9-2-2020

Trusting In Order to Inspire Trustworthiness

Michael Pace

Trusting In Order to Inspire Trustworthiness

Noname manuscript No.
(will be inserted by the editor)

Trusting in order to inspire trustworthiness

Michael Pace

the date of receipt and acceptance should be inserted later

Abstract This paper explores the epistemology and moral psychology of "therapeutic trust," in which one trusts with the aim of inspiring greater trust-responsiveness in the trusted. Theorists have appealed to alleged cases of rational therapeutic trust to show that trust can be adopted for broadly moral or practical reasons and to motivate accounts of trust that do not involve belief or confidence in someone's trustworthiness. Some conclude from the cases that trust consists in having normative expectations and adopting vulnerabilities with respect to the trusted; others that trust involves accepting (without necessarily believing) that someone will prove trustworthy. Although there are, I argue, some genuine cases of rational therapeutic trust, some prominent examples confuse trusting with entrusting and are actually counterexamples to the adopted vulnerabilities and acceptance accounts they have been taken to support. An alternative account, which construes trust in terms of being confident enough to take salient risks on someone's trustworthiness, makes better sense of therapeutic trust.

1 Introduction

In a famous scene from *Les Misérables*, Bishop Myriel gives Jean Valjean a valuable set of silver that Valjean has been caught stealing, trusting him to use the gift to become an honest man. Although Valjean is not, at that time, the sort of person for whom such trust is warranted, Bishop Myriel's trust inspires him to become the sort of person who is. The scene depicts the way that trust in others can inspire dramatic moral transformations. Powerful non-fiction cases are fairly easy to come by. Consider, for example, the way leaders

M. Pace
Philosophy Department, Chapman University, One University Drive, Orange, CA 92866, USA
E-mail: pace@chapman.edu

of non-violent resistance movements, such as Gandhi and Martin Luther King, Jr., trusted their followers to be capable of acts of extreme self-sacrifice, and thereby inspired them to live up to the high expectations embodied in their trust. I suspect many readers can think of personal stories in the same key, when the trust of a parent, teacher, or mentor has inspired them to become the sort of person worthy of the trust given to them.

That trust sometimes inspires greater trustworthiness is uncontroversial. More controversially, some philosophers have suggested that there are rational cases of "therapeutic trust," in which one trusts with the explicit aim of inspiring trustworthiness—harnessing, as reasons for trust, considerations about how one's own trust might inspire greater trust-responsiveness. Karen Jones suggests the following oft-cited example:

> [A] mother might trust her teenage daughter to look after the house for the weekend even though the daughter has failed in the past to meet such responsibilities with trustworthiness. The mother might think that by displaying her trust and not arranging to have the daughter stay at a friend's house or to have the neighbors keep an eye on the place, she can elicit trust-responsiveness in her daughter. The mother might lack confidence that the daughter will respond to trust with trustworthiness on this occasion, but she hopes that, eventually, her trust will be repaid with trustworthiness. [Jones, 2004, 16]

In Jones' description, the mother trusts the daughter with the house for the weekend because she wants to encourage her to become more trustworthy; her reasons for trust include the thought that trusting the daughter with the house will potentially inspire her to become more trustworthy in the long term (though probably not this weekend).

More generally, we can stipulate that a case of therapeutic trust is any case in which something akin to the following "therapeutic thought" serves as a reason for trust:

> Therapeutic Thought: If I were to trust S to ϕ, then my trust would likely inspire S to be more trust-responsive.

A second proposed example of therapeutic trust comes from Richard Holton:

> Suppose you run a small shop. And suppose you discover that the person you have recently employed has just been convicted of petty theft. Should you trust him with the till? It appears that you can really decide whether or not to do so. And again it appears that you can do so without believing that he is trustworthy. [Holton, 1994, 63]

Holton suggests that you could decide to trust the employee with the till because you think that your trust will be morally beneficial for the employee, encouraging or inspiring him to be more trust-responsive.[1]

[1] I follow Karen Frost-Arnold [2014] in interpreting this example as a case of therapeutic trust, although the motivations Holton suggests the shopowner might have for trusting do not obviously involve inspiring trustworthiness. He suggests the shopowner might trust in

Although Victor Hugo does not describe Bishop Myriel's reasons for trusting Jean Valjean in much detail, we can construct a third potential case of therapeutic trust by adding to the story that Bishop Myriel trusted Valjean, at least in part, because he predicted the inspirational effects his trust would likely have on Valjean.

The aim of this paper is to explore the epistemology and moral psychology of therapeutic trust, and to consider what the phenomenon has to teach about the nature and rationality of trust. Theorists such as Jones and Holton have appealed to the examples above as intuitively clear cases in which trust is rationally adopted for practical or moral reasons. So construed, the cases seem to be counterexamples to a prominent "evidence-only" view about reasons for trust, according to which only evidence of someone's present trustworthiness can properly serve as a reason for trusting them. However, as we shall see, making sense of rational therapeutic trust is more difficult than is sometimes thought, and we should not presuppose that all of the examples just given can ultimately be understood as genuine cases of rational trust.

The discussion will proceed as follows. In Section 2, I clarify the concept of interpersonal trust at issue and distinguish it from acts of entrusting with which it is easily confused. Section 3 presents what I take to be the strongest arguments for the view that only evidence of trustworthiness serves as reasons for trust, building on Hieronymi's [2008] contention that considerations about the usefulness of trust are the "wrong kind of reasons" for trust. When combined with plausible accounts of how trust might reasonably be expected to inspire greater trustworthiness, the arguments suggest a wrong kind of reasons challenge to explain how exactly the therapeutic thought can

I then evaluate two different strategies for responding to the arguments by appeal to cases of rational therapeutic trust. The two strategies reflect an important difference between the Bishop Myriel and mother-daughter cases. Bishop Myriel, we are supposing, expected that his trust in Valjean to use the silver to make himself a better man would inspire Valjean to do the very thing he is trusting him to do (use the silver to make himself a better man). In such "self-fulfilling prophecy" cases, as we will call them, the therapeutic thought is that trusting someone to ϕ will make it more likely that they will ϕ; trust is expected to be something like a self-fulfilling prophecy, causally contributing to its own fulfillment. Self-fulfilling prophecy cases contrast with "long-term" cases of therapeutic trust, such as the case of Jones' mother, who expects trusting the daughter with the house to have only long-term effects. (Holton's shopowner case can be stipulated to be either kind of case, depending on whether you expect that your trust in the employee will prompt him not to steal on this occasion.)

I argue that only self-fulfilling prophecy cases provide resources for an adequate response to the wrong kind of reasons challenge. In Section 4, focusing

order to "draw [the employee] back into the moral community" or because the shopowner thinks this is the way one ought to treat an employee. [Holton, 1994, 63] Other theorists define therapeutic trust less narrowly so that it can include other moral benefits to the trusted besides increasing trust-responsiveness.

on a version of the Bishop Myriel case, I argue that the therapeutic thought can make it rational to believe or boost one's confidence in someone's trustworthiness. In such cases, the therapeutic thought does not serve as evidence of trustworthiness but is epistemically relevant.

Most theorists in the literature focus on long-term cases such as Jones' mother-daughter case, arguing that therapeutic trust can be adopted for non-evidential reasons precisely because it is a kind of *non-doxastic* trust, which does not require an outright belief that someone will prove trustworthy. Thus, Holton says that you can trust the employee with the till without believing that he won't steal, and Jones suggests that the mother trusts despite having little confidence that the daughter will take good care of the house. Some go as far as to say that therapeutic trust can be adopted without any confidence at all that the trusted will prove trustworthy. For example, H.J.N. Horsburgh (who coined the "therapeutic trust" label) claims that therapeutic trust "is placed in response to what is believed to be a moral need, and so does not presuppose *any* favorable estimate of those in whom it is reposed." [Horsburgh, 1960, 349, emphasis mine]

Based in part on long-term cases, theorists have proposed two prominent alternatives to the idea that trust consists in belief or confidence of trustworthiness. Jones concludes from the mother-daughter case that trust consists in willingly adopting vulnerabilities with respect to the trusted and having *normative* expectations about the way they should behave. [Jones, 2004, McGeer, 2008, McLeod, 2015] Others argue that trust can consist in *accepting* or presupposing, without believing, that the trusted person will prove trustworthy. [Frost-Arnold, 2014, Holton, 1994]

In Section 5 I argue that there are strong reasons to doubt that there are rational cases of long-term therapeutic trust. Jones' mother-daughter case, in particular, is best construed as a case in which the mother merely entrusts the house to the daughter without trusting her. When properly understood, the case is a counterexample to the adopted vulnerabilities and acceptance accounts of non-doxastic trust that many take it to support.

Section 6 suggests a promising alternative account that takes trust to consist in being confident enough to take certain kinds of risk on someone's proving trustworthy. The confident-enough-to-risk account allows a significant but limited role for non-evidential considerations, including some therapeutic considerations.

2 Preliminaries

2.1 Trusting-With and Trusting-To

At least two important locutions in English can be used to pick out the attitude of interpersonal trust:[2]

[2] We also sometimes speak of trust as a two-place relation: 'A trusts B'. I set aside the question of the relation between 2-place trust and trusting people to perform certain actions,

1. 'A trusts B to ϕ', where ϕ is an action. (For example, "Bishop Myriel trusted Jean Valjean to use the silver to become a better man." "The mother trusted the daughter to care for the house.")
2. 'A trusts B *with* X', where X is a thing of value to A. (E.g., "Bishop Myriel trusted Valjean with the silver." "The mother trusted the daughter with the house.")

Both the trust-to and trust-with locutions can be used to describe our central examples, and both have played a role in the trust literature. It will thus be helpful to make a few points about their relation.[3]

First, the trust-to locution is a more general way of describing the attitude of trust. [Holton, 1994, 63] The attitude involved in trusting someone with a thing of value is a special case of trusting the person to ϕ, namely that of trusting the person *to care properly for the valuable thing*. Trust-to cannot similarly be explained in terms of trust-with, however, since not all cases of trusting someone to ϕ involve giving (or being disposed to give) a thing of value to the trusted. For example, there is no obvious *thing* one entrusts when one is trusting a friend to tell the truth or to bring wine to the party one is hosting.

Second, both locutions exhibit an interesting (and, so far as I know, unremarked upon) shift in meaning between the present tense construction ('A trusts B to ϕ' or 'A trusts B with X') and the present continuous ('A is trusting B to ϕ' or 'A is trusting B with X'). Compare, for example:

> *Present Tense:* The mother trusts the daughter to care for the house. The mother trusts the daughter with the house.
> *Present Continuous Tense:* The mother is trusting the daughter to care for the house. The mother is trusting the daughter with the house.

The claims in the present continuous tense are logically stronger, implying that the trustor performs an action or actions that manifest the attitude of trust. To say that the mother *is trusting* the daughter to care for the house for the weekend implies that the mother is acting on her trust in the daughter to care for the house. The present tense constructions attribute the attitude of trust but do not imply that the trustor takes any such actions (though they may carry a strong conversational implicature to this effect). It can be true that a mother trusts the daughter with the house (or, trusts her to take proper care of it) even though she has never left the daughter alone for the weekend and has invited the daughter's grandparents over this weekend while she's away. "It's not that I don't trust you with the house," she might explain to her daughter. "I do trust you with it, but your grandparents really wanted to come and spend time with you."

---

though I am somewhat inclined to think that the former can be understood (though probably not cleanly defined) in terms of the latter. However, see Faulkner [2015], Domenicucci and Holton [2017] for arguments that the two-place trust relation is more fundamental.

[3] The first has been the dominant focus of analysis in the recent literature, although Annette Baier [1986], the wellspring of contemporary philosophical discussion of trust, takes trusting people with things of value as her *analysandum*.

The present continuous tense is useful for describing actions that manifest the attitude of trust, actions that the trustor performs *because* they trust. In the case of your trusting someone with X, the actions that manifest trust involve your giving or entrusting X to someone's care. Thus, saying that the mother is trusting the daughter with the house implies that the mother entrusts the house to the daughter's care *because* she trusts the daughter to properly care for it. (More generally, if A is trusting B with X, then A entrusts X to B because A trusts B to care properly for X.)

2.2 Trusting vs. Entrusting

Note well: here and throughout I will use "trust" and "entrust" to mark an important distinction that is present in ordinary English usage, but has the potential to cause confusion. The *Pocket OED* defines "entrust" as follows:

> 1. assign the responsibility for doing something to (someone). "I've been entrusted with the task of getting him safely back."
> 1.1 put (something) into someone's care or protection. "You persuade people to entrust their savings to you."

So defined, entrusting does not require having the attitude of trust.[4] One can assign responsibility for doing something or put something into someone's care without trusting the person to fulfill the responsibility or properly care for the thing. You might, for example, *en*trust a secret to someone whom you do not trust to keep the secret. Perhaps you want the information to be leaked, or you want to catch the person in the act of spreading the secret. Likewise, a mother might entrust the house to a teenage daughter for the weekend without *trusting* her with the house. In a much less happy case than the one Jones describes, for example, a mother might entrust the house to the daughter without any therapeutic motive, expecting her to fail and hoping to shame her when she does.[5]

The distinction between trusting and entrusting suggests a challenge that an adequate theory of trust must meet: it must properly distinguish between actions that manifest a genuine attitude of trust and those that are merely cases of entrusting. Section 5 will later suggest that some recent accounts of trust fail to meet this challenge.

[4] Readers are invited to take this as stipulative if they doubt that ordinary English usage always distinguishes quite so cleanly between entrusting as an action and trusting as an attitude.

[5] Hieronymi [2008] uses "entrust" in a different way, to mean having an attitude of trust that does not involve outright belief in someone's trustworthiness. This does not accord with ordinary usage, which allows for cases in which one entrusts but lacks trust (or even distrusts) and cases in which one both entrusts and believes the other will prove trustworthy.

2.3 Trust and Reliance

A further point of clarification, common in the recent literature on trust, concerns the difference between trust and reliance;[6] reliance is insufficient for trust.[7] Whereas we commonly rely on inanimate objects (e.g., we rely on a rope to hold us), we trust inanimate objects only in a metaphorical usage. Further, we can rely on persons without trusting them, by merely incorporating predictions of their actions into our practical plans. For example, you might predict that an opponent in a game will make a bad move and so rely on them to do so, but you will not thereby count as trusting them to make the bad move. Moreover, whereas a mafioso trusts his underlings to the extent that he expects them to act out of loyalty, he merely relies on their predictable actions to the extent to which he motivates them by threats, bribes, or the making of offers that cannot be refused. If he genuinely trusted them he would not take himself to need these methods of motivation.

What is this difference between trust and reliance? We will assume (again, following most of the recent literature) that trusting someone to ϕ, unlike relying on the person, involves some kind of expectation not only that the person will perform ϕ but also that they will do so for the right kinds of reasons—that they will, as we sometimes say, "prove trustworthy" in ϕ-ing.[8] Proposals for what it is to prove trustworthy, acting for the "right kind of reasons," have included 1) acting out of goodwill toward the trustor, [Baier, 1986]; 2) following through on a commitment to ϕ [Hawley, 2014]; and 3) being motivated to ϕ by the thought that the trustor is relying, depending, or counting on you. [Faulkner, 2014, 2007, Jones, 1996, McGeer and Pettit, 2017].

Although we need not settle on a general account of what it is to expect someone to prove trustworthy in performing an action, it is worth noting that the distinction between trusting and entrusting suggests a plausible account for cases of trusting-with. When one trusts someone with a thing of value, the expectation seems to be that the person will properly care for X, at least in part, because they have been *entrusted* with X. Entrusting someone with X is a normatively richer action than merely giving X; it involves assigning

[6] [Baier, 1986, Holton, 1994, Jones, 1996]. Theorists differ as to whether this distinction is always present in English usage. Faulkner suggests that "trust" has two senses; he uses "affective trust" to name the sense that contrasts with reliance and "predictive trust" for a usage that he takes to be tantamount to mere reliance. McGeer and Pettit [2017] suggest that the concept of trust that contrasts with reliance may have become a philosopher's term of art.

[7] Many in the literature hold that reliance is nevertheless necessary for trust, or even that trust is a special case of reliance. See Thompson [2017] for criticism of this idea. We need not take a stand on this issue here.

[8] Someone's "proving trustworthy" in performing an action is not intended to imply that the person has the general trait of trustworthiness. In the intended usage, an untrustworthy person might prove trustworthy in performing an action on a single occasion, while remaining generally untrustworthy. This, I think, accords with standard English usage. Even so, I will often prefer "proves trust-responsive" to "proves trustworthy" in order to avoid the misleading impression that it is the general character trait that is in view.

responsibility for properly caring for X. Entrusting people thus places normative demands on them, giving them reasons to perform the actions that they would not otherwise have. When one trusts someone with X, one expects the fact that one has entrusted them to "weigh with them as a reason" to care properly for X.[9]

3 The Wrong Kind of Reasons Challenge

3.1 Two Arguments for the Evidence-Only View

With these preliminary points about the attitude of trust in hand, let us turn to considering the kinds of reasons that can support trust. I will take for granted that one main kind of reason for trusting someone involves evidence that someone will prove trust-responsive.[10] Let us say that an *evidential reason* for a proposition is a consideration that raises the epistemic probability of a proposition given one's total evidence. (We leave open a wide range of options for understanding epistemic probability and what constitutes one's total evidence.) Paradigm cases of evidential reasons for thinking someone will prove trustworthy include track-record arguments supporting the claim that someone is generally trustworthy or trustworthy in some particualar domain. Online ratings of others, for example, give us evidence that others will follow through on their commitments in the marketplace and serve as reasons on the basis of which many of us are willing to trust total strangers.

Pamela Hieronymi [2008] argues for the stronger claim that *only* such evidence of trustworthiness can serve as reasons for trust. Thus, Hieronymi:

> [O]ne trusts another to do something to the degree that one harbours a trusting belief that the other will do that thing. Like any belief, this trusting belief (if it is supported by reasons at all) can only be supported by reasons which one takes to bear on its truth....[R]easons that show trust good, important, etc., will not bear on the truth of the trusting belief—they will not bear on whether the person in question will do the thing in question. [Hieronymi, 2008, 235]

This passage actually suggests two distinct arguments. The first, and more widely discussed, begins with a claim about the nature of outright trust: To trust someone to do something is just to have an outright belief that they will prove trust-responsive in doing the thing in question. Call this the Doxastic Account of Trust:

> Doxastic Account of Trust: A trusts B to ϕ if and only if A believes that B will prove trustworthy in ϕ-ing.

[9] I borrow the language of "weighing with someone as a reason" from [McGeer and Pettit, 2017, 15].

[10] One important stream of work on trust, which takes inspiration from Richard Moran's work on testimony, challenges the idea that evidence can serve as a reason for trust. [Moran, 2005, McMyler, 2011, Marušić, 2015, 2017] Unfortunately, I cannot adequately engage these arguments here. See Simpson [2018] for relevant criticism.

A Doxastic Account of Trust is not uncommon.[11] Hieronymi gives an influential argument for it (to which we will respond in Section 5.1) based on an example in which you decide to show up at a restaurant at which you've agreed to meet her, even though you have serious doubts and are agnostic about whether she will show up. Hieronymi says that if she were to find out that you did not believe that she would show up, she could properly complain that you did not trust her:

> Certainly your actions are somehow more trusting than those of someone who, in the face of such doubt, did not come to the restaurant at all. Nevertheless, I could rightly complain that your lack of confidence betrays a lack of trust. [Hieronymi, 2008, 218]

Given this Doxastic Account of Trust, the conclusion that only evidence for trustworthiness can serve as genuine reasons for trust is entailed by the following Evidence-Only Thesis for Belief:

> Evidence-Only Thesis for Belief: The only reasons that can make it rational to believe p are evidential reasons for p.

This epistemological thesis is widely held,[12] though not uncontroversial.[13] A familiar way of motivating the Evidence-Only Thesis for Belief begins with the observation that practical incentives seem impotent in producing belief, for someone rational, in the face of strong evidence against a proposition. If offered $10,000 to believe that the moon is made of green cheese, for example, you could not use this consideration as a reason to believe that it is true that the moon is made of green cheese. [Alston, 1988] Whereas we can change our beliefs in the light of new evidence, considerations about the desirability or usefulness of belief seem to be the wrong kind of reasons for belief.

The Doxastic Account of Trust and the Evidence-Only Thesis for Belief together entail the following Evidence-Only Thesis for Trust:

> Evidence-Only Thesis for Trust: The only reasons that can make it rational for A to trust B to ϕ are evidential reasons for thinking that B will prove trustworthy to ϕ.

Although this "wrong kind of reasons" challenge is framed in terms of outright trust and outright belief, it is notable that Hieronymi talks in terms of degrees of trust and belief in the passage cited above. One trusts, she says, "*to the degree that* one harbors a trusting belief." The passage suggests a second (and

[11] On Richard Hardin's view, for example, "The declarations 'I believe you are trustworthy' and 'I trust you' are equivalent." [Hardin, 2004, 10] Theorists who hold that trusting someone to ϕ requires believing that they will ϕ include [Baker, 1987, Hieronymi, 2008, McMyler, 2011, Marušić, 2015, 2017].

[12] It is endorsed by Shah and David Velleman [2005], Kelly [2003], Hieronymi [2008], Martin [2013], among others. Also related are Uniqueness theses, such as the one defended by Feldman [2007], according to which a body of evidence justifies only one doxastic attitude *vis-a-vis* belief, disbelief, and withholding.

[13] See Sylvan [2016] for discussion. Pragmatic encroachment accounts of justification deny it. See [Fantl and McGrath, 2009, Pace, 2011].

under-discussed) wrong kind of reasons challenge, targetting the claim that non-evidential reasons can properly *strengthen* trust.

To see it, begin with the idea that strength of trust is fundamentally a matter of one's degree of confidence that someone will prove trustworthy:

> Degree of Confidence Account of Strength of Trust: How strongly A trusts B to ϕ depends solely on A's degree of confidence that B will prove trustworthy to ϕ.

The account is intended to appeal only to a non-technical conception of degrees of confidence; we can think of a degree of confidence as the likelihood of a proposition from a subject's perspective, leaving open further ways of understanding the notion. So construed, there is considerable initial plausibility to explaining strength of trust in terms of confidence that someone will prove trustworthy. After all, natural replies to questions about strength of trust — e.g., "How strong is your trust in me to do what I promised?" — are quite naturally framed in terms of degrees of confidence: "I'm completely confident that you will do what you promised," or "I am not at all confident that you will do what you promised."

A common epistemological claim concerning degrees of confidence, somewhat less controversial than the Evidence-Only Thesis for Belief, is the following:

> Evidence-Only Thesis for Degrees of Confidence: The only reasons that can make it rational to increase one's degree of confidence in a proposition are evidential considerations in favor of the proposition.

Given the Degrees of Confidence Account of Strength of Trust and the Evidence-Only Thesis for Degrees of Confidence, we get the following thesis that only evidence of trustworthiness can strengthen trust:

> Evidence-Only Thesis for Strength of Trust: The only reasons that can make it rational for A to strengthen trust in B to ϕ are evidential considerations suggesting that B will prove trustworthy to ϕ.

In sum, we have seen two wrong kind of reasons challenges to the idea that there are non-evidential reasons for trust. One relies on a view about the nature of outright trust—that it consists in an outright belief about someone's proving trustworthy—and concludes on epistemic grounds that only evidential considerations can serve as reasons for trust. The other relies on a claim about strength of trust—that it consists in one's degree of confidence that someone will prove trustworthy—and concludes on epistemic grounds that only evidence can rationally strengthen trust.

How might one respond to these arguments by appeal to cases of therapeutic trust? Recall that cases of therapeutic trust were stipulated to be cases in which the following therapeutic thought (or its ilk) serves as a reason for trusting someone to ϕ:

> Therapeutic Thought: If I were to trust B to ϕ, then my trust would likely inspire B to be more trust-responsive.

In self-fulfilling prophecy cases, one expects that trusting someone to ϕ would inspire them to prove trust-responsive to ϕ. In long-term cases, one expects only that trust would inspire greater trust-responsiveness at some time in the future.

Self-fulfilling prophecy and long-term cases make possible quite different responses to the wrong kind of reasons challenges. In Section 4, we will consider a response based on self-fulfilling prophecy cases of therapeutic trust that targets the epistemic premises (the evidence-only theses for belief and degrees of confidence). In Section 5, we consider a response, based on long term cases, that denies the doxastic account of trust. (As we will see, though, the strategy founders in answering the wrong kind of reasons challenge to strength of trust.)

3.2 Therapeutic Trust To the Rescue?

Before discussing these responses, though, let us consider an initial reason to doubt that therapeutic trust will escape the wrong kind of reasons challenge.

Notice, first, that in cases of therapeutic trust, the trustor expects the attitude of trust—and not merely the act of entrusting—to elicit greater trust-responsiveness. If the point made in Section 2.3 is correct, ordinary cases of trust routinely, and perhaps constitutively, involve an expectation that the act of entrusting someone with an object or a task will have an effect on them, making it more likely that they will act appropriately. Therapeutic trust involves the stronger expectation that the trusted will be moved also by the recognition that you trust them, in a manner similar to the examples of inspirational trust that began this paper. The inspirational effects in such cases would be undermined if the trustee failed to believe that the trust is genuine. For example, if Valjean were to believe that Bishop Myriel was merely pretending to trust, or if the daughter were to believe this of the mother, the motivational effects would disappear.

How might trust reasonably be expected to inspire the trusted? It will be helpful to consider in more detail how trust inspires in ordinary, non-therapeutic cases. Suppose we consider the Bishop Myriel example, as we did at the outset, as a paradigm case in which trust inspires, without adding the more complex assumption that Bishop Myriel based his trust on therapeutic considerations. What is it about Bishop Myriel's trust that inspired Valjean to become more trust-responsive?

McGeer and Pettit [2017] propose two main mechanisms to explain how trust inspires trustworthiness, each of which is plausibly at work in the Bishop Myriel example.[14] First, manifesting trust can be a source of encouragement

[14] McGeer and Pettit also discuss a third, "request-based" mechanism. [McGeer and Pettit, 2017, 24] Manifesting trust in someone, they suggest, often involves an explicit or implicit request that they do the thing in question. If the trusted does not decline the request, they thereby make a tacit promise to do the thing in question, and they thus gain whatever extra motivational oomph comes from making a promise as opposed to merely declaring an intention. (This mechanism is also nicely illustrated in *Les Mis.* Bishop Myriel implores Valjean: "Do not forget, never forget, that you have promised to use this money in becoming

for the trusted, much like having someone shout "You can do it!" as you are biking up a steep hill. The thought that you have been sincerely trusted to do something can boost your confidence in your own capacity to prove trust-responsive, thereby strengthening it. [McGeer and Pettit, 2017, 22] Applied to the Myriel-Valjean case, for example, Valjean's belief that Bishop Myriel trusted him with the silver plausibly encouraged Valjean to prove trust-responsive by helping him to think of himself as someone capable of doing so.[15]

Second, manifesting trust can inspire in virtue of the way trust shows respect or esteem to the trusted as someone who is trust-responsive. (Cf. Pettit [1995]) The trusted might be motivated to be more trust-responsive in order to maintain the respect and esteem of the trustor. Thus, we might plasubily suppose that Valjean was motivated to use the silver to become a better man, in part, because he wanted to maintain the respect and esteem Bishop Myriel showed in trusting him. Having the respect and esteem of those whom we respect and esteem can be a powerful motivator.[16]

These mechanisms, which plausibly explain trust's power to inspire trustworthiness, also suggest a *prima facie* reason to worry that therapeutic trust will not escape the wrong kind of reasons problem. Each mechanism seems to require, for its operation, that one's manifest trust communicates a belief or high degree of confidence in the trusted to prove trust-responsive. What the trusted regards as encouraging or esteeming seems to be the apparent fact that the trustor has a confident belief that they will prove trustworthy. As McGeer suggests in earlier work, when we are inspired by the trust of a respected parent or mentor, "the galvanizing thought that drives us forward is...'I want to be as she already sees me to be.'" [McGeer, 2008, 249]

But if the only kind of trust that could reasonably be expected to inspire greater trust-responsiveness is trust that manifests confident belief in the trusted's current trust-responsiveness, then it is difficult to see how ther-

an honest man." The admonition surprises Valjean, "who had no recollection of ever having promised anything." McGeer and Pettit's discussion plausibly explains why Myriel was right; in voluntarily accepting the gift of the silver, Valjean implicitly makes just such a promise.) However, as I see it, this mechanism is not relevant to therapeutic trust, since it does not require the attitude of trust for its operation. It is the *en*trusting of the silver, the till, or the secret, etc., that generates the implicit promise and consequent obligation, whether or not it is accompanied by the attitude of trust. Valjean's accepting the gift would have amounted to promising even if Bishop Myriel had not genuinely trusted Valjean to use the silver to become a better man.

[15] Upon giving the silver, Bishop Myriel says, "Jean Valjean, my brother, you no longer belong to evil, but to good." [Hugo, 2008, Kindle location 2201]. One might take this as an explicit pronouncement of the sort of encouragement that is already demonstrated in his trusting Valjean with the silver. (This interpretation, admittedly, ignores Myriel's next claim—"I have bought your soul for God"—which suggests that other more controversial theological reasons may be at work. Thanks to Teresa Morgan for pointing this out.)

[16] Such positive effects may be contingent on the overall context and one's particular psychology. It may be that distrust can, in some circumstances, also have therapeutic effects, prompting the trustor to aim at gaining or winning back someone's trust. There may be room here for a notion of therapeutic distrust, when one manifests a distrustful attitude with the aim of trying to elicit more trust-responsiveness. (Cf. Tsai [2017] for a discussion of the potential therapeutic effects of showing disrespect.)

apeutic trust could serve as a reason for trust. Someone with good reason to think that the encouragement or esteem mechanisms would be operative will need to manifest confident belief to engage the mechanism. As we have seen, though, the evidence-only theses for belief and confidence suggest that only evidence of trust-responsiveness could serve as proper reasons to believe or boost one's confidence in someone's trust-responsiveness.

Of course, someone skilled at deception might attempt to secure the hoped for therapeutic effects by faking trust or by acting as if one's trust is stronger than it is. However, cases of therapeutic trust in the literature are intended to be both rational and morally admirable, and we will take for granted that any genuine case of rational therapeutic trust would not involve deception (including self-deception) regarding the strength of the attitude or the reasons for it. We will assume that any case of rational therapeutic trust would pass a version of Annette Baier's "expressibility test" for morally decent trust relations. It would survive the reflective awareness of both the trustor and trustee concerning the actual reasons on which the trust is based.[17] To communicate a strength of trust one lacks or to aim at adopting trust that one deems irrational would violate the expressibility requirement. Thus, in order for the mother-daughter or Myriel-Valjean cases to count as rational cases of therapeutic trust, their entrusting of the house, the silver, must communicate a genuine attitude of trust, the sincerity and inspirational effects of which would survive full disclosure of the reasons on which the trust is based.

Thus, we have a wrong kind of reasons challenge for rational therapeutic trust. It looks as if the sort of trust capable of inspiring trustworthiness requires confident belief in the trusted to prove trustworthy. But confident belief is just what the epistemic theses suggest cannot be adopted without evidence.

4 Self-Fulfilling Prophecy Cases of Therapeutic Trust

In this section, I argue that self-fulfilling prophecy cases make possible a successful reply to the wrong kind of reasons challenges. Some such cases are counterexamples to the claim that only evidence can properly boost confidence, related to the Jamesian theme that in some special circumstances a belief can "help to create its own fact." [James, 1896]. In such cases, the Therapeutic Thought can rationally boost confidence and even serve as a reason for belief, even though it is not evidence of trust-responsiveness.

To see this, let us propose some further details regarding the deliberations that led Bishop Myriel to trust Jean Valjean with the silver. Suppose, for example, that even though Bishop Myriel initially has a low confidence in the proposition, *If Valjean is given the silver he will use it to become a better man*, he also knows (or has good reason to believe) the following version of the Therapeutic Thought:

[17] [Baier, 1986]. See [McGeer, 2008] for a defense of the claim that Baier's expressibility test is also appropriate as a test of the rationality of trust.

> Therapeutic Thought: If I were to manifest confident trust in Valjean to use the silver to become a better man, Valjean will likely use the silver to become a better man.

We might suppose, as Victor Hugo's depiction of Bishop Myriel suggests, that Myriel has developed a keen insight into the psychology of people in Valjean's situation, gleaned from years of experience working with the poor and oppressed. His wisdom and insight allows him to predict the effects his trust would likely have on Valjean. (In other possible cases, the trustor may be a parent or mentor with more personal knowledge of the trusted's psychology.)

The Therapeutic Thought, in this case, is not an evidential reason for Myriel to think that Valjean will prove trustworthy. Evidential reasons, as we defined them, raise the likelihood of a proposition given one's total evidence. But at this time in his deliberation, Bishop Myriel's total evidence does not include that he has a high confidence that Valjean will prove trustworthy. He is only entitled to think that Valjean would prove trustworthy if he, Bishop Myriel, were to take a leap of faith beyond the current evidence.

Although the therapeutic thought is not an evidential reason for thinking that Valjean will prove trustworthy with the silver, neither is it epistemically irrelevant. Let us say that an *epistemic reason* is a reason that bears on whether some state is epistemically rational, where epistemic rationality is a matter of having beliefs or degrees of confidence that are normatively appropriate from the point of view of accurate representation. One who is epistemically rational does well with respect to the goal of believing things if and only if they are true, and the goal of having a confidence that fits one's total evidence. If the therapeutic thought is warranted for Bishop Myriel, he can see in advance that *if* he were to believe Valjean will prove trust-responsive, the belief would likely be accurate. He can also see in advance that if he were to boost his confidence in the proposition that Valjean will prove trustworthy, his confidence would fit the evidence he would then have (which would include his increased confidence). The therapeutic thought in this case is thus an epistemic as well as a practical reason; it is a consideration that supports boosting one's confidence in the proposition that the trusted will likely prove trustworthy, as an appropriate way of pursuing the goal of accurate representation.[18]

Thus, if one is warranted in thinking that manifesting trust in someone to ϕ will make it significantly more likely that they will prove trustworthy to ϕ, one will thereby have a non-evidential but epistemic reason to boost one's prior confidence that the trusted will prove trustworthy. If the increase would be enough for one's confidence to count as an outright belief, one will also have a non-evidential but epistemic reason to adopt doxastic trust.

[18] One might be skeptical that it is psychologically possible to bootstrap your way to trust in the way self-fulfilling prophecy cases require. However, as Sharadin [2016] (who argues against an evidence-only account of justified belief based on a similar scenario) points out, this is an empirical consideration, as yet untested, and there seems to be no non-question-begging philosophical reason to deny that such reasons can motivate.

5 Long-Term Cases and Non-Doxastic Trust

As we noted above, theorists who appeal to therapeutic trust cases have typically focused on long-term cases of therapeutic trust, such as Jones' home alone teenager case. Such cases, they contend, show that the Doxastic Account of Trust is false; trust can be adopted even in cases in which one does not have an outright belief that the trusted will prove trustworthy. Below we will see that such appeals fail to provide an adequate response to the wrong kind of reasons challenge targetting strength of trust, and we will see strong reasons to doubt that there are long-term cases of rational therapeutic trust. In the next sub-section, though, I argue that although the strategy that appeals to long term therapeutic trust is misguided, it is not misguided in rejecting the Doxastic Account of Trust.

5.1 In Defense of Non-Doxastic Trust

Some points that we have seen already suggest that there are non-doxastic cases of trust, *pace* Hieronymi's defense of the Doxastic Account of Trust. Recall Hieronymi's claim that if she found out that you were agnostic about her showing up to the restaurant, she could "rightly complain that your lack of confidence betrays a lack of trust." In reply, a defender of nondoxastic trust should acknowledge Hieronymi's point that your lack of confidence betrays *some* lack of trust. But why think that outright trust must be ideally strong? A natural rejoinder you could make to Hieronymi's complaint is that your confidence was strong enough: "I did trust you. My trust in you to show up was strong enough that I came to the restaurant. If I had not trusted you to show up, I would not have come."

This type of rejoinder, which highlights your trusting *enough* to take some salient risk, is even more compelling in some cases of trusting-with-X. Even if Bishop Myriel was not confident enough to believe outright that Valjean would use the gift of the silver to become a better man, he could defend his claim to have trusted Valjean with the silver by noting that he trusted Valjean enough to take an enormous risk in giving him the silver.

We can make the same point by considering a variation of the mother-daughter case. In the variation, we will take back Jones' stipulation that would make this a case of therapeutic trust, namely that the mother trusts the daughter with the house in order to inspire long-term trustworthiness. We will suppose that the mother is agnostic about whether the daughter will take good care of the house this weekend but is confident enough to be willing to take a risk on giving the house to the daughter. If the daughter questions her trust after the weekend is over, she could honestly make the following response:

> **Mother:** "I did not know whether or not you would prove trustworthy with the house. I was agnostic about that, but I trusted you to take care of the house anyway. I trusted you enough to risk giving it to you

> for the weekend without checking up on you. If I had not trusted you to take care of the house, I wouldn't have given it to you!"

Two points about the mother's response support the idea that this is a genuine case of non-doxastic trust. First, the second sentence conjoins a claim to trust with a claim not to believe that the trusted will prove trustworthy. If trust always involves outright belief, one would expect the second sentence to strike us as infelicitous, but in the context of the explanation there seems nothing linguistically odd about the mother's assertion. Second, the counterfactual in the last line is evidence that the mother entrusted the house to her daughter *because* she trusted her to care for it properly. She thus meets the criteria for trusting-with-X endorsed in section 2.1.

In Section 3.3, we worried that McGeer and Pettit's encouragement and esteem mechanisms might require doxastic trust in order to non-deceptively inspire greater trust-responsiveness. It seemed that one might need to manifest confident belief in the trusted in order to encourage or show esteem in a way that inspires. However, the present cases suggest that non-doxastic trust can, in principle, also show esteem and encourage in way that inspires long-term trustworthiness and passes the expressibility test. Even if he were to find out that Bishop Myriel was agnostic about whether he would use the silver to become an honest man, Valjean might be inspired by the thought that Myriel was confident enough to take the big risk of giving him the silver.

Likewise, we can imagine circumstances in which the daughter is inspired to be more trustworthy in the future by the fact that the mother was at least confident enough to risk being vulnerable, even though the daughter knew all along that the mother was not fully confident and did not believe outright that she would take good care of the house. In the Hallmark movie ending to this version of the case, we can imagine the daughter expressing gratitude for the mother's trust later in life:

> **Daughter:** "I know that you had your doubts and didn't really believe I'd come through, but I appreciate that you at least trusted me enough to risk giving me the house for the weekend anyway. Your trust inspired me. Thanks Mom."

Notice that even though the case as described is one in which non-doxastic trust has therapeutic effects, we have not yet vindicated this as an authentic case of rational therapeutic trust. In order for it to count as a case of therapeutic trust, we must add to the description that the mother trusts her daughter with the house based on a long-term therapeutic thought such as the following:

> Long-Term Therapeutic Thought: If I trust the daughter with the house, she will be more likely in the long-run to become trust-responsive.

Making sense of how the long-term therapeutic thought might serve as a reason for non-doxastic trust requires further clarification, and we will see that there are challenges to the idea that it can serve as a reason for trust.

Moreover, merely defending the possibility of non-doxastic trust is not, by itself, a sufficient response to the wrong kind of reasons challenge. Such a defense refutes the Doxastic Account of Trust, a central premise in one of the arguments of Section 3, but it leaves untouched the challenge framed in terms of strength of trust. What is needed is an account of what non-doxastic trust is and how exactly non-evidential considerations, such as the therapeutic thought, can serve as a reason for it. Does the therapeutic thought serve as a reason for non-doxastic trust by strengthening trust? If so, we owe a response to the wrong kind of reasons challenge that targets strength of trust. But if not, we need an account of how a consideration could serve as a reason for trust without strengthening trust.

In Sections 5.2 and 5.3, we will consider and criticize the two main proposals in the literature for saying what non-doxastic trust is and how the therapeutic thought might be thought to serve as a reason for it.

5.2 Two Accounts of Non-Doxastic Trust

What, then, is non-doxastic trust? In particular, what plays the cognitive role in cases of non-doxastic trust, substituting for belief? Two accounts have been prominent in the literature. A first proposal is defended by Jones in her discussion of the home alone teenager case:

> [The mother] is willing to accept vulnerability (e.g., complaints from the neighbors and a huge cleaning task) in the hope that a policy of trusting, consistently displayed, will bear fruit in the long run. The mother might have no expectation that the daughter will look after the house well—the past track record makes such predictive expectations unwarranted. But the mother does have normative expectations of the daughter that she look after the house well. Should the daughter fail to do so, she will respond with resentment and reproach; she will feel let down. [Jones, 2004, 16-7]

Jones identifies two relevant features that might be thought to substitute, in therapeutic cases, for belief or high confidence in the daughter to properly care for the house: 1) the mother's normative expectations that taking care of the house is what the daughter *should* do; and 2) the mother's willingness to embrace vulnerabilities and forego opportunities to lessen the vulnerabilities.[19]

The account sketched seems to take the following form:

> Adopted Vulnerabilities Account: A is trusting B to ϕ iff 1. A believes that B should ϕ; and 2. A willingly accepts vulnerability to losses if B does not ϕ.

The Adopted Vulnerabilities Account is non-expectational. Thus, Jones is committed to thinking that the mother can trust the daughter with the house even

[19] Cf. [Baier, 1986, 2013, McGeer, 2008]

when she is quite confident, or even believes, that the the daughter will not take proper care of the house on this occasion.[20]

A different and widely-held account suggests that the heart of non-doxastic trust is accepting that the trusted will come through, where acceptance is taken to be a cognitive attitude distinct from belief.[21] In developing the Acceptance Account, Karen Frost-Arnold borrows the notion of acceptance from Michael Bratman, who uses the following example to motivate the idea:

> The three of us need jointly to decide whether to build a house together. We agree to base our deliberations on the assumption that the total cost of the project will include the top of the estimated range offered by each of the sub-contractors. We facilitate our group deliberations and decisions by agreeing on a common framework of assumptions. We each accept these assumptions in this context, the context of our group's deliberations, even though it may well be that none of us believes these assumptions or accepts them in other, more individualistic contexts. [Bratman, 1992, 7 quoted in Frost-Arnold, 2014, 1964]

To accept a proposition is to resolve to use it as a premise in one's practical reasoning, within a limited context. Bratman argues that acceptance differs from belief in that it is dependent on practical and intellectual contexts, can be shaped by reasons other than a concern for the truth, can be directly and voluntarily chosen, and is not subject to demands for consistency across different contexts. In Bratman's example, we as a group accept that the total cost will be an amount at the top of the estimated range, in order to protect ourselves from dangers of going overbudget. When faced with group decisions that depend on the total cost—Do we have enough money to add an extra bathroom? What will be our tax liability?—we use the accepted proposition in our deliberations, even though none of us believes that it accurately represents the cost (nor even that it is the most accurate estimate available). Within the context of our deliberations, we "take offline" our beliefs about the cost and and use the accepted proposition.

This view of acceptance is incorporated into the following Acceptance Account of Trust:

> Acceptance Account of Trust: A trusts B to ϕ [in context C] iff A either believes or accepts that B will ϕ and this belief or acceptance is the basis of A's practical reasoning [in C]. [Frost-Arnold, 2014, 1964]

The Acceptance Account allows for doxastic and nondoxastic cases of trust. In non-doxastic cases, the trustor is said to accept, without believing, that a person will prove trustworthy. For example, the mother might trust the daughter to take care of the house by accepting, for purposes of practical reasoning, the proposition that the daughter will take good care of the house

[20] This is a reversal from earlier work in which she takes trust to involve an expectation of trustworthiness. See Jones [1996].

[21] [Alonso, 2016, Faulkner, 2014, Frost-Arnold, 2014, Hawley, 2014]. Cf. [Holton, 1994] for a related account in terms of "presupposing" rather than accepting.

this weekend, even though she does not believe this. When faced with decisions about which the daughter's care for the house is relevant — Should I call and check up on her? Should I take Monday off in case the house is in disarray? — the mother might take offline her beliefs about what the daughter will do and reason from the accepted proposition. "My daughter will take good care of the house," she might say to herself, "so there is no need to call to check on her or take Monday off."

The Acceptance Account comes in expectational and non-expectational versions. The non-expectational version holds that it is possible to accept that someone will ϕ, and thus trust them, even when one is quite confident or even takes oneself to know that they will not.[22] According to the more widely-held expectational version, accepting that someone will prove trust-responsive is compatible with agnosticism but not with outright *disbelief* concerning their proving trust-responsive.[23]

5.3 The Risk Constraint on Trust

The Adopted Vulnerabilities Account and the Acceptance Account share a common problem: both incorrectly imply that some cases of mere entrusting are genuine cases of trusting. The problem is especially apparent for versions of the views that are non-expectational, placing no epistemic constraints on acceptance. Below I will suggest that one case-in-point is the mother in Jones' version of the home-alone teenager example, but it will be helpful to focus first on a different example, involving Gandhi's non-violent resistance.

When Gandhi and his followers placed themselves in positions in which they expected the authorities to brutally enforce unjust laws—when they refused to defend themselves against the onslaught of excessive force—they acted in a way that meets the Adopted Vulnerability Account's criteria for trusting the authorities. To put some flesh on the case, consider one of Gandhi's earliest acts of non-violent resistance (powerfully depicted by a scene in Richard Attenborough's film), in which Gandhi publicly burns registration cards that Indians were forced to carry and that symbolized their second-class status under South African Apartheid. While placing cards in the fire, the leader of the police strikes Gandhi down with a baton. Gandhi struggles to get up and place more cards in the fire...and is met with another blow. The process repeats, until Gandhi can no longer move. Each effort to burn additional cards involved very deliberately adopting (and actively rejecting ways of avoiding) vulnerabilities of just the sort described by the Adopted Vulnerabilities Account. And, of course, Gandhi had normative expectations; he no doubt believed that the

[22] Facundo Alonso [2016] suggests that one can accept and thus rely on or trust someone to ϕ based *solely* on practical reasons, even when one *knows* that the person will not ϕ. (Alonso actually makes this claim about "reliance," but he maintains that reliance is necessary for trust and constitutes the cognitive component of trust.) Cf. Thompson [2017], who expresses sympathy for the view that trust, but not reliance, is non-expectational.

[23] Cf. [Holton, 1994, 71]; [Frost-Arnold, 2014, 1968]; [Faulkner, 2007, 316].

authorities *should* not use such excessive force. Gandhi thus meets the conditions put forward by the Adopted Vulnerabilities Account for trusting the authorities to care for his life and well-being.

Moreover, if there are no epistemic constraints on acceptance — if one can accept what one knows ain't so — we can stipulate for the example that Gandhi adopted the relevant vulnerabilities by accepting (for practical purposes, in the limited context of the protest) that the authorities would justly care for his life and well-being. Suppose that each time he deliberated about whether to consign another card to the fire, he took offline his belief that the police would probably beat him, resolving to act on the proposition that they would properly care for his life and well-being this time. (This might, perhaps, be a psychologically effective strategy for bracing oneself to adopt the extreme vulnerabilities required for non-violent resistance.) Given this stipulation about his psychology, Gandhi meets the conditions for trust articulated by the non-expectational version of the Acceptance Account.[24]

Both accounts thus seem committed to the claim that Gandhi was trusting the police with his life and well-being in acting as he did. However, this is a clear case of entrusting without trusting. In Section 2, we noted that in genuine cases of trusting someone with X, one entrusts X to them *because* one trusts them to care properly for X. It is clear from the example, though, that Gandhi did not entrust his life and well-being to authorities because he trusted them to care properly for these things. To the contrary, it was Gandhi's *distrust* of the officials to care for his life and well-being that was his reason for entrusting them. His goal in entrusting was to publicly expose their injustice.

Although Gandhi clearly did not trust the authorities with his life and well-being, there is a trust-relevant explanation of Gandhi's actions that involves trust in the authorities, and highlighting it will prove useful for analyzing proposed cases of long-term therapeutic trust. As Ryan Preston-Roedder [2013, 671] argues, Gandhi's commitment to nonviolence rested on a kind of faith in those in positions of power to be moved by their consciences (eventually) to repent of their injustice if presented with poignant, public examples of it. Athough Gandhi did not trust the authorities to act justly, he arguably did trust them to be capable of one day becoming trustworthy in this way, if given the right sort of experiences. The relevant distinction is one that McGeer [2008] helpfully describes as a difference between having a "first" and "second line of trust." The first line of trust involves trusting someone to ϕ. The second line of trust involves trusting them to one day become the sort of person who can properly be trusted to ϕ; it can serve as a fall-back position that warrants acts

[24] So far as I can see, the two accounts are also vulnerable to a counterexample based on the case (mentioned in Section 2.2) of a vindictive parent who believes the daughter will wreck the house but entrusts her with it in order to shame her for her untrustworthiness. The goal of shaming the daughter might lead the parent to adopt the relevant vulnerabilities and to ground her weekend deliberations on the assumption (which she believes to be false) that the daughter will take good care of the house. The parent in this case clearly does not genuinely trust the daughter with the house.

of entrusting when the first line of trust is weak (or even, as in Gandhi's case, when one distrusts).

With this distinction in hand, consider the following hypotheses:

> *First Line of Trust Hypothesis:* Gandhi entrusted his life and well-being to the authorities because he trusted them to properly care for his life and well-being.
> *Second Line of Trust Hypothesis:* Gandhi entrusted his life and well-being to the authorities because he trusted them to become the sort of people who can properly be trusted with his life and wellbeing (if given the right sort of experiences).

Although the first line of trust hypothesis is false, the second line of trust explanation of Gandhi's actions is, arguably, deep and important. It plays a crucial role in explaining why Gandhi and his followers were willing to place themselves in circumstances in which they expected the authorities to enforce unjust laws, and why they often refused to defend themselves against the onslaught of excessive force. Without trust in oppressive authorities to be capable of changing for the better, non-violent forms of resistance can seem pointless. No surprise, then, that debates about whether resistance should take a non-violent form often focus on whether the second line of trust is warranted.[25] However, we should not confuse the plausible claim that Gandhi had the second line of trust in the authorities (trusting them to one day become the sort of people who can be properly trusted with his life and wellbeing) with the implausible claim, which is implied by the non-expectational version of the Acceptance Account and the Adopted Vulnerabilities Account, that Gandhi had the first line of trust (trusting the authorities with his life and well-being).

A similar analysis applies to Jones' version of the mother-daughter case, in which the mother believes the daughter will not take good care of the house. Suppose that the mother engages in the following deliberation, which, according to the accounts under consideration, should result in her trusting the daughter with the house for long-term therapeutic reasons (and which she helpfully frames using the terminology of the accounts):

> **Mother:** "I know my daughter will not take good care of the house if I entrust her with it. However, it would be good for her long-term moral development if I trust her to care for it. After she wrecks the house, I will be able to point out that she let me down, even though I *trusted* her with the house. The regret she feels will hopefully prompt her to be more trustworthy in the future. It is worth the cleanup, neighborhood complaints, etc. to do this for her. So, I will accept that she *will* take

[25] Further, the second-line of trust suggests a plausible explanation of a key advantage often touted by defenders of non-violent resistance—that non-violence makes it more likely to secure justice and a lasting peace. One reason for this advantage may well involve the way that manifesting the second line of trust in someone shows respect for them as a person capable of changing for the better.

> care of the house, using this proposition in my practical reasoning all this weekend, and I will freely adopt the vulnerabilities that result."

I submit that if these reasons for entrusting the house became known to the daughter, she could properly complain that the mother did not genuinely trust her with the house. Notice that the case is crucially different from the version of the mother-daughter case given as an example of non-doxastic trust in Section 5.1. In that case, we supposed that the mother was agnostic but willing, in entrusting the house to the daughter, to take a risk on the daughter's proving trustworthy with the house, a risk she would no doubt have regretted taking if the daughter had wrecked the house. In the present case, by contrast, the mother preferred to entrust the house to the daughter regardless of whether the daughter proved trustworthy (indeed, even though she believed the daughter would not prove trustworthy).

The discussion here suggests an important general constraint on actions that manifest the attitude of trust, a constraint that is not satisfied in cases of mere entrusting. Actions that manifest trusting someone to ϕ involve taking a particular kind of practical risk on someone's proving trustworthy, a risk one takes when one would prefer to perform the action if the person proves trustworthy to ϕ, but not otherwise. Thus, we have the following Risk Constraint on Trust:

> Risk Constraint on Trusting-to: An action involves A's trusting B to ϕ only if the action involves taking a practical risk on B's proving trustworthy to ϕ (that is, A prefers to perform the action if B proves trustworthy to ϕ, but not otherwise).

Two points about the concept of practical risk at play in the Risk Constraint will be germane to our discussion.[26] First, this kind of risk is practical as opposed to epistemic. One can take a practical risk on someone's proving trustworthy even if one is completely confident that the other will prove trustworthy. Indeed, one who is very confident will typically be willing to take very big practical risks on the trusted's proving trustworthy, performing actions that would have disastrous consequences if the trusted fails to come through. In another, epistemic sense, the trustor will not regard these actions as risky, being fully confident that the trusted will not let her down. (Cf. [Pettit, 1995, 208])

Second, taking a practical risk in this sense requires more than merely making oneself vulnerable to costs that would be incurred if the person fails to prove trust-responsive. There is a natural but weaker sense in which Gandhi risks being beaten when he puts a card in the fire, and the mother in Jones' case risks damages to the house and to her reputation among the neighbors

[26] In her account of faith, Lara Buchak argues that acting on faith that a proposition X is true requires taking this kind of practical risk on X's being true. Buchak gives a more formal definition of taking a risk on the truth of a proposition: "[A]n act [A] constitutes an individual's taking a risk on X just in case for some alternative act B, A is preferred to B under the supposition that X, and B is preferred to A under the supposition that $\overline{X}$." [Buchak, 2014, 54]; cf. [Buchak, 2012].

when she entrusts the house to the daughter. But neither takes a practical risk in the stronger sense of performing an action that they prefer to do if the entrusted proves trustworthy *but not otherwise*. Even on the supposition that the person would fail to prove trustworthy, each still preferred, all things considered, to perform the entrusting actions and suffer the consequences.

So far, our criticisms have not explicitly targetted the expectational version of the Acceptance Account, according to which acceptance is compatible only with agnosticism and not outright disbelief concerning whether someone will prove trustworthy. Yet this theory, too, violates the Risk Constraint. To better see this, consider another variation of the mother-daughter case. Suppose that the mother is agnostic about whether the daughter will prove trustworthy but entrusts the daughter with the house based solely on the following dominance reasoning (which, according to all versions of the Acceptance Account, should result in a long-term therapeutic case in which the mother trusts the daughter with the house):

> **Mother:** I am agnostic about whether my daughter will prove trustworthy with the house if I give it to her this weekend. But whether or not she proves trustworthy, it would be better for me to trust her to take care of the house. Maybe she will take good care of the house, and if so, great. But if she wrecks the house, it will be good for her long-term moral development if I trust her to care for it. (I will be able to point out that she let me down even though I *trusted* her with the house, and perhaps the regret she feels will make her more trustworthy in the future.) It would be worth the cleanup, neighborhood complaints, etc. to do this for her. So, I will accept (for purposes of practical reasoning in this limited context), that she will take care of the house.

Again, if these reasons became known to the daughter, she could properly complain that the mother was not genuinely trusting her with the house. The mother decided to entrust the daughter with the house based on the idea that it would be preferable to do so regardless of whether she proves trust-responsive. Despite appearances, her entrusting the house did not involve taking the relevant kind of practical risk on the daughter to care for the house. Further, the mother fails the counterfactual test that, in the case of the agnostic mother from Section 5.1, served as evidence that she entrusted the house because she trusted the daughter to care for it. She could not honestly claim, "If I had not trusted you to take care of the house, I wouldn't have given it to you;" she would have given the house regardless of how much she trusted the daughter to care for it.[27]

[27] Might this be a case of overdetermination, in which the mother entrusted her daughter both because of the dominance reasoning but also because she trusted her to care for the house? In some cases this may be plausible, but not all. We can stipulate that the mother in this case suffers from weakness of will, so that she would not be willing to give the house to the daughter if it meant taking even a slight practical risk. In such a case her entrusting the house clearly does not manifest trust, although the Acceptance Account implies that it does.

None of this is intended to disparage either the mother's parenting in these cases or the important role that the second line of trust might play in educating for trust-responsiveness. As with the Gandhi case, we can distinguish first and second line of trust hypotheses intended to explain why the mother entrusted the daughter with the house:

> *First Line of Trust Hypothesis:* The mother entrusted the daughter with the house because she trusted her to take proper care of the house.
> *Second Line of Trust Hypothesis:* The mother entrusted the daughter with the house because she trusted her to one day become the sort of person who can properly be trusted to take care of the house, if given the right moral opportunities.

The first line of trust hypothesis must be true in order for the mother to count as trusting the daughter with the house, and I've argued that it is false in Jones' case and the one just considered. However, the second line of trust hypotheses is, arguably, true and important in these cases. It's plausible to suppose that, in entrusting the house to the daughter, the mother was trusting the daughter *to one day become someone who can be trusted with the house*, taking a practical risk on the daughter's moral development that satisfies the Risk Constraint's condition for manifesting the second line of trust.

Further, even if it is common knowledge that the mother does not trust the daughter with the house but merely *en*trusts her with it, the second line of trust manisted in her entrusting may yet inspire in a way that passes Baier's expressibility test. One can imagine the mother forthrightly explaining her reasons for entrusting the house and the daughter responding positively.

(In the Hallmark movie ending to this version of the case, the daughter later expresses her gratitude:

> **Daughter:** "I know that you did not really trust me to take good care of the house. But I appreciate that you entrusted me with the house anyway, because you trusted me to one day become the sort of person who is trustworthy. Your trust inspired me. Thanks Mom.")

However, notice two points about the mother's second line of trust. First, the second line of trust plausibly involves a belief or high confidence in the daughter to one day become the sort of person who is trustworthy; it is doubtful that non-doxastic trust would have these inspirational effects. Second, the mother's second line of trust is not a case of long-term therapeutic trust. To construe it as a case of therapeutic trust requires adding the stipulation that the mother trusted her daughter to one day become trustworthy in order to inspire her to one day become trustworthy. But adding the stipulation would make this a self-fulfilling prophecy type of case. Thus, we do not here have a vindication of the possibility of long-term therapeutic trust.

6 Trust as Confidence-Enough-To-Risk

We have criticized the Adopted Vulnerabilities and Acceptance Accounts of trust on the grounds that neither does justice to the Risk Constraint on Trust.

Trusting requires being willing to take practical risks on someone's proving trustworthy, acting in ways one prefers to act if the trusted proves trustworthy, but not otherwise. In our defense of nondoxastic trust (Section 5.1), we noted that it is possible to be agnostic about someone's proving trustworthy and yet trust the person *enough* to take salient risks on their doing so. However, in the limiting case in which placing something into someone's care does not involve taking a practical risk (because one would prefer to give it whether or not they prove trustworthy), entrusting the person does not manifest trust.

A more promising account identifies trust with having a confidence sufficient to motivate one to take salient practical risks on someone's proving trust-responsive. On such a view, the attitude of trust consists in having confidence in someone to prove trustworthy, and such confidence might in some cases fall short of outright belief. What it must do, however, is ground a willingness on the part of the trustor to take some practical risks on the trusted's proving trustworthy.[28]

On this view, in order for your entrusting someone with X to count as trusting them with X, two conditions must be met. First, entrusting X must satisfy the Risk Constraint; it must be an action that you prefer to take if the trusted proves trustworthy to care properly for X, but not otherwise. Second, your confidence in the person to properly care for X must be sufficient to motivate you to take the risk of entrusting X. More generally, we can give the following account of actions that manifest trust in someone to ϕ:

> Confident-Enough-To-Risk Account of Trusting: An action involves A trusting B to ϕ only if a) the action involves taking a practical risk on B's proving trustworthy to ϕ (that is, A prefers to perform the action if B proves trustworthy to ϕ, but not otherwise); and b) A performs the action because she is confident enough that B will prove trustworthy to ϕ.[29]

The account suggests that there are two distinct kinds of reasons for trusting: *confidence-boosting* and *licensing* reasons. Confidence-boosting reasons are reasons that should rationally lead one to become more confident that the trusted will prove trustworthy. The central kind of confidence-boosting reason involves evidence of trust-responsiveness. Such evidence not only makes it epistemically rational to raise one's confidence, it also makes it practically rational to take greater risks on the person's trustworthiness. If the considerations of Section 4 are correct, some self-fulfilling prophecy cases of therapeutic trust may involve a distinct, non-evidential kind of confidence-boosting reason. One

[28] Recall from Section 2.3 that cases of trusting differ from cases of merely relying in that they involve some kind of expectation that the person will not just ϕ but will *prove trustworthy* in doing so. As I see it, this expectation is part of the content of the confidence one has when one trusts, although it may be possible to develop a view that separates out one's confidence that someone will ϕ and further attitudes that constitute an expectation that they will prove trustworthy in doing so. Thanks to an anonymous referee for prompting me to clarify this.

[29] See [Gambetta, 1988] for an early account that identifies trust with being confident enough to take risks.

can rationally boost one's confidence in someone to prove trustworthy to ϕ based on reasons to think that one's trust in them would make it more likely that they will prove trustworthy.

Licensing reasons, by contrast, are considerations that suggest that the likelihood of someone's proving trustworthy is *enough* to motivate taking certain risks on their proving trustworthy.[30] Licensing reasons do not rationally boost one's confidence that someone will prove trustworthy; rather, they suggest that the confidence one already has is sufficient to license taking practical risks on the person's doing so.

The account fits well with the common observation that trust seems to depend on the potential cost of misplaced trust in one's practical context. As Frost-Arnold points out,

> I may trust my friend not to tell my secret when exposure would do me little harm, while not trusting her with the same secret in an environment where revelation would seriously damage my reputation. [Frost-Arnold, 2014, 1966]

In a similar vein, Katherine Hawley suggests that deliberating about whether to trust depends on practical stakes, with more evidence of trustworthiness required when misplaced trust would not be so costly:

> Can you trust the newspaper's claim that the weather will be sunny all week? Yes, if you're just trying to decide whether to hang out the washing to dry. No, if you're deciding whether to take out insurance for an expensive outdoor wedding....When the stakes are low, we don't need much evidence; when there's more at stake, we need more. [Hawley, 2012, 5]

One type of licensing reason, suggested by these examples, will involve considerations that show that in one's practical context there is less practical risk to trusting, because the cost of misplaced trust would be less than one initially thought. Circumstances might change so that one now knows that one's secret, even if exposed, will not damage one's reputation as much as one initially thought. Or, riffing on Holton's shopkeeper example, you might learn that there is less money in the till than you first thought. Whereas your confidence in the employee not to steal might not be great enough for you to be willing to risk $10,000, it might be enough now that you know there is only $100 there. This consideration might lead you to trust the employee with the till without increasing your confidence in the employee.

Notice that the licensing reasons in these cases *do not* strengthen trust. Rather, they suggests that the strength of trust one already has is sufficient for taking the risk. For our purposes, this point has two important consequences. First, it allows us to see how licensing reasons of this sort can be non-evidential while not running afoul of the wrong kind of reasons challenge for strength of trust. One might accept the Evidence-Only Thesis for Strength

[30] I borrow the term from Adrienne Martin [2013], who employs the term in her discussion of reasons for hope.

of Trust, holding that only evidence of trust-responsiveness can serve as reasons that strengthen trust, while still holding that one can trust based in part on licensing reasons.

Second, although licensing reasons challenge the view that only evidence can serve as reasons for trust, they also cast doubt on whether there are rational cases of long-term therapeutic trust. For, if the long-term therapeutic thought is to serve as a licensing reason, it will do so by showing that trusting someone is less risky than one might have thought. The long-term therapeutic thought might lead one to expect that the short-term costs associated with misplaced trust would be compensated by a long-term good that the trustor values: the trusted's developing greater trust-responsiveness. The overall cost of misplaced trust to the trustor will thus be less. (In the counterexamples we presented to the Adopted Vulnerabilities and Acceptance Accounts, the considerations show that entrusting someone with something of value would be preferable even if the person fails to prove trustworthy, and so does not require any degree of trust.)

The problem is that these considerations threaten to shortcircuit the mechanisms by which trust might reasonably be expected to inspire greater trustworthiness, lessening the encouragement and esteem that could sincerely be conveyed by one's trust. Both the encouragement and esteem mechanisms work by conveying something about one's confidence—either that one has a confident belief that the trusted will prove trustworthy, or that one is at least confident enough to take a serious practical risk on the trusted's proving trust-responsive. To the extent that one's entrusting is motivated by a desire to produce long-term therapeutic effects, though, the act of entrusting will not manifest the strength of trust that one had hoped would inspire.

Again, this is not to say that there are never therapeutic goods associated with entrusting someone with X when one does not trust them with X. As we saw above, the daughter in Jones' case might be inspired by the second line of trust that the mother manifests in entrusting the house to her. Further, the daughter might draw moral inspiration from the mother's willingness to sacrifice her own self-interest for the sake of the daughter's moral good. The mother's entrusting displays an admirable ordering of values, which places the daughter's long-term moral development above being able to avoid the inconveniences of a wrecked house. The daughter might, ironically, be inspired by the idea that the mother did not regard entrusting her with the house as involving a big practical risk and so did not regard the situation as one requiring strong trust. This point, while true and important, suggest that attitudes other than trust can have therapeutic effects, and an adequate moral psychology should distinguish this source of moral inspiration from the very different cases in which trust inspires.

Acknowledgements For helpful feedback and discussion, thanks to Dan McKaughan, Teresa Morgan, and two anonymous referees for this journal. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions

expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

References

Facundo M. Alonso. Reasons for reliance. *Ethics*, 126(2):311–338, 2016.

William P Alston. The deontological conception of epistemic justification. *Philosophical Perspectives*, 2:257–299, 1988.

Annette Baier. Trust and antitrust. *Ethics*, 96(2):231–260, 1986.

Annette Baier. What is trust? In David Archard, Monique Deveaux, Neil Manson, and Daniel Weinstock, editors, *Reading Onora O'Neill*, pages 175–185. Routledge, New York, 2013.

Judith Baker. Trust and rationality. *Pacific Philosophical Quarterly*, 68(1):1, 1987.

Michael E. Bratman. Practical reasoning and acceptance in a context. *Mind*, 101(401):1–16, 1992. ISSN 401.

Lara Buchak. Can it be rational to have faith? In Jacob Chandler and Victoria Harrison, editors, *Probability in the philosophy of religion*. Oxford University Press, Oxford, 2012.

Lara Buchak. Rational faith and justified belief. In Timothy O'Connor and Laura Frances Callahan, editors, *Religious faith and intellectual virtue*, pages 49–73. Oxford University Press, 2014.

Jacopo Domenicucci and Richard Holton. Trust as a Two-Place Relation. In Paul Faulkner and Thomas W. Simpson, editors, *The Philosophy of Trust*, pages 149–160. Oxford University Press, 2017.

Jeremy Fantl and Matthew McGrath. *Knowledge in an Uncertain World*. Oxford University Press, Oxford, 2009.

Paul Faulkner. A genealogy of trust. *Episteme*, 4(3):305–321, 2007.

Paul Faulkner. The practical rationality of trust. *Synthese*, 191(9):1–15, 2014.

Paul Faulkner. The attitude of trust is basic. *Analysis*, 75(3):424–429, 2015.

Richard Feldman. Reasonable religious disagreements. In *Philosophers without gods: Meditations on atheism and the secular*, pages 194–214. Oxford University Press, Oxford, 2007.

Karen Frost-Arnold. The cognitive attitude of rational trust. *Synthese*, 191 (9):1957–1974, 2014.

Diego Gambetta. Can we trust trust? In Diego Gambetta, editor, *Trust: Making and breaking cooperative relations*, pages 213–237. Blackwell, Oxford, 1988.

Russell Hardin. *Trust and trustworthiness*. Russell Sage Foundation, April 2004.

Katherine Hawley. *Trust: A Very Short Introduction*. Oxford University Press, Oxford, 1 edition edition, September 2012. ISBN 978-0-19-969734-2.

Katherine Hawley. Partiality and prejudice in trusting. *Synthese*, 191(9): 2029–2045, 2014.

Pamela Hieronymi. The reasons of trust. *Australasian Journal of Philosophy*, 86(2):213 – 236, 2008.

Richard Holton. Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72(1):63 – 76, 1994.

H. J. N. Horsburgh. The ethics of trust. *Philosophical Quarterly*, 10(41): 343–354, 1960.

Victor Hugo. *Les Misérables*. 2008. URL `https://www.gutenberg.org/ebooks/135`.

William James. *The will to believe*. Longmans, Green, 1896.

Karen Jones. Trust as an affective attitude. *Ethics*, 107(1):4–25, 1996.

Karen Jones. Trust and terror. In Margaret Urban Walker and Peggy DesAutels, editors, *Moral psychology : Feminist ethics and social theory*, Feminist constructions. Rowman & Littlefield Publishers, Lanham, MD, 2004.

Thomas Kelly. Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research*, 66(3):612–640, 2003.

Adrienne Martin. *How we hope: A moral psychology*. Princeton University Press, Princeton, NJ, 2013.

Berislav Marušić. *Evidence and agency: Norms of belief for promising and resolving*. Oxford University Press, Oxford, 1 edition edition, November 2015.

Berislav Marušić. Trust, reliance and the participant stance. *Philosophers' Imprint*, 17, 2017.

Victoria McGeer. Trust, hope and empowerment. *Australasian Journal of Philosophy*, 86(2):237 – 254, 2008.

Victoria McGeer and Philip Pettit. The empowering theory of trust. In Paul Faulkner and Thomas W. Simpson, editors, *The Philosophy of Trust*, pages 14–34. Oxford University Press, 2017.

Carolyn McLeod. Trust, 2015. URL `https://plato.stanford.edu/archives/fall2015/entries/trust/`.

Benjamin McMyler. *Testimony, trust, and authority*. Oxford University Press, Oxford, 2011.

Richard Moran. Getting told and being believed. *Philosophers' Imprint*, 5(5): 1–29, 2005.

Michael Pace. The epistemic relevance of moral considerations: Justification, moral encroachment, and James' "The will to believe". *Noûs*, 49(2):239–268, 2011.

Philip Pettit. The cunning of trust. *Philosophy and Public Affairs*, 24(3): 202–225, 1995.

Ryan Preston-Roedder. Faith in humanity. *Philosophy and Phenomenological Research*, 87(3):664–687, 2013.

Nishi Shah and J. David Velleman. Doxastic deliberation. *Philosophical Review*, 114(4):497–534, 2005.

Nathaniel P. Sharadin. Nothing but the evidential considerations? *Australasian Journal of Philosophy*, 94(2):343–361, 2016.

Thomas W. Simpson. Trust, belief, and the second-personal. *Australasian Journal of Philosophy*, 96(3):447–459, 2018.

Kurt Sylvan. Epistemic reasons I: Normativity. *Philosophy Compass*, 11(7): 364–376, 2016. doi: 10.1111/phc3.12327.

Christopher Thompson. Trust without reliance. *Ethical Theory and Moral Practice*, 20(3):643–655, 2017.

George Tsai. Respect and the efficacy of blame. In David Shoemaker, editor, *Oxford studies in agency and responsibility*, volume 4, pages 248–275. Oxford University Press, Oxford, 2017.