

## Trust and the Trickster Problem

Pre-print draft

Abstract:

In this paper, I articulate and defend a conception of trust that solves what I call “the trickster problem.” The problem results from the fact that many accounts of trust treat it similar to, or identical with, relying on someone’s good will. But a trickster could rely on your good will to get you to go along with his scheme, without trusting you to do so. Recent philosophical accounts of trust aim to characterize what it is for one person to trust another so as to avoid this problem, but no extant account successfully does so.

I argue that connecting trust to important, normatively defined relationships like friendship, romantic partnerships and parenting shows us something important about trust. The clearest cases of trust are found within the confines of normatively defined relationships like these, suggesting that there is a normative element to trust. Trusting someone involves not just believing that another person’s good will covers your interactions. Trusting involves believing that, at least in a certain domain of interaction, you are *entitled* to rely on that person’s good will. This account solves the trickster problem, because a trickster is not entitled to his victim’s good will.

“Most successful business executives intertwine their personal and professional lives. But those two strands of Mr. Madoff’s life were practically inseparable. He sometimes used his 55-foot fishing boat, Bull, as a floating entertainment center for clients. He used his support of organizations like the Public Theater in Manhattan and the Special Olympics to build a network of trust that began to stretch wider and deeper into the Jewish community.”

"Trust ... [is] probably the most difficult hurdle."  
--Elizabeth Edwards

In this paper, I articulate and defend a conception of trust that solves what I call “the trickster problem.” The problem results from the fact that many accounts of trust treat it similar to, or identical with, relying on someone’s good will. But a trickster like Bernie Madoff could rely on your good will to get you to go along with his scheme, without trusting you to do so.<sup>1</sup> Therefore, any account of trust that treats trust as relying on another’s good will is subject to the trickster problem. Recent philosophical accounts

of trust aim to characterize what it is for one person to trust another so as to avoid this problem, but no extant account successfully does so. One reason for this is that while theorists interested in trust have standardly used the phenomenon of feeling betrayed as a test for when trust is present—because it is assumed that when your trust is violated, feelings of betrayal will be warranted—no one has examined the phenomenon of betrayal itself. So while it is generally agreed that feeling betrayed is fundamentally different from feeling let down or disappointed, there has been little attention to the distinction itself. Clarifying the distinction between feelings of betrayal and feelings of disappointment will point us in the direction of an account of trust that can successfully deal with the trickster problem.

I'll begin in the first section by laying out some commonly accepted ideas about trust—its relation to the ideas of reliance and good will and also the idea that a violation of trust licenses feelings of betrayal. I'll then introduce Karen Jones' account of trust, which I regard as the best account in the current literature. Despite the attraction of her account, however, I'll demonstrate that it is importantly flawed in identifying trust with an affective attitude, specifically, hope. Jones' attempt to analogize trusting to hoping leads her account to treat people as trusting others when there is clearly no trust. I show, instead, that trusting someone is better understood as a type of belief.

In the second section, I show that even this improved account of trust cannot get the right answer to the trickster problem. It cannot say, correctly, that Madoff does not trust his victim to do as he wants. I then turn to an attempt to solve the trickster problem. Pamela Hieronymi and Richard Holton have argued that understanding trust as relying on someone from the participant stance can deal with the trickster worry. However, given

plausible assumptions about what the participant stance involves, this idea does not solve the problem.

In the third section, I examine the phenomenon of feeling betrayed more closely and develop an account that understands the unique character of feeling betrayed in terms of feeling that an important relationship has been ruptured or violated. Feeling disappointed, I claim, lacks this content. If this is so, then it makes sense to connect the account of trust on offer with important, normatively defined relationships like friendship, romantic partnerships and parenting. The clearest cases of trust are found within the confines of relationships like these, suggesting that there is a normative element to trust. Trusting someone involves not just believing that another person's good will covers your interactions. Trusting involves believing that, at least in a certain domain of interaction, you are entitled to rely on that person's good will, that the other party owes you good will. This account solves the trickster problem, because a trickster like Bernie Madoff is not entitled to his victim's good will. It is his realization of this fact that explains both why Madoff does not trust his victim to go along with his scheme and why it would be inappropriate for him to feel betrayed if his victim does not do as he wishes.

## **I. Trusting**

### **A. The Basic Model**

Let me say a bit more about the area I mean to investigate, which is complicated by the fact that the word 'trust' is used on different occasions to mean very different things. English allows both of the following usages: "We trust you have enjoyed your flight with

Air New Zealand," and "Othello's trust in Iago was misplaced."<sup>2</sup> In the first instance, 'trust' is being used to mean something like 'expect' or 'hope.' In the second, 'trust' picks out interpersonal trust, which is my target in this paper. I won't, then, be directly concerned to account for usages of 'trust' that deviate from the interpersonal phenomenon. Examples of the kind of trust I have in mind include the trust that Bernie Madoff's investors had in him as their friend and financial advisor and the trust that Elizabeth Edwards previously had toward her husband John, before his affair was revealed.<sup>3</sup>

As a first gloss, then, what is it for one person to trust another in the way that Madoff's clients trusted him and Elizabeth Edwards trusted John? Philosophers interested in trust have standardly answered this question by following Annette Baier in grafting their accounts of trust onto a characterization of another attitude: reliance.<sup>4</sup> Reliance is a mental attitude characterized by an expectation that something will happen. As such, it is an attitude that we can have toward other human beings, but also toward inanimate objects, plants, animals, natural events, etc. You rely on the grocery store to have food stocked and the bus to be on time. You also rely on your computer to start up when you push the power button. Insofar as Madoff's clients trusted him, they relied on him and insofar as Elizabeth once trusted John, she relied on him as well.

Trusting someone, however, is thought to involve more than simply relying on that person. Baier notes that "The comedian, the advertiser, the blackmailer, the kidnapper-extortioner, and the terrorist" (Baier 234-5) all rely on the other people with whom they interact to do certain things, but cannot be said to trust them. Part of our reason for saying that these people do not trust those they rely on has to do with what

responses make sense for them if their reliance is disappointed. They may be annoyed, angry, or perhaps surprised, in a similar manner to how you might feel if your computer did not turn on when you wanted to use it. In all these cases the person in question feels what we would describe as disappointment, or perhaps annoyance, but does not feel betrayed, as would be appropriate if a trust had been violated. So it is generally assumed that an important test for the presence of trust in a given situation is whether a person would reasonably feel betrayed.

Why can trust be betrayed, but not reliance? The canonical answer, again due to Baier, is that trust involves relying on the good will of another person, as opposed to relying on that person's "dependable habits," where these habits could involve ill will or other kinds of motives distinct from good will. An advertiser may expect to see sales for a product increase as the result of the typical formula of putting young, attractive people in the product shoot, but a customer does not betray her if the customer does not respond to her ad in the expected way. This is because the advertiser's reliance on the customer's reaction to the product pitch does not depend on the customer's good will, but on the dependable transference of the customer's attraction toward the models onto the product.

The idea that trust involves relying on someone's good will seems intuitive and has been adopted by most theorists of trust. Trust is also thought to involve assumptions about the other person's competence. As Karen Jones notes, "some people have very good will but very little competence, and the incompetent deserve our trust almost as little as the malicious" (Jones 6-7). It is generally agreed, then, that to trust someone involves relying on her good will and competence.<sup>5</sup>

## B. Jones' Improvement

Baier's account of trust as reliance on a person's good will doesn't get us far enough, however. General moral claims require that people bear us some level of good will. If we anticipate they will act as morality demands, we might reasonably rely on their good will. But trusting is more selective. Trust requires a more personal connection with a person than simply expecting that the person's good will covers your interactions. Trusting someone also involves assumptions about that person's motivations—their reasons for acting toward you with good will and what that implies about the circumstances in which you can expect it.

Karen Jones' account of trust captures this idea. Jones understands trust as

[A]n attitude of optimism that the good will and competence of another will extend to cover the domain of our interaction with her, together with the expectation that the one trusted will be directly and favorably moved by the thought that we are counting on her. (4)

The idea that the person trusted is being directly and favorably moved by the thought that the one trusting is counting on her should be understood as the person trusted giving significant weight to the fact that she is being counted on. It is not required that this motivation trump all other considerations—depending on the circumstances, other considerations may carry the day. However, as Jones notes,

one would not trust if one thought that the fact that one was counting on

someone, while always being taken into account, would nonetheless be reliably overridden by other considerations. Were that the case, then, from the point of view of the trustee, the other would appear unwilling to give enough weight to the thought that she was counting on her. (8)

What this reflects is that the person doing the trusting expects that the trustee will be moved by the fact that the truster is counting on her and not simply by a generally benevolent motivation toward all people, for example. Jones' account captures the idea that trusting someone involves taking that person to be motivated in a particular manner, namely, by the fact that the one trusting is counting on her.

### C. The Affective Problem with Jones' Proposal

My points of disagreement with Jones' account of trust concern the nature of the attitude that is directed at the good will and competence of another person when you trust someone and what that implies about when we can sensibly say that one person trusts another. All parties to the debate believe that trust involves an attitude toward the good will and competence of someone else; Jones thinks that trust requires that attitude to be an *affective attitude of optimism*. She believes this because she wants her account of trust to be able to account for the fact that when we trust someone we have a tendency to interpret that person's behavior in a positive light. For example, Jones mentions the fact that in Shakespeare's play, *Othello*, Othello's trust in Iago makes Othello unable to perceive that Iago is attempting to harm him, even though the signs that Iago presents a danger are obvious for someone not already heavily biased in Iago's favor.

Jones believes that this tendency to look on the bright side can best be accounted for by an affective attitude, and that in this sense “trusting is more like hoping than like predicting” (Jones 15). Jones wants to draw an analogy between her account of trust, which is attempting to capture our tendency to positively interpret others, and accounts of the emotions that emphasize emotional states as being partly constituted by tendencies of interpretation.<sup>6</sup> For example, Jones claims that the object of resentment might be thought to be a “manipulative exploiter.” If this is correct, then if a person already resents her neighbor, it will be hard to get that person to attend to evidence that her neighbor is not a manipulative exploiter.

This is insufficient evidence, however, to support the conclusion Jones wants to draw about the similarity between trust and hope. It is true that trusting someone can result in a tendency to discount evidence of behavior that would undermine trust, but there is no need to posit that an attitude like hope is involved in order to generate this psychological result. This is because the kind of bias involved in trust—a bias in favor of the person trusted—can be accounted for by the fact that biases in judgment can occur as the result of a person holding either cognitive or affective attitudes.<sup>7</sup> For example, the *halo effect* is a cognitive bias that tends to make perceptions of specific traits in an object or person fit previous global evaluations of that same object or person. Finding that a person is likable, in general, tends to make you perceive him as having other attractive qualities. Finding someone to be physically attractive may lead you to also perceive him as kind, or finding someone to be courageous might also lead you to perceive her as having good judgment.<sup>8</sup> The halo effect, or some other cognitive bias, could account for Othello’s inability to perceive Iago’s machinations without presuming that Othello’s

attitude toward Iago is characterized by hope. It is equally plausible that another attitude, like a belief, can also support Othello's trusting bias.

Jones' affective proposal is also problematic because the account leaves open the possibility that you could trust someone when you hope they will do something but do not believe that they will. But this is false. As Hieronymi points out, "to the extent that you lack confidence in a person's trustworthiness in some matter, to that extent it also seems correct to say that you do *not* trust that person in that matter" (214). To trust someone, you must have more confidence in them than is captured by saying that you hope that person's good will and competence will govern your interactions.

We can see this point on display by considering Elizabeth Edwards' attitudes toward John after his infidelity is revealed. Given the nature of his work, John travels frequently and spends a significant amount of time away from Elizabeth. During those time periods, Elizabeth has to decide how to react to their newly impaired relationship. Suppose that there are really only two viable options: to try to go on much as before, throwing herself into her own work, or to try to obsessively monitor John's actions from afar by calling him and his aides, persistently asking for updates to try to determine whether he is with any other women.

To the extent that Edwards does not monitor John's actions and behaviors because she *believes* that his good will and competence will prevent him from being unfaithful, Elizabeth trusts John.<sup>9</sup> Suppose that John has come clean about all his past infidelities and guarantees that he will no longer seek out new paramours. If Elizabeth finds that she believes him, she has some reason to trust John. If she believes that John will not waver when the next chance arises because he will be moved by the thought of their relationship

and the fact that she is counting on him, then Elizabeth has a reason based on trust not to worry excessively and spend her time monitoring his behavior.

But Elizabeth could also decide for other reasons that monitoring John is not the thing to do. She might actually be uncertain about whether or not his entreaties have been genuine and whether, even if they were, he is competent enough to override sexual urges toward other women. Given her uncertainty, she might feel that the only thing to do is to turn him loose and see what happens. If he doesn't betray her again then she will have evidence that will allow her to trust him more easily. Thinking to herself, "I have to try to trust him if we're going to get through this," she might steel herself and try to focus on other things. Or, alternatively, Elizabeth might be focused less on the health of their relationship and more on the fact that if she obsesses about John while he's gone, she isn't able to put on a brave face for their children, who are also very upset by their father's transgressions. She might be *hoping* that he is able to live up to his commitment to her, without really believing this is the case.

If Elizabeth decides not to monitor John while he is away for any of the latter reasons, she will not have made her decision because she trusts John. Instead, she will have decided not to monitor him because, given the costs and benefits associated with monitoring, it makes more sense to try to act normally while taking a gamble on John's future behavior. In this case, Elizabeth *entrusts* the future success of their relationship to John, but she does not do it because she trusts him.

The lesson here is that it is possible to entrust someone with the care of something, even yourself, for reasons that go beyond the reasons that ground trust. In this sense, entrusting is like acting (that is, literally performing an action) and can be undertaken for

all the familiar and myriad reasons that we act. (Think of the connotation of “entrusting,” namely, handing something over to someone else.) Trust, on the other hand, is a more selective attitude. The considerations that support a well-grounded attitude of trust are a subset of the reasons that would support entrusting. It makes sense to attribute trust to someone only when that person believes that the trustee will come through, not when she decides to take a gamble on the trustee’s future behavior.

The idea that trust involves believing in others can be captured by modifying Jones’ proposal as follows:

To trust someone is to believe that the good will and competence of that person will extend to cover the domain of your interaction with her because the one trusted will be directly and favorably moved by the thought that you are counting on her.

This formulation is preferable to Jones’ account for two reasons. First, it gets the core attitude of trust right. Trusting someone involves believing certain things about that person, not hoping that those things are the case. Second, this formulation emphasizes that when we trust, we believe the good will and competence of another person will govern her interactions with us because we take her to be moved by the fact that we are counting on her. Jones’ original account decouples the reason that the person trusting is counting on the trustee from the belief that the trustee will respond with good will. It thus allows that the trustee might respond with good will, and that the trustee is moved in the right way, but that the trustee’s good will is not produced by the trustee being moved

in the relevant fashion. It is better to emphasize that when we trust, we expect the trustee's good will to be produced by her realization that we are counting on her.

## **II. The Trickster Problem and the Participant Stance**

### **A. The Trickster Problem**

The above account of trust is plausible and is, I think, better than any other account of trust on offer. Even so, it is vulnerable to what I call "the trickster problem." To see the problem, suppose that Madoff has found that emphasizing how a deal will help him in some way makes his potential victims more likely to take the bait. He tells a particular potential victim that he has a great investment opportunity and most of the capital already raised to go forward, but not quite enough. Madoff makes it clear to his mark that he needs her money to make the deal happen, or their great opportunity will be missed. He beseeches her, on behalf of himself and his other investors, to invest in the scheme with him. The trickster problem is this: in the described scenario, can't Madoff believe that because his potential investor will be directly and favorably moved by the thought that he is counting on her, that her good will and competence will extend to cover the domain of their interactions?

Since the above example is based on the way he actually operated, it is clear that Madoff could have these attitudes toward his potential victim. Therefore, our working account of trust would have it that Madoff is trusting his potential mark to go along with his scheme. This is surprising and problematic because this case then runs afoul of one of the canonical assumptions about trust: that trust that is violated tends to produce, and warrants, feelings of betrayal.

Though the case described above fits our current account of trust, it does not seem correct to say that Madoff would feel betrayed if this potential victim turned him down. He would most probably feel disappointed, but regardless of what he might feel, it strains credulity to say that it would be appropriate for Madoff to feel betrayed. We expect him to simply shrug his shoulders and move on to find his next victim. If he does not, but instead rails against his betrayal at the hands of his potential victim, we would think he was being unreasonable.

Our current proposal holds that when we trust another person we believe that because she will be directly and favorably moved toward us we can count on her good will and competence governing our interactions in a particular domain. But that proposal does not, apparently, include everything we mean to capture when we speak of trusting another person. In our example, Madoff expects that the good will and competence of his potential victim will be operative in their interactions because he believes she will be directly and favorably moved by the fact that he is counting on her, but we do not believe that he trusts her. It seems he merely relies on her to do what he wants, even though he satisfies the conditions of our current account of trust.

## B. The Participant Stance

The most prominent attempt to solve the trickster problem involves appealing to an idea developed by P.F. Strawson in his famous piece “Freedom and Resentment.” Part of Strawson’s argumentative strategy in that piece involves drawing our attention to what he calls the *reactive attitudes*, a class of attitudes that we typically have toward persons but not toward nonpersons. Strawson’s reactive attitudes are emotions like “gratitude,

resentment, forgiveness, love, and hurt feelings,” (62) but the current proposal extends Strawson’s idea to other psychological states. Trust, on this account, is also a reactive attitude.

Pamela Hieronymi has insightfully explored Richard Holton’s idea that trusting someone involves relying on that person from the “participant stance.”<sup>10</sup> It is generally agreed, as Hieronymi explicitly states, that the participant stance is the stance we take toward other responsible agents—that is, persons—and that this stance is inappropriate toward nonpersons. But reliance, we have noted, is simply characterized by the expectation that something will happen. Certainly it can be appropriate to rely on nonpersons or even inanimate objects to do certain things. What would it mean to expect a person to do something *from the participant stance*? Also, it is clear that we can simply rely on, rather than trust, other persons. What would the participant stance add to reliance to create trust, such that the resultant attitude would be inappropriate with respect to nonpersons?

Hieronymi’s tentative answer to these questions regarding relying on someone from the participant stance involves appealing to an analogy between two different grounds you might have for a belief that someone will do something. Consider, on the one hand, one kind of evidence you might have that someone else would do something: that person’s stated intention to do that thing. If someone asks why you believe that your friend Jim will mow his lawn, it makes perfect sense to say that Jim told you that he intends to mow it.

Now, it is possible to regard the intentions of other people as evidence in favor of certain predictions about what another person will do. And clearly, as the case of Jim’s

mowing illustrates, we sometimes do this. But, perhaps surprisingly, it does not make sense for us to do this in our own case. Following Richard Moran, Hieronymi notes that our own intentions to do things do not serve as evidence for us that we will do what we intend (225). When we believe that we will act as we intend to act, our belief is not based upon our intention, itself, for our intentions are not stable facts about us. Since it is up to us what we intend, Hieronymi argues that we should not regard our intention as evidence of how we will act, as it is always possible that we will change our minds about what to do. Suppose you are idly wondering whether you are going to mow the lawn today. You will not be able to answer this question for yourself by simply noting that earlier you formed the intention to mow. Now that you are attending to countervailing considerations, like how hot it is outside, your estimation of the reasons for and against mowing should come to the fore while your previous intention should recede into the background of your deliberation.

Hieronymi exploits this difference in plausible available responses to the question “why should I believe that an agent will perform a particular action?” in her account of the difference between trust and reliance. When you trust someone, she believes, this involves thinking that another person will do something because of the reasons that you believe that person has “in something like the way that one might form a belief about one’s own future on the basis of one’s own practical reasons” (226). Reliance, in contrast, is distinguished from trust by lacking this distinctive focus on the other person’s reasons.

### C. Problems with the Participant Proposal

Suppose, then, that a unique focus on another person's reasons defines the participant stance. But it is not clear how to vindicate Hieronymi's idea that mere reliance on another person does not involve the participant stance, while trust does. This is because many cases of simply relying on another person will involve an implicit reference to that person's reasons, given that we are dealing with another agent. For example, I rely on other motorists to allow me to proceed first when I arrive at a 4-way stop before they have. If I consider why I rely on them it is because I have certain beliefs about their reasons: I believe they want to avoid accidents and being cited by the police, and that these reasons are sufficient to get them to follow most traffic laws. But simply bringing the content of the reasons of others to mind does not amount to trusting them. Nor would it be reasonable for me to feel betrayed if one of the other drivers broke the law and entered the intersection before me. While anger or exasperation might be reasonable for me to feel toward the driver, it would not be appropriate to feel betrayed.

Perhaps the idea is, instead, that trusting involves a special kind of justification for my beliefs about the other person's reasons. This way of developing the idea is supported by Hieronymi's appeal to an analogy with testimony. When you testify to me that something is true, you offer a kind of guarantee. You put yourself on the hook to some degree, morally speaking, for what results from my new belief based on your testimony. Further, you personally stand behind my belief in a way that you would not if I came to believe by some other manner.<sup>11</sup> Hieronymi might then object, based on the appeal to testimony, that the personal assurance that is characteristic of trust is not present in my driving example. I have beliefs about what the other driver's reasons are,

to be sure, but they are not based on any specific information about their beliefs. Perhaps if my belief about the content of their reasons was based on their testimony, or some other kind of uniquely personal assurance, then I could be said to trust the other drivers.

But this proposal also does not seem sufficient for trust. Imagine that I am hired by the city for a compliance project: I am to stand at 4-way stops and question drivers who stop about why they do so. Most drivers tell me that they stop because of a fear of getting a ticket and because they hope to avoid accidents. On the way home after a day of interviews, my path takes me to a 4-way stop. To my surprise, I notice that someone who answered my questions earlier in the day is stopped there as well. In recalling the fact that this subject told me that she obeys traffic laws because she is concerned about accidents, I do not thereby trust to follow those rules. If she starts into the intersection even though I have the right of way, I certainly might become angry with her, as I might with any driver who cuts me off, but I would not feel betrayed.

Further difficulty for understanding trust as reliance from the participant stance comes from attempting to use it to solve the trickster problem. When Madoff relies on his potential victim to go along with his plan, we can suppose that he has many true beliefs about the kinds of considerations that she tends to view as reasons. As a good-hearted person, she tends to support the plans of others people and she has even more of a soft spot for friends of her friends. Madoff knows that she tends to be frugal with her money, but also will take a chance from time to time on a more risky investment. Further, he correctly surmises that she finds his current list of clients very impressive, which gives her confidence in his excellence as an investment advisor.

All of these considerations make it quite reasonable for Madoff to think that his

potential victim will go along with his scheme so long as she doesn't know that the scheme is fraudulent. But even if Madoff is relying on her because he is making reasonable assumptions about her reasons, and even if his victim has told him that these are her reasons for participating, that simply indicates to him that she may be an easy mark for his scam. None of these facts would warrant Madoff feeling betrayed if she refused to invest with him.

These observations suggest that relying on another person because of personal knowledge of the content of her reasons is not sufficient to create trust, nor is it sufficient to warrant betrayal should the reliance be let down. This is true even when the other person has testified to you about her reasons for doing what you are relying on her to do. On Hieronymi's account, it would seem that I trust the other driver and Madoff trusts his victim, therefore supporting feelings of betrayal if I and Madoff are let down. But as we've seen, this is at odds with plausible intuitions about the cases. Therefore, we are forced to conclude that there is a problem with Hieronymi's account of trust as reliance from the participant standpoint. That account is also not sufficient to solve the trickster problem.

### **III. Solving the Trickster Problem**

#### **A. Feeling Betrayed**

A clue to developing an account of trust that will solve the trickster problem can be gained by reflecting on what it is to feel betrayed. Since Baier's formative paper, theorists interested in trust have been appealing to the idea that when someone's trust is violated, it makes sense for that person to feel betrayed, not to merely feel disappointed.

But there has been little attention to the distinction between feeling betrayed and feeling disappointed. Here is a proposal that makes sense of the distinction and also has important ramifications for an account of trust.

Both feeling betrayed and feeling disappointed involve emoting in a particular, negatively tinged way. As understood in the current context, feeling betrayed and feeling disappointed are both emotional phenomena, in the sense of involving felt instances of emotion, not dispositions or traits. But these reactions are not basic emotional responses like anger, guilt, or joy.<sup>12</sup> When a person “feels betrayed” or “feels disappointed” her underlying emotional state is consistent with a plurality of more basic emotional responses, including anger, disgust and sadness.<sup>13</sup> It is perfectly coherent to feel betrayed or disappointed in virtue of feeling angry about what someone has done. It is also possible to feel betrayed or disappointed in virtue of feeling sad or disgusted about what has transpired.

These observations suggest that the distinction between feeling betrayed and feeling disappointed does not stem from the emotional experience that undergirds them. For example, we would recognize a difference between disappointment and betrayal even in cases where the person who is disappointed and the person who feels betrayed both feel very sad because of what happens.

What distinguishes feelings of betrayal from feelings of disappointment is that the clearest cases of feeling betrayed seem to involve feeling angry, disgusted or sad at the rupture of an important relationship. Being disappointed lacks this specific content. We would not say that Elizabeth Edwards feels disappointed that John committed adultery. We describe her as feeling betrayed because the harm she has suffered is specific to her

relationship with John. We also describe Bernie Madoff's clients as feeling betrayed because they took him to be more than just an investment advisor, but also their friend. Other things equal, feelings of betrayal tend to be more intense, given that they are responsive to the breach of relationships that are centrally important to our lives. But the difference between feeling betrayed and feeling disappointed continues to assert itself even if we suppose that the intensity of the sadness is equivalent across an episode of betrayal and an episode of disappointment. This is because feeling betrayed involves an implicit reference to the external situation—the breach of a relationship—that warrants the emotion.

#### B. Relationships of Trust

Because betrayed involves being emotionally bothered by the fact that an important relationship has been violated or ruptured, our account points in an important direction. It is no accident that the attitude that is characteristic of trust—believing that, because a person will be directly and favorably moved toward us, we can count on their good will and competence—is an attitude that is a constitutive part of many different socially salient relationships.<sup>14</sup> We expect that attitude from friends, romantic partners, parents and others whom we count on. What has been missing from previous accounts of trust is that in the context of a relationship characterized by trust we believe not just that we can expect good will from someone, but we take ourselves to be *entitled* to it. It is our belief that we are entitled to that person's good will, and the fact that we take the other party to owe us good will, which legitimates our feeling betrayed when trust is violated.<sup>15</sup>

Let me say a bit more about how I understand the relevant interpersonal relationships. In a healthy friendship or romantic partnership, for example, the parties involved have standing dispositions to treat each other well, spend time with each other, and help each other. A healthy friendship between two people will also involve distinctive ways of caring and responding to each other. A true friend gets upset at the things that negatively impact you and cares about your projects. When you are a party to a relationship of this kind, one of your overarching presumptions about the other person is that she will act toward you with good will because she will be directly and favorably moved by the fact that you are counting on her.

However, it is not simply a happy fact that the other person bears you good will and that you realize this. In the context of a healthy interpersonal relationship, the fact that the other person will act toward you with good will because she will be directly and favorably moved by the fact that you are counting on her becomes normative. That is why, if you are treated poorly by a friend, you can respond to the slight by saying “Aren’t we friends?” By doing this, you call your friend’s attention to the relationship the two of you share and how it gives you each reasons to foster and care for one another. By slighting you, your friend breaches the terms of the relationship. In order to make sense of this idea, we must take relationships like friendship to be partly normatively defined. It is not just that the parties to the relationship have attitudes that coalesce happily. They have a normative claim on these attitudes from each other.<sup>16</sup>

What is striking about the normatively constituted relationships I have been discussing is that one of the overarching norms of all of these relationships is what I call the trust norm. According to the trust norm, when you are a party to a normatively

defined relationship of the right kind, the other person who is a party to that relationship owes you good will because she should be directly and favorably moved by the fact that you are counting on her. In order to qualify as a proper friend, romantic partner or parent, you must see the other party as having a claim on your good will. In this way, these relationships presuppose trust.

### C. The Trickster Problem, Solved

The account of normatively defined relationships can help us solve the trickster problem. It explains why Madoff does not trust his potential victim to do as he wishes and why he will not feel betrayed if his mark does not go along with his scheme. Given his deception, Madoff is not *entitled* to take his victim's good will toward him to get her to go along with his scheme. The flip side of this claim is that Madoff's victim does not *owe* him her good will.

What is key to solving the trickster problem and makes sense of our intuitions about the previous examples, is that trust involves the belief that the good will of another person will cover our interactions with her because she owes us good will due to the terms of our interpersonal relationship with her.<sup>17</sup> Madoff is not actually coming to his victim under the umbrella of an interpersonal relationship that involves trust. He may well be presenting himself that way—and, indeed, his presentation of himself as a trusted friend was a large part of his success. Madoff's pretense at friendship attempts to lead his potential victim to act because of her good will toward him, not merely to act because of a potential benefit to herself. Given that he comes to his potential victim as a friend, it may appear to her that she has some reason to be motivated by the fact that he is counting

on her to act with good will toward his plans. Since we can imagine her being persuaded of his good intentions toward her under the guise of friendship, we can imagine that Madoff can then have the beliefs characteristic of trust that involve relying on her good will and concern for him. But while he can have reasonable beliefs about his victim's attitudes, he is not entitled to treat her attitudes with the significance they would normally have for a friend. He is not entitled to believe something further—that his victim owes him her good will. And since he realizes he is not entitled to treat her attitudes as evocative of friendship, he should not feel betrayed if she does not go along with his plan.

#### **IV. Objections, Replies, and Conclusion**

In this paper I have offered an account of trust that aims to solve the trickster problem. It does this by calling attention to a neglected feature of trusting—its normative aspect. Some previous accounts of trust tried to specify the requisite mental attitudes on the part of the truster that, taken together, would add up to trust. Other accounts appealed to the notion of the participant stance. Neither, on its own, could solve the problem.

My account is, in some ways, a hybrid of these two approaches. I also appeal to mental attitudes on the part of the truster, namely a belief that the other person will act toward you with good will because she is directly and favorably moved by the thought that you are counting on her. But that is not enough, as the trickster case exemplified by Bernie Madoff demonstrates. The missing ingredient is normative. Someone who reasonably trusts another has a special sort of normative claim on the good will and concern of the trusted person.

To explain this claim I have appealed to the examples provided by a number of

interpersonal relationships like friendships and romantic partnerships. It is important to emphasize that I do not exclude other relationships from providing the normative backing that is required for trust. All that is required is that a requisite relationship includes, as one of its constitutive norms, what I call the Trust Norm. The trust norm states that

Because it is partly constitutive of a normatively defined relationship you are a party to, the other person who is a party to that relationship owes you good will because she should be directly and favorably moved by the fact that you are counting on her.

If a relationship includes the trust norm as one of its constitutive norms and two people are parties to that relationship, then the normative backing of that relationship will entitle at least one of the parties to the good will and concern of the other. In many such relationships, of course, the relevant trust norm is bivalent.

The trust norm does the relevant work in solving the trickster case, since it shows that the trickster's victim does not owe him good will and concern. But it also helps to explain the phenomenon of feeling betrayed, as contrasted with feeling disappointed. Feeling betrayed involves feeling violated, feeling that the relationship on which you thought you were entitled to count was breached. This response makes sense in the context of a relationship that is normatively defined and includes the trust norm as one of its constitutive norms.

It might be objected at this point that while my account of trust has the virtue of generating the right result in the trickster case, it comes at a cost, for my account might

appear to preclude attributions of trust *outside* socially salient relationships like friendships, romantic partnerships, and the like.<sup>18</sup> After all, don't we sometimes trust others with whom we are not party to such a relationship? For example, we might trust the doctor we see for regular checkups, or we might trust our drycleaner to return our clothes clean and on time.<sup>19</sup> If my account cannot countenance trusting in such cases, then even though the view can handle the trickster problem, it runs afoul of common attributions of trust.

Fully addressing this concern requires that I make several points. The first is that nothing in my account precludes other relationships, like that between a doctor and patient, from being ones where we recognize trust as extant and appropriate. I do not here offer a theory of which relationships are appropriately characterized by the trust norm—I simply appeal to examples like friendships, partnerships and parenting that seem to be correctly characterized by it. One of the virtues of my account is that it allows any relationship properly characterized by the trust norm to count as a trusting relationship, not just the exemplars I have mentioned. The doctor-patient relationship may well be a case properly characterized by the trust norm. I take no stand on that here, but nothing in my account precludes it. It is not obvious, however, that what I have said so far is enough to handle the drycleaner case. After all, surely we can imagine someone trusting her drycleaner to perform competently and being angrily disappointed when her new shirt is returned with a noticeable tear, even though the drycleaner-customer relationship does not seem appropriately characterized by the trust norm.

My response to the drycleaner case is two-fold. For one, it is not clear that the drycleaner case counts as a case of trust in the sense importantly connected with

betrayal.<sup>20</sup> Evidence for this treatment is in the natural way the customer's reaction is described: as feeling angrily disappointed about the tear in the shirt, but not feeling betrayed. If we continue to make use of feeling betrayed as a marker of the existence of trust, it is safe to say that while the customer may have a normative expectation of the drycleaner, we need not suppose it is a specifically trusting one. On the other hand, we can imagine the case such that the customer really does feel *betrayed* at the rip in the shirt. If so, then it is important to recall that if my analysis of feeling betrayed is correct, in order to count as feeling *betrayed*, the customer must be feeling something other than simply disappointment, anger, or sadness. On my account, the additional element is the feeling that an important relationship has been ruptured.

I allow that feeling could be present in the drycleaner case, but only on the supposition that the customer takes herself to have a relationship with her drycleaner properly characterized by the trust norm. Such a case is easily imaginable: suppose the customer has been going to that drycleaner for a significant amount of time and they relate about other matters than merely the number of garments to be laundered, the method to be used, and the price of such services. Suppose they exchange more than simple pleasantries about each other's personal lives—they volunteer personal information and express real concern for one another. But in filling in the case like this, we are making the customer's feelings of betrayal understandable in virtue of the customer's taking herself to be a party to a relationship properly characterized by the trust norm. Her friendly patronage of the drycleaner makes her come to think of the drycleaner as, in some way, more than a mere economic acquaintance. Of course, the customer might be unreasonable in construing her relations with the drycleaner in that manner (it

will depend on the details of their interactions). If so, she still trusts, though she is unreasonable in doing so.

This last point is worth expanding: my account does not require that someone *actually* be a part to a relationship of the requisite kind in order to count as trusting—it is sufficient for trust that someone *believes* himself to be a party to such a relationship. For example, imagine a case where one lover trusts another not to stray, while he himself has no compunction about cheating.<sup>21</sup> We can certainly imagine that the straying lover might feel betrayed should his partner come to share his taste for infidelity. Given his own transgressions, his trusting would be unreasonable because his unwillingness to be faithful means he is not entitled to fidelity from his partner. He is not entitled to his partner being directly moved by the fact that he is counting on the partner not to cheat. My account of trust can correctly characterize this case as unreasonable, but extant, trust because it does not require that the relationship *is* as the one trusting believes it to be—though reasonable trust does require this.

Of course, on my view, believing that you are a party to a relationship characterized by the trust norm is not all there is to trusting. You could believe that you are entitled to count on the good will of another person because of your participation in an interpersonal relationship with them properly characterized by the trust norm, but not believe that they will come through. This is just the situation that Elizabeth Edwards faces in response to John's infidelity. She knows that she is entitled to John being directly and favorably moved to act toward her with good will, but she is not at all sure that he will act in that way. To the extent that Elizabeth Edwards doubts that John will act as she is entitled for him to act, she fails to trust him.

This is why the correct account of trust requires both normative and nonnormative attitudes. We need the normative part to exclude Bernie Madoff from counting as trusting his potential victim and we need the nonnormative part to properly characterize Elizabeth Edwards' struggle to trust John. Our final account of trust holds that

To trust someone is (i) to believe that because a person will be directly and favorably moved toward us we can count on her good will and competence governing our interactions in a particular domain and (ii) to believe that we are entitled to her good will because we are a party to a normatively characterized relationship with that person.

Trust, then, is a rather complex attitude. It is no wonder that it has taken so long to solve the trickster problem.<sup>22</sup>

## Notes

---

<sup>1</sup> For years, Bernie Madoff ran the world's largest Ponzi scheme. Madoff continually took on new investors, whose money would be funneled to previous investors at incredibly high rates of return. When he ran out of new investor prospects, the scheme collapsed. For more on Madoff's scheme, see (Henriques).

<sup>2</sup> The examples are from (Jones).

<sup>3</sup> Shortly after John ended his campaign for the Democratic Party's nomination for the US Presidency, it was revealed that he had been having an affair with another woman. Elizabeth Edwards has since written a book detailing how she dealt with the revelation of the affair. For more, see (Seelye).

<sup>4</sup> Annette Baier first connected and distinguished trust and reliance in her "Trust and Antitrust," which began serious inquiry into trust as philosophical project .

<sup>5</sup> In what follows I largely ignore complications introduced by the claim that warranted trust requires that the other person be competent. I agree with Jones that trust is not warranted toward those who are incompetent. However, I leave complexities involving competence aside because it seems clear that if you properly trusted someone and that person became incompetent, feelings of betrayal would not be warranted toward that person. I'm interested in cases where competent, trusted people do things that warrant feelings of betrayal.

<sup>6</sup> Jones has in mind an account of the emotions from (Rorty; Calhoun; de Sousa; De Sousa).

<sup>7</sup> Some of the best known research into cognitive biases is collected in (Kahneman, Slovic, and Tversky).

<sup>8</sup> The halo effect was first empirically demonstrated in (Thorndike). Thorndike asked commanding officers to rate soldiers that served under them and found that the officers' ratings showed a high degree of correlation between positive and negative traits, respectively, in each individual soldier.

The halo effect has been confirmed by many other experiments. See, for example, (Nisbett and Wilson).

The halo effect is regularly invoked today to explain other social phenomena. For example, it is touted as the reason for the reemergence of Apple's Mac computers, based on the halo effect generated by positive associations with Apple's iPod MP3 players.

<sup>9</sup> I here follow a critique of Holton deployed in (Hieronymi).

<sup>10</sup> I focus here on Hieronymi's discussion rather than Holton's because Holton's account doesn't go very far toward explaining the phenomenon in question. Holton thinks of trusting someone as being vulnerable to betrayal in your interactions with them. On this account, when someone feels betrayed, that person's betrayal is a reactive attitude. Disappointment, it seems, is not a reactive attitude for Holton.

Holton notes that while this distinction doesn't provide a reductive analysis of trust, it allows us to point to a distinction between cases where there is trust and cases where there is no trust (67). But this is little progress, regardless of whether there is a reduction of the concept. For we have no explanation of *why* feeling betrayed is appropriate when one's trust is violated or why betrayal, but not disappointment, is a reactive attitude.

<sup>11</sup> While I am skeptical that the appeal to testimony can directly address the trickster problem, the idea of “being on the hook” for something or “standing behind” something or someone has a normative ring to it that has an affinity with my preferred account. See sections 3 and 4, below.

<sup>12</sup> The basic emotions are thought to make up a class from which all other emotions are derived. There is some dispute about exactly which emotions are basic, but a representative list might include some or all of the following: amusement, anger, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness, satisfaction, sensory pleasure and shame. This list of 15 emotions is from (Ekman 45-60). Ekman defended a shorter list of six basic emotions in earlier work (happiness, surprise, anger, sadness, fear, disgust/contempt).

<sup>13</sup> When Elizabeth Edwards learned of her husband’s affair, she apparent felt all three of these emotions in the course of feeling betrayed. She writes: “I cried and screamed, I went to the bathroom and threw up” (Seelye).(Walker)

<sup>14</sup> The idea that the relevant relationship be one that is recognized as “socially salient” is meant to exclude relationships that hold between two people simply in virtue of sharing some property. For example, two people might share the relationship of having the same number of letters in their last name. (Scheffler 198).

<sup>15</sup> The term “entitled” is natural here, but I want to contrast its current use with another use of “entitlement,” that is, a rule-governed institutional right. For a person to be entitled, in my sense, is for that person to have a normative claim based on an interpersonal relationship, not an institution. The claim on someone’s good will based on the relationship is supported by norms that need not be explicit rules. However, the structure of the claim does bear some similarities with claims of institutional entitlement.

Further, I take no stand here on what explains the fact that we are sometimes entitled to another person being moved in the right way to act toward us with good will. The account of trust I offer is neutral between an account of interpersonal relationships that is “voluntaristic” in treating the relevant entitlements as arising only from the voluntary actions of people and an account of these relationships that allows for legitimate claims on the good will of others being supported by considerations other than voluntary actions. For more on the distinction, see (Scheffler 191).

<sup>16</sup> I draw the example of appealing to the terms of a relationship from (Gilbert 149-153). Gilbert’s target analysis is the notion of a joint commitment, rather than trust, though she notes there are important similarities. I follow her in thinking that trust, like a joint commitment, is normative.

<sup>17</sup> “Terms” can mislead. I do not assume that the relevant terms need to be explicitly stated or agreed to.

<sup>18</sup> I thank an anonymous reviewer for raising this objection.

<sup>19</sup> I owe the example of the drycleaner to Margaret Walker’s discussion of trust (Walker 82).

<sup>20</sup> While I don’t deny that you can correctly describe yourself as trusting your drycleaner with your clothes, I want to insist that this is a different phenomenon, supported by the normative expectations that flow from the drycleaner’s role in the economic system. Margaret Walker is in agreement: while she would attribute trust both in the case of

---

friendship and reliance on a drycleaner, her view seems to be that the normative basis of such reliance is not the same (Walker 81-82).

<sup>21</sup> I am indebted to an anonymous reviewer for suggesting discussion of a case like this.

<sup>22</sup> I am indebted to helpful comments from audiences at the the UCLA Ethics Writing Workshop, Rocky Mountain Ethics Congress, Eastern American Philosophical Association Meeting, Feminist Ethics and Social Theory Conference, and the Southern California Law and Philosophy Group. I am particularly thankful to Andrea Scarpino, Pamela Hieronymi, and Steve Munzer, who all provided particularly valuable feedback and encouragement.