to establish or repair trust. Improving the trustworthiness of a person, group, or institution that is distrusted by someone (or some group) whose distrust stems from the lasting effects of a moral injury will fail to move one to trust unless the actual cause is addressed. It also becomes evident that dismissing distrust that is not tied to evidence of untrustworthiness as "irrational" may induce additional harm to those whose reasons for distrust stem from an experience of moral injury.[152] Restoring trust will require addressing the violence that incited the moral injury, but the secondary and tertiary harms must also be addressed before trust can be expected. We cannot expect to establish or repair trust until the party that we want to gain the trust of feels safe, recognized, and in control of their identity. How to best repair or heal the secondary and tertiary levels of harm will vary with respect to the form of the moral injury (such as the three Brennan identifies).[153] Similarly, the time required for recovery will also look different based on the larger trajectory or shape of a person's life.

Brennan's analysis of moral injury identifies two key variables—the nature of the moral injury and the degree of severity of harm it caused—that are relevant for understanding how the typical development of self-trust and trust in the world can be disrupted. However, his helpful

[152] Of course, the fact that the perpetrator committed an act that constituted a moral injury is evidence for the victim of the act of such violence that the perpetrator is untrustworthy (among other things). My point here is that the lingering effects of moral injury may make distrust of relevantly similar parties (who themselves were not the original perpetrator) justified. This is also not to say that there are no cases where distrust that does not track the evidence could be dismissed as irrational, but rather that what counts as a "rational" reason to trust, or distrust should not be tied only to evidence about trustworthiness. For further discussion of this point see Jones (2004).
[153] For instance, for Susan Brison, recovery from her traumatic rape and attempted murder did not follow immediately from the arrest and prosecution of her assailant. Regaining the ability to connect with others, a feeling of safety, and a sense of control over her identity was only possible through a process of narratively reconstructing herself. She describes the process of narratively reconstructing herself in *Aftermath: Violence and the Remaking of a Self*. (2002). Another attempt to regain control of one's identity can also be seen through Jean Améry's decision to adopt "Jean Améry" as his new name after escaping Nazi persecution in 1945.

analysis does not address another key variable: the developmental stage at which the moral injury occurs. To understand this variable, we must venture into interdisciplinary territory. Because my developmental account parallels the biological, psychological, and social development of human beings, it is at home in this territory, which remains foreign to the alternative accounts of trust discussed earlier.

Bessel van der Kolk's research on trauma and its impact on the brain, mind, and body provides further insight about how the developmental stage at which the moral injury occurs affects the kind impact it has on the victim. Unlike those whose trust in the world and self-trust is shattered or eroded later in their adult lives, those who experience trauma in the form of abuse and/or neglect in childhood grow up without having the feeling of being able to trust their surroundings and those providing care to them. As a result, self-trust and trust in the world are prevented from fully developing. This stunting of the development of self-trust and trust in the world further inhibits their ability to form secure attachments, even with those who do not abuse or neglect them (van der Kolk 2014, 152, 160-165). Children who grow up with abusive caregivers find ways to cope with these experiences. These coping mechanisms often involve the repression or misremembering of their traumatic experiences. Van der Kolk reports that "erasing awareness and cultivating denial are often essential to survival, but the price is that you lose track of who you are, of what you are feeling, and of what and whom you can trust" (van der Kolk 2014, 137). These early disruptions in the development of self-trust and trust in the world leave survivors of childhood trauma without reliable heuristics about who to trust. According to van der Kolk, the fact that they do not learn these norms partly explains why people with a history of childhood trauma often find themselves in abusive relationships later in life. The task of healing self-trust and trust in the world in these cases is going to look different from the kinds

of cases that are Brennan's primary focus. There may be some overlap with respect to the concrete practices involved, however, creating the space for self-trust and trust in the world to completely mature for the first time is a different task than re-establishing these forms of trust in a person who once possessed them, but then had them shattered or chipped away by an experience later in life. It will also look different from the adjustment, coping, or grieving that accompanies the typical narrowing of self-trust that occurs as we age. The same is true of our interpersonal relationships. Coming to trust someone for the first time feels different than coming to trust someone again after they've betrayed your trust. I find it promising that establishing, maintaining, and repairing trust are distinct at each of these various levels (within the self, interpersonally, socially).

Research in trauma studies shows us that the nature the of the trauma is yet another variable to consider. Up until this point, I have discussed forms of trauma and moral injury that are the result of an intentional act of violence inflicted by one agent onto another. However, trauma can also result from living in or through dangerous conditions. Survivors of natural disasters and those who live in war zones or other areas where resources are scarce often persist in feeling a lack of safety or security even after the event has passed or their circumstances have changed.[154] Though the experience of living through temporarily hazardous conditions (like a tornado or hurricane) may not qualify as "moral injury" in the sense that Brennan describes, the symptoms of survivors and the impact on their sense of trust in the world can be similar.[155] This

[154] Vanderbilt University Medical Center's ethics grand rounds panel in April 2024 discussed the kind of moral injury that doctors, nurses, and healthcare service staff experience in their workplace and potential ways of addressing it. The focus of the panelists' was on the moral injuries experienced by hospital staff during the early days of the Covid-19 pandemic. These presentations by Sunil K. Geevarghese, Jason Nieuwsma, Sherry Perry, and Jessica B. Williams, brought yet another class of moral injury to my attention.

[155] The increased frequency of natural disasters caused by climate change presents us with a unique (in the sense of being unprecedented) case. On the one hand, the natural disaster itself cannot have a will, intention, or agency attributed to it. However, given our collective knowledge about the links between climate change, its harmful impacts on the environment, and the increased frequency of natural disasters, entities that contribute

resemblance in the impact on survivors' trust in the world affirms my earlier argument that in order to fully understand our tendencies to trust and distrust we must acknowledge that they are sensitive to our external environment, understood not just as the presence of other human beings, but also our material conditions.

Like the other variables, the condition of the external environment is also something we must appreciate within the context of restoring self-trust and trust in the world. My developmental account allows us to extrapolate the framework of restoring individuals' self-trust and trust in the world onto the practical problem of distrust in broader social and political contexts. In this way, it points to a new path we can explore when trying to address these issues. Just as one must know about individual's past experiences in order to determine whether their distrust is justified or warranted, one must also know the history of the relationship between the community and institution in order to issue accurate judgements about the status of the (dis)trust between them. Neglecting the history of relations between various social groups, institutions, and governments is likely to result in missteps in attempts to foster or repair trust between these groups. Though the missteps may appear innocuous or even well-intentioned, they can have disastrous effects, including further entrenchment of the very distrust such actions had aimed to reduce or eliminate.[156]

---

disproportionately to climate change are arguably blameworthy or responsible for them. Increasingly, survivors of these disasters and people who have incurred harms from living in adversely impacted areas have identified companies and countries driving climate change (and those occupying leadership roles within them) as the perpetrators of violence. Awareness of the indifferent and callous attitude these alleged perpetrators have can make the experience of such events feel dehumanizing. This additional dimension would thus qualify them as a kind of moral injury by Brennan's lights. In a recent class action lawsuit against the state of Montana, the courts ruled in favor of sixteen individuals (aged between five and twenty-two) who charged state agencies with violating their right to a clean environment for approving fossil fuel development without considering its impact on climate change (NPR Associated Press, "Judge Sides with Young Activists in First-of-Its-Kind Climate Change Trial in Montana."). The fictional film *How to Blow Up a Pipeline*, based on Andreas Malm's book of the same name, provides another example of this kind. Though it is fictional, the characters have plausible backstories that represent a diverse array of reasons that might motivate victims of climate injustice to engage in acts of property destruction.

[156] See Walker in *Moral Repair* (2006) for several detailed case studies of reconciliation committees and their varying degrees of effectiveness.

Though actual experiences of trauma are singular, having a sense of the variables at play allows us to roughly classify trauma and moral injury into different categories. The possibility of such classification can, in turn, allow us to identify patterns of healing and types of actions that are effective for fostering, maintaining, and repairing these two foundational forms of trust. I believe we can also discover more about the practices and principles that reliably foster, maintain, and repair trust, through the exploration of healthy, typical developments as well. For instance, by looking at the way trust is fostered and maintained in healthy friendships, physician-patient, teacher-student, and parent-child relationships. I plan to explore both paths in my future research.

Conclusion:

In this closing chapter, I've explored the way trust is interwoven in the process of identity formation by closely examining the links between self-trust and trust in the world and our capacity for memory and action. Building on Brennan's concept of recognition trust, I've shown that the typical course of development for self-trust and trust in the world begins with the need-meeting interactions that form the recognition trust we first experience in infancy and, if we are fortunate, throughout the rest of our lives. Then, drawing from the trauma studies literature, I surveyed several variables associated with atypical developments that prevent or disrupt one's self-trust or trust in the world (and therefore constrain one's ability to form a stable personal identity). In the process, I considered how these insights can guide and shape our understanding of what is required to establish and repair trust.

The developmental account of trust I offer, unlike the doxastic, affective, and ordinary language accounts reviewed earlier, was designed with the embodied experience of trust in mind. To attend fully to how trust feels, we need to consider the interaction between our bodies and our

external environments. What this detour into the trauma studies literature has aimed to show is that addressing the practical problem of distrust, whether at the interpersonal or at broader levels, requires an interdisciplinary approach. This includes research from other disciplines, but more than that, we also need to learn from the experiences of scholars and others who are actively involved in or tasked with the fostering, maintaining, and repairing of trust within their communities.

References

Addams, Jane. *Democracy and Social Ethics / Jane Addams; Introduction by Charlene Haddock Seigfried.* Urbana, Illinois: University of Illinois Press, 2002.

———.*Newer Ideals of Peace. / Jane Addams; Introduction by Bernice A. Carroll and Clinton F. Fink.* Urbana, Illinois: University of Illinois Press, 2007.

———.*The Long Road of Woman's Memory / by Jane Addams.* New York, New York: The Macmillan Company, 1916.

Almassi, Ben. "Relationally Responsive Expert Trustworthiness." *Social Epistemology* 36, no. 5 (2022): 576–85. https://doi.org/10.1080/02691728.2022.2103475.

Améry, Jean. *At the Mind's Limits: Contemplations by a Survivor on Auschwitz and Its Realities / by Jean Améry; Translated by Sidney Rosenfeld and Stella P. Rosenfeld.* Bloomington, Indiana: Indiana University Press, 1980.

Aristotle. *The Nicomachean Ethics / Sir David Ross and Lesley Brown (Eds)*. Revised edition. Oxford World's Classics. Oxford, United Kingdom: Oxford University Press, 2020.

Baier, Annette. *Death and Character : Further Reflections on Hume / Annette C. Baier*. Cambridge, Massachusetts: Harvard University Press, 2008. https://doi.org/10.4159/9780674268913.

———. *Reflections on How We Live / Annette C. Baier.* Oxford ; Oxford University Press, 2009.

———. *Moral Prejudices: Essays on Ethics / Annette C. Baier.* Cambridge, Massachusetts: Harvard University Press, 1994.

———. "Trust and Antitrust." *Ethics* 96, no. 2 (1986): 231–60.

Bennett, Matthew. "Demoralizing Trust." *Ethics* 131, no. 3 (2021): 511–38. https://doi.org/10.1086/712564.

———. "Judging Expert Trustworthiness: The Difference Between Believing and Following the Science." *Social Epistemology* 36, no. 5 (2022): 550–60. https://doi.org/10.1080/02691728.2022.2106459.

Bernstein, J. M. "Trust: On the Real but Almost Always Unnoticed, Ever-Changing Foundation of Ethical Life." *Metaphilosophy* 42, no. 4 (2011): 395–416. https://doi.org/10.1111/j.1467-9973.2011.01709.x.

Branch, T. Y. "Enhanced Epistemic Trust and the Value-Free Ideal as a Social Indicator of Trust." *Social Epistemology* 36, no. 5 (2022): 561–75. https://doi.org/10.1080/02691728.2022.2114114.

——— and Gloria Origgi. “Social Indicators of Trust in the Age of Informational Chaos.” *Social Epistemology* 36, no. 5 (2022): 533–40. https://doi.org/10.1080/02691728.2022.2121622.
Brennan, Johnny. “Can Novices Trust Themselves to Choose Trustworthy Experts? Reasons for (Reserved) Optimism.” *Social Epistemology* 34, no. 3 (2020): 227–40.
———. “Recognition Trust.” *Philosophical Studies* 178, no. 11 (November 2021): 3799–3818. https://doi.org/10.1007/s11098-021-01626-z.
Brison, Susan J. *Aftermath: Violence and the Remaking of a Self / Susan J. Brison.* Princeton, New Jersey: Princeton University Press, 2002. https://doi.org/10.1515/9781400841493.
Camporesi, Silvia, Maria Vaccarella, and Mark Davis. “Investigating Public Trust in Expert Knowledge: Narrative, Ethics, and Engagement.” *Journal of Bioethical Inquiry* 14, no. 1 (2017): 23–30. https://doi.org/10.1007/s11673-016-9767-4.
Cavarero, Adriana. *Relating Narratives: Storytelling and Selfhood*. Routledge, 1997.
Chang, Ruth. “Are Hard Choices Cases of Incomparability?” *Philosophical Issues* 22, no. 1 (2012): 106–26. https://doi.org/10.1111/j.1533-6077.2012.00239.x.
———. “Hard Choices.” *Journal of the American Philosophical Association* 3, no. 1 (2017): 1–21. https://doi.org/10.1017/apa.2017.7.
———. “The Possibility of Parity.” *Ethics* 112, no. 4 (2002): 659–88. https://doi.org/10.1086/339673.
Conklin, Sherri Lynn, Irina Artamonova, and Nicole Hassoun. “The State of the Discipline: New data on Women Faculty in Philosophy”. *Ergo* 6, no 30 (2019-2020). https://doi.org/10.3998/ergo.12405314.0006.030
Cunningham, Kelly Ann. “Cooperation and Climates of Trust”. *Social Theory and Practice.* (forthcoming)
Darwall, Stephen L. “Two Kinds of Respect.” *Ethics* 88, no. 1 (1977): 36–49. https://doi.org/10.1086/292054.
D’Arms, Justin, and Daniel Jacobson. “The Moralistic Fallacy: On the ‘Appropriateness’ of Emotions.” *Philosophical and Phenomenological Research* 61, no. 1 (2000): 65–90.
D’Cruz, Jason. “Humble Trust.” *Philosophical Studies* 176, no. 4 (2019): 933–53. https://doi.org/10.1007/s11098-018-1220-6
———. Manuscript. (forthcoming)
———. “Trust and Distrust” in *The Routledge Handbook of Trust and Philosophy/ Judith Simon (eds),* 41-51 New York, New York: Routledge, 2020.
———. “Trust, Trustworthiness, and the Moral Consequence of Consistency.” *Journal of the American Philosophical Association* 1, no. 3 (2015): 467–84. https://doi.org/10.1017/apa.2015.3.
———. “Trust within Limits.” *International Journal of Philosophical Studies* 26, no. 2 (2018): 240–50. https://doi.org/10.1080/09672559.2018.1450080.
Dover, Daniela. “The Conversational Self.” *Mind* 131, no. 521 (2022): 193–230. https://doi.org/10.1093/mind/fzab069.
Englund, Tomas. “The Potential of Education for Creating Mutual Trust: Schools as Sites for Deliberation.” *Educational Philosophy and Theory* 43, no. 3 (2011): 236–48.
Ewing, Eve L. *Ghosts in the Schoolyard : Racism and School Closings on Chicago’s South Side / Eve L. Ewing.* Chicago, Illinois: The University of Chicago Press, 2018.
Faulkner, Paul. “Finding Trust in Government.” *Journal of Social Philosophy* 49, no. 4 (2018): 626–44. https://doi.org/10.1111/josp.12262.
———. “On Telling and Trusting.” *Mind* 116, no. 464 (November 1, 2007): 875–902. https://doi.org/10.1093/mind/fzm875.

———. “The Practical Rationality of Trust.” *Synthese* 191, no. 9 (2014): 1975–89. https://doi.org/10.1007/s11229-012-0103-1.
Flanigan, Jessica. *Pharmaceutical Freedom : Why Patients Have a Right to Self Medicate / Jessica Flanigan. [Electronic Resource]*. New York, New York: Oxford University Press, 2017. https://doi.org/10.1177/1043463103154004.
Gawande, Atul *Being Mortal Medicine and What Matters in the End/ Atul Gawande*. New York, New York: Metropolitan Books, Henry Holt and Company, 2014.
Goldhaber, Daniel, dir. *How to Blow Up a Pipeline*. Screenplay by Ariela Barer, Jordan Sjol, and Daniel Goldhaber. (2022; Neon) Hulu.
Govier, Trudy. “Distrust as a Practical Problem” *Journal of Social Philosophy* 23, no. 1 (1992): 52–63. https://doi.org/10.1111/j.1467-9833.1992.tb00484.x.
Hardin, Russell. “The Street-Level Epistemology of Trust.” *Politics & Society* 21, no. 4 (1993): 505–29. https://doi.org/10.1177/0032329293021004006.
Hardwig, John. “Epistemic Dependence.” *The Journal of Philosophy* 82, no. 7 (1985): 335–49. https://doi.org/10.2307/2026523.
Hawley, Katherine. “Trust, Distrust and Commitment: Trust, Distrust and Commitment.” *Noûs* 48, no. 1 (2014): 1–20. https://doi.org/10.1111/nous.12000.
———. “Trustworthy Groups and Organisations,” in *The Philosophy of Trust/ Paul Faulkner and Thomas Simpson (eds).* Oxford, Great Britian: Oxford University Press, 2017.
Held, Virginia. *The Ethics of Care: Personal, Political, and Global/ Virginia Held.* Oxford, United Kingdom: Oxford University Press, 2006.
Hieronymi, Pamela. *Freedom, Resentment, and the Metaphysics of Morals/Pamela Hieronymi*. Princeton, New Jersey: Princeton University Press, 2020.
———. “The Reasons of Trust.” *Australasian Journal of Philosophy* 86, no. 2 (June 2008): 213–36. https://doi.org/10.1080/00048400801886496
Holton, Richard. “Deciding to Trust, Coming to Believe.” *Australasian Journal of Philosophy* 72, no. 1 (March 1994): 63–76. https://doi.org/10.1080/00048409412345881.
Honneth, Axel. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. MIT Press, 1996.
hooks, bell. *All About Love: New Visions/ bell hooks.* New York, New York: Harper Collins Publishers, 2001.
———.*Teaching to Transgress : Education as the Practice of Freedom / bell hooks.* New York: Routledge, 1994.
James, William. “What Is an Emotion?” *Mind* 9, no. 34 (1884): 188–205. http://www.jstor.org/stable/2246769.
Jeske, Diane. “Feminism, Friendship, and Philosophy.” *Canadian Journal of Philosophy* 32 (2002): 63–82. https://doi.org/10.1080/00455091.2002.10717583.
———. “Friendship and Reasons of Intimacy.” *Philosophy and Phenomenological Research* 63, no. 2 (2001): 329–46. https://doi.org/10.1111/j.1933-1592.2001.tb00106.x.
———. “Friendship and the Grounds of Reasons.” *Les Ateliers de l’éthique* 3, no. 1 (April 13, 2018): 61–69. https://doi.org/10.7202/1044605ar.
Jones, Karen. “Emotional Rationality as Practical Rationality.” In *Setting the Moral Compass/ Cheshire Calhoun (eds),* 333–52. New York, New York: Oxford University Press, 2003. https://doi.org/10.1093/oso/9780195154740.003.0019.
———. “Trust as an Affective Attitude.” *Ethics* 107, no. 1 (October 1996): 4. https://doi.org/10.1086/233694.

———. “Trust, Distrust, and Affective Looping.” *Philosophical Studies* 176, no. 4 (April 2019): 955–68. https://doi.org/10.1007/s11098-018-1221-5.
———. “Trustworthiness.” *Ethics* 123, no. 1 (October 2012): 61–85. https://doi.org/10.1086/667838.
Jaiswal, Jessica. “Whose Responsibility Is It to Dismantle Medical Mistrust? Future Directions for Researchers and Health Care Providers.” *Behavioral Medicine (Washington, D.C.)* 45, no. 2 (2019): 188–96. https://doi.org/10.1080/08964289.2019.1630357.
Katz, Ralph V, B. Lee Green, Nancy R. Kressin, Sherman A. James, Min Qi Wang, Cristina Claudio, and Stephanie Luise Russell. “Exploring the ‘Legacy’ of the Tuskegee Syphilis Study: A Follow-up Study From the Tuskegee Legacy Project.” *Journal of the National Medical Association* 101, no. 2 (2009): 179–83. https://doi.org/10.1016/S0027-9684(15)30833-6.
———, S. Stephen Kegeles, Nancy R. Kressin, B. Lee Green, Sherman A. James, Min Qi Wang, Stefanie L. Russell, and Cristina Claudio. “Awareness of the Tuskegee Syphilis Study and the US Presidential Apology and Their Influence on Minority Participation in Biomedical Research.” *American Journal of Public Health (1971)* 98, no. 6 (2008): 1137–42. https://doi.org/10.2105/AJPH.2006.100131.
Kelp, Christoph, and Mona Simion. “What Is Trustworthiness?” *Noûs* 57, no. 3 (2023): 1-17. https://doi.org/10.1111/nous.12448.
Keren, Arnon. “Trust and Belief.” in *The Routledge Handbook of Trust and Philosophy / Judith Simon (eds),*109–20 New York, New York: Routledge, 2020.
Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity/ Christine M. Korsgaard.* Oxford, United Kingdom: Oxford University Press, 2009.
Kirton, Andrew. “Matters of Trust as Matters of Attachment Security.” *International Journal of Philosophical Studies* 28, no. 5 (2020): 583–602. https://doi.org/10.1080/09672559.2020.1802971.
Krishnamurthy, Meena. “(White) Tyranny and the Democratic Value of Distrust.” *The Monist* 98, no. 4 (2015): 391–406. https://doi.org/10.1093/monist/onv020.
Lenard, Patti T. “The Decline of Trust, The Decline of Democracy?” *Critical Review of International Social and Political Philosophy* 8.3 (2005), 363–78
Lagerspetz, Olli. *Trust: The Tacit Demand / by Olli Lagerspetz.* Library of Ethics and Applied Philosophy, Dordrecht, The Netherlands: Kluwer Academic Publishers, 1998.
Lanphier, Elizabeth, and Leah Lomotey-Nakon. “Birth, Trust and Consent: Reasonable Mistrust and Trauma-Informed Remedies.” *Journal of Medical Ethics* 49, no. 9 (2023): 624–25. https://doi.org/10.1136/jme-2023-109210.
Malm, Andreas. *How to Blow up a Pipeline : Learning to Fight in a World on Fire / Andreas Malm.* First edition paperback. London, Great Britian: Verso, 2021.
Mason, Cathy. “The Epistemic Demands of Friendship: Friendship as Inherently Knowledge-Involving.” *Synthese* 199, no. 1–2 (2021): 2439–55. https://doi.org/10.1007/s11229-020-02892-w.
Misak, Cheryl J. “Narrative Evidence and Evidence-Based Medicine: Narrative Evidence and EBM.” *Journal of Evaluation in Clinical Practice* 16, no. 2 (2010): 392–97. https://doi.org/10.1111/j.1365-2753.2010.01407.x.
McGeer, Victoria. “Trust, Hope and Empowerment.” *Australasian Journal of Philosophy* 86, no. 2 (June 2008): 237–54. https://doi.org/10.1080/00048400801886413.
McKneally, Martin F, Esther Ignagni, Douglas K Martin, and Jason D’Cruz. “The Leap to Trust: Perspective of Cholecystectomy Patients on Informed Decision Making and Consent.” *Journal of*

the American College of Surgeons 199, no. 1 (2004): 51–57. https://doi.org/10.1016/j.jamcollsurg.2004.02.021.
McKneally, Martin F., Douglas K. Martin, Esther Ignagni, and Jason D'Cruz. "Responding to Trust: Surgeons' Perspective on Informed Consent." *World Journal of Surgery* 33, no. 7 (2009): 1341–47. https://doi.org/10.1007/s00268-009-0021-7.
McLeod, Carolyn. *Self-Trust and Reproductive Autonomy/ Carolyn McLeod*. Cambridge, Massachusetts: The MIT Press, 2002. https://doi.org/10.7551/mitpress/6157.001.0001.
Mill, John Stuart. *Utilitarianism/ John Stuart Mill, George Sher (eds)*. Indianapolis, Indiana: Hackett Publishing Company, 2001 [1806-1873].
Morton, Jennifer M. *Moving Up without Losing Your Way: The Ethical Costs of Upward Mobility/ Jennifer M. Morton.* Princeton, New Jersey: Princeton University Press, 2019.
Mullin, Amy. "Trust, Social Norms, and Motherhood." *Journal of Social Philosophy* 36, no. 3 (2005): 316–30. https://doi.org/10.1111/j.1467-9833.2005.00278.x.
Murphy-Hollies, Kathleen. "Self-Regulation and Political Confabulation." *Royal Institute of Philosophy Supplement* 92 (2022): 111–28.
Nguyen, C. Thi. "Trust as an Unquestioning Attitude." in *Oxford Studies in Epistemology Volume 7 1st ed. / Tamar Szabó Gendler, John Hawthorne, and Julianne Chung (eds)*, 214–244. Oxford, Great Britian: Oxford University Press, 2022. https://doi.org/10.1093/oso/9780192868978.003.0007.
Nickel, Philip J. "Trust and Obligation-Ascription." *Ethical Theory and Moral Practice* 10, no. 3 (2007): 309–19.
Niker, Fay, and Laura Specker Sullivan. "Trusting Relationships and the Ethics of Interpersonal Action." *International Journal of Philosophical Studies* 26, no. 2 (2018): 173–86. https://doi.org/10.1080/09672559.2018.1450081.
Nunnlay, Shayla. *Trust in Black America/ Shayla Nunnlay*, New York: New York University, 2012.
Nussbaum, Martha Craven. *Upheavals of Thought : The Intelligence of Emotions / Martha C. Nussbaum. [Electronic Resource]*. Cambridge: Cambridge University Press, 2001.
O'Neill, Onora. "Integrity and Quality in Universities: Accountability, Excellence and Success" *British Academy Review* 20 (2012):41–44.
———. "Questioning Trust." in *The Routledge Handbook of Trust and Philosophy / Judith Simon (eds)*, 17-27 New York, New York: Routledge, 2020.
Östlund, Ruben, dir. *Triangle of Sadness.* Screenplay by Ruben Östlund (2022; Neon) Hulu.
Patrizio, Ross F. "Trust's Meno Problem: Can the Doxastic View Account for the Value of Trust?" *Philosophical Psychology* 37, no. 1 (2023): 1–20. https://doi.org/10.1080/09515089.2023.2206837.
Pedwell, Carolyn. *Revolutionary Routines : The Habits of Social Transformation / Carolyn Pedwell.* Montreal, Canada: McGill-Queen's University Press, 2021.
Petherbridge, Danielle. "Recognition, Vulnerability and Trust." *International Journal of Philosophical Studies* 29, no. 1 (2021): 1–23. https://doi.org/10.1080/09672559.2021.1885135.
Pettit, Philip. "The Cunning of Trust." *Philosophy and Public Affairs* 24, no. 3 (July 1995): 202–25. https://doi.org/10.1111/j.1088-4963.1995.tb00029.x.
Pham, Adam, and Clinton Castro. "The Moral Limits of the Market: The Case of Consumer Scoring Data." *Ethics and Information Technology* 21, no. 2 (2019): 117–26. https://doi.org/10.1007/s10676-019-09500-7.

———. Alan Rubel, and Clinton Castro. “Social Media, Emergent Manipulation, and Political Legitimacy.” in *The Philosophy of Online Manipulation, 1st ed.*, 353–69. New York, New York: Routledge, 2022. https://doi.org/10.4324/9781003205425-21.
Potter, Nancy Nyquist. *How Can I Be Trusted? : A Virtue Theory of Trustworthiness / Nancy Nyquist Potter*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 2002.
———. “Interpersonal Trust” in *The Routledge Handbook of Trust and Philosophy / Judith Simon (eds)*,243-255. New York, New York: Routledge, 2020.
Pouryousefi, Sareh, and Jonathan Tallant. “Empirical and Philosophical Reflections on Trust.” *Journal of the American Philosophical Association* 9, no.3 (2022): 1–21. https://doi.org/10.1017/apa.2022.14.
Press, The Associated. “Judge Sides with Young Activists in First-of-Its-Kind Climate Change Trial in Montana.” *NPR*, August 14, 2023, sec. Climate. https://www.npr.org/2023/08/14/1193780700/montana-climate-change-trial-ruling.
Simpson, Thomas W. *Trust: A Philosophical Study/ Thomas W. Simpson.* Oxford, Great Britian: Oxford University Press, 2023.
———. “Trust, Belief, and the Second-Personal.” *Australasian Journal of Philosophy* 96, no. 3 (2018): 447–59. https://doi.org/10.1080/00048402.2017.1382545.
———. “What Is Trust?” *Pacific Philosophical Quarterly* 93, no. 4 (2012): 550–69. https://doi.org/10.1111/j.1468-0114.2012.01438.x.
Smith, Sandra Susan. “Race and Trust” *Annual Review of Sociology* 26 (2010): 453-75.
Strawson, Peter. “Freedom and Resentment.” *Proceedings of the British Academy* 48 (1962): 187–211.
Talisse, Robert B. *Overdoing Democracy : Why We Must Put Politics in Its Place / Robert B. Talisse.* New York, New York: Oxford University Press, 2019.
Um, Sungwoo. “Trustfulness as a Risky Virtue.” *Journal of Humanities, Seoul National University* 81, no. 2 (2024): 315–44. https://doi.org/10.17326/JHSNU.81.2.202405.315.
Vallier, Kevin. *Trust in a Polarized Age / Kevin Vallier.* New York, NY: Oxford University Press, 2021.
van der Kolk, Bessel A. *The Body Keeps the Score: Brain, Mind, and Body in the Healing of Trauma/ Bessel van der Kolk.* New York, New York: Penguin Books, 2015.
Walker, Margaret Urban. *Moral Repair: Reconstructing Moral Relations after Wrongdoing / Margaret Urban Walker.* United Kingdom ; New York, New York: Cambridge University Press, 2006.
Warren, Mark, ed. *Democracy and Trust*. Cambridge, United Kingdom ; New York, New York: Cambridge University Press, 1999.
———. *Trust and Democracy Vol 1 Eric M. Uslaner (eds).* Oxford, United Kingdom: Oxford University Press, 2017. https://doi.org/10.1093/oxfordhb/9780190274801.013.5.
Wee, Cecilia. “Xin, Trust, and Confucian Ethics” *Philosophy East and West* 61, no 3 (2011): 516-533.
Williams, Bernard Arthur Owen., Jonathan. Lear, and A. W. Moore. *Ethics and the Limits of Philosophy / Bernard Williams; with a Commentary on the Text by A.W. Moore and a Foreword by Jonathan Lear*. London, Great Britian: Routledge, 2011.
Wolf, Susan. “Happiness and Meaning: Two Aspects of the Good Life.” *Social Philosophy and Policy* 14, no. 1 (1997): 207–25. https://doi.org/10.1017/S0265052500001734.
———. “Self-Interest and Interest in Selves.” *Ethics* 96, no. 4 (1986): 704–20. https://doi.org/10.1086/292796.

Yao, Vida. “The Good Fit.” *Philosophy and Phenomenological Research*, no. 2 (2023): 414–29.