Title
The reasons of trust

Journal
Australasian Journal of Philosophy, 86(2)

ISSN
0004-8402

Author
Hieronymi, Pamela

Publication Date
2008-06-01

DOI
10.1080/00048400801886496

Peer reviewed

This is an Accepted Manuscript of an article published in *The Australasian Journal of Philosophy* 86, no. 2 (June 2008): 213–36, available online: http://www.tandfonline.com/10.1080/00048400801886496.

The Reasons of Trust

Pamela Hieronymi
hieronymi@ucla.edu
August 31, 2007

Trust is required for any collective enterprise; a psychologically healthy person must be capable of trusting others; trusting relationships are a vital component of a fulfilling life. Trusting is sometimes a way to build greater trust in a particular relationship, to build the self-esteem of the one trusted, or to avoid hurting someone's feelings. It may be required by one's role as parent, teacher, or friend. Trust is, in each of these ways, useful, valuable, important, or required.

I will argue to a conclusion I find at once surprising and intuitive: although many such considerations show trust useful, valuable, important, or required, these are not the reasons for which one trusts a particular person to do a particular thing.[1] The reasons for which one trusts a particular person on a particular occasion concern, not the value or importance or necessity of trust itself, but rather the trustworthiness of the person in question in the matter at hand. In fact, I will suggest, the degree to which you trust a particular person to do a particular thing will vary inversely with the degree to which you must rely, for the motivation or justification of your trusting response, on reasons that concern the importance or value or necessity of having such a response.

The claim is surprising, for two reasons. First, it is generally surprising whenever we realize that reasons that genuinely show some response good or important or required are nonetheless not reasons for which we can have that response. It seems to us that, if certain reasons show something good (etc.) to do, and if it is the sort of thing we can do for reasons, then we should be

[1] Throughout, I will be focusing on trusting someone to do something—on trust as a three-place relation, so to speak (*x* trusts *y* to ϕ). Interesting work has been done on what it is to trust someone, in general, or to be someone who is generally trusting of others. These are not my topic (though they are obviously and interestingly related).

able to do it for *those* reasons. But, surprisingly, this is not so. Here I argue that it is not so for trust.[2]

Second, the claim is surprising because it seems false. In ordinary usage, 'trust' covers a wide range of cases, allowing us to correctly say that we have decided to trust someone whom we do not think trustworthy, because, e.g., the demands of parenting or politics or pragmatics require it. But, if we can trust those whom we do not think trustworthy, for such reasons, then the surprising-yet-intuitive claim is false: we do, after all, trust particular people, to do particular things, for reasons that concern only the importance of having a trusting response.

Though I do not want to deny that we sometimes trust those whose trustworthiness we doubt, and that our response is often motivated by reasons that concern the importance of trust, I nonetheless want to insist that there is something intuitive about the surprising-yet-intuitive claim: 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. And thus it seems that, to the extent that you *must* rely, for the justification or motivation of your trusting response, on reasons that appeal, not to the trustworthiness of the person in question, but rather to the importance of trusting, to that extent you do not trust the person, and so, to that extent, your response was not (fully) trusting.

My aim, then, is to vindicate what is intuitive about the surprising-but-intuitive claim (the claim that reasons showing trust useful, valuable, important, or required are not the reasons for which one trusts a particular person to do a particular thing) without denying the ordinary

[2] Elsewhere I have argued that belief, intention, certain emotions, and virtuous actions and attitudes provide other examples of responses that can be shown good or useful or important or required for reasons for which one cannot have the response. I became interested in trust primarily as a particularly useful example of this larger class. I consider belief, intentions, and the reasons for them in [Hieronymi, 2005, ; Hieronymi, 2006]; I consider resentment in [Hieronymi, 2001, ; Hieronymi, 2004]; I consider virtuous and vicious action in [Hieronymi, in progress].

phenomena that seem to show it false. Vindicating what seems intuitive will require refining our ordinary notion of trust in a way that is, I think, illuminating.

Accordingly, much of this paper is spent supporting what some would call a 'purist's' notion of trust, according to which one person trusts another to do something only to the extent that the one trustingly believes that the other will do that thing. (I will provide some, though not a complete, account of *trusting belief*.) I will suggest that this purist's notion of trust is not merely a technical stipulation, but rather a natural refinement of our ordinary notion.

Of course, it is widely (though not universally) granted that you cannot believe for reasons that you do not take to bear on the truth of the belief. If this is granted, it follows that you cannot believe that someone will do something for reasons that you do not take to bear on whether that person will. Since reasons that show trust good, important, etc., do not bear on whether the person in question will do the thing in question, one cannot trustingly believe that someone will do something for reasons that show trust good, important, etc.[3] Even though reasons that show trust good, etc., might be among one's reasons for acting, and even though one's action might be a trusting one, these reasons will not support one's trusting belief. Thus, if we adopt the refined notion of trust, reasons that concern the importance of trust are not reasons for which one person trusts another to do something. In fact, I will suggest, surprisingly but I think intuitively, to whatever degree you need to supplement your confidence in the person with reasons that show the trusting response important, to that extent you fall short of (fully) trusting the person in question.

[3] One might think that a trusting belief is just an ordinary belief that is *accompanied by* some other attitude or affective state, where that accompanying attitude or affect could be supported by reasons that do not bear on the truth of the belief. In footnote 16 I argue that a trusting belief should, instead, be understood as a belief that is grounded in certain sorts of reasons (the reasons of trust).

AN INITIAL EXAMINATION OF TRUST

As noted, to vindicate the surprising-but-intuitive claim, we need to refine our notion of trust. We should start with an account of trust to be refined. I will borrow, initially, from Richard Holton [1994].

As a first approximation, consider reliance. When you rely on someone to do something, you 'work the supposition that she will do it into your plans'.[4] Trust is similar. In fact, at first glance, one might think them the same. But, as Holton notes, a trickster can rely on you to do something without trusting you. Trust requires something more: a vulnerability to betrayal if let down. The trickster might be disappointed if you did something other than what he was relying on you to do, but he would not have been betrayed. I will take vulnerability to betrayal as a kind of touchstone for trust: whereas misplaced reliance is merely disappointed, a trust is betrayed. Thus, one trusts only if one in some way risks betrayal, should one be disappointed. Otherwise, one merely relies, or perhaps only acts-as-if one trusts.[5]

Holton notes that feelings of betrayal belong to a class of attitudes, like resentment and gratitude, which we take toward people, but not toward objects. P. F. Strawson has named this class the 'reactive attitudes', and noted that they are bound up with ascriptions of responsibility [1962]. I will follow Holton in calling the 'stance' from which the reactive attitudes are appropriate, that is, the stance we take towards those we regard as responsible agents, the 'participant stance'.

---

[4] The formula is Holton's.

[5] By distinguishing trust from mere reliance, I eschew what Karen Jones [1998] calls 'risk-assessment accounts' of trust. Although it is fairly common (among non-risk-assessment accounts) to distinguish trust from mere reliance by appeal to vulnerability to betrayal, it is worth noting that 'betrayal' must be used in a somewhat technical sense: in ordinary usage, one is betrayed only in cases of moderate to great importance; it seems a stretch or exaggeration to say, in small things, that you were betrayed. Still, even in small matters, it seems there is a difference between the disappointment of reliance and the disappointment of trust. I take this to be a conceptual or 'grammatical' point (rather than a psychological one): whereas misplaced reliance is merely disappointed, trust is the sort of thing that can be betrayed. (Cf. [Baier, 1986, 235]: 'trusting can be betrayed, or at least let down, and not just disappointed'.) More will be said about betrayal below.

On Holton's account, trust is *reliance from the participant stance*. Thus far, I agree.

Consider, next, whether trusting someone to do something requires believing that that person will do that thing.[6] Initially I think it seems plausible, perhaps commonsensical, to claim that trust requires confidence, that is, requires something like this belief. But in certain cases it seems we can decide to trust someone to do something even though we do not believe that he or she will. Holton considers what happens in a 'trust circle'. (A trust circle is a trust-building exercise common to drama classes in which one person stands in the centre of a circle of people, is spun around, eyes closed, until she loses her bearings, and then, arms down and knees straight, falls backward to be caught by the others.) Holton notes that, when in the circle, 'there is a moment at which you weigh up whether or not to let yourself fall . . . It feels as though you are *deciding* whether or not to trust'. He advocates taking this feeling 'at face value'. He then goes on to ask,

> Does my decision to trust the others entail that I *believe* they will catch me? If it does, does this in turn mean that when I decide to trust them, I also decide to believe that they will catch me? I think not. In order to trust I do not need to believe . . . Mightn't I be most uncertain that I will be caught, but decide to trust anyway? [1994, 63]

According to Holton, this decision to rely from the participant stance can be sufficient for trust. One need not believe.

Holton wants to allow for trust without belief exactly because he recognizes that certain reasons that count in favour of trusting will not serve as reasons for believing. These are just the sort of reason we have already considered (that trust is required for the well-functioning of your collective enterprise, that you want to build greater trust in the relationships at hand, or think your current role requires it). Such reasons count in favour of having a trusting response but, it seems, will not support the belief that people in question will do the thing in question. And so,

[6] Throughout I will assume (with Holton) that, if *x* trusts *y* to ϕ, the relevant belief to consider is the belief that *y* will ϕ. An anonymous reviewer rightly points out that this may be too thin of an account of the content of the relevant belief, as *x* need not believe that *y* will ϕ *come what may*. The relevant belief(s) may be quite complex.

Holton reasons, if one can decide to trust for such reasons, deciding to trust cannot entail deciding to believe.

Of course, it is exactly this sort of reason that I hope to show problematic, precisely because such reasons will not support the belief that is required for trust. Nonetheless, I want to agree that one could, in the trust circle, do something that is reasonably called 'deciding to trust'. In fact, I think at least two different things could be going on at the moment of 'weighing up' that feels like decision.

First, as you 'weigh things up', you might be making up your mind whether you trust these people. This is like making up your mind whether you believe that something is true. Indeed, it seems that the best way to describe this first possibility is to say that when you weigh up whether you trust them, asking yourself, 'do I trust these people?' (rather than asking whether to let yourself fall), you are making up your mind about what to believe—whether to believe they will catch you. While you certainly cannot 'just' decide to believe that they will catch you (just because you want to, e.g., or because it would be good for your team dynamic), you may nevertheless be able to form a judgment, on the spot, regarding your confidence in them. Thus, as you weigh things up, you might conclude that you do trust these people—that, in fact, you believe they will catch you. I want to call this first possibility *full-fledged* trust, 'full-fledged' because it includes the belief that you will be caught.

Of course, even if you believe you will be caught, you may still, as you let yourself fall, feel fear; but this is fear you will disavow. Fear, famously, can be felt even when you believe there is no danger. In this case it is seen as an irrational fear, something to which you are, in some way, subject. In the face of your fear, you might rehearse to yourself evidence to support your belief

that you will be caught (this game is played all the time, people always catch one another, etc.).[7] That is, you might do the same things you would do in confronting your irrational fear of the dark by rehearsing to yourself your reasons for believing that your environment is actually safe. Nevertheless, in the moment of weighing up, you decided whether or not to fall by making up your mind whether or not you believe you will be caught. Believing they will catch you, you disavow your fear, putting forward whatever degree of effort necessary to overcome it, and let yourself fall.

There is, certainly, another possibility, which Holton has in mind. As you hesitate and weigh up, you may find that you are uncertain whether or not they will catch you. You have no definite beliefs on this point, and, what is more, you cannot come to any. You are in a state of doubt. From this state of doubt, you must now decide whether to fall. You may then decide to *entrust* yourself to them, and let yourself fall.

In this case, you should feel fear. And, in this case, you cannot simply disavow and overcome that fear. It is not irrational. Rather, in this case, the uncertainty that grounds your fear represents your all-things-considered assessment. The risk of falling might be outweighed by other considerations, but the fear cannot be disavowed or simply overcome or convinced to agree with you.[8] In confronting your fear, you will not marshal evidence to support a belief that you will be caught; rather you will tell yourself either why falling is worth the risk or why the risk of injury is not a legitimate consideration in this case. In spite of your lack of confidence, you tell yourself why you should *fall*.

[7] I suspect that the reasons you rehearse to your fear are not the reasons that underwrite your trust, since (I will suggest) trust is not grounded in 'evidence' or probabilities. But this should be expected: You are trying to convince or overcome the part of you that *does not* trust, so you rightly rehearse reasons of a different sort.

[8] Once outweighed or silenced, the fear might persist at a level that one does not deem appropriate. In this case it will again seem irrational, and so again need to be overcome.

It seems plausible to suppose, as Holton does, that what steps in to get you over the resistance of your doubt, in this case, is your desire to build the relationship in question by exhibiting trust, or a feeling that you are obliged to act in a trusting way in this situation, or the thought of how hurt these people would be if you did not trust them. In the face of your doubt, you must have some reason that makes falling seem worth the risk. This 'backstop' reason is often one that makes some explicit reference to the importance, value, or usefulness of having the trusting response.

Though, in practice, these two possibilities (entrusting and full-fledged trust) may fade into one another (entrusting may well be a way to cultivate trust), and though, in many cases, it may be very hard to say which is happening, I will nevertheless maintain that there is a difference between trusting with the belief that you will be caught and entrusting in the face of doubt.[9] Further, I will argue that we should take the first, the one that entails belief, as the full-fledged and primary sort of trust. 'A decision to rely from a participant stance in the face of one's doubt' is too weak a characterization.

Suppose that, in the morning, you and I agree to meet for dinner at a certain time at a certain restaurant, to plan an upcoming event. Later in the day you learn that all my friends have decided to go to my favourite restaurant to celebrate a surprise promotion bestowed on one of them. You now doubt whether I will keep my engagement with you. You are not certain I will not, but then you are not certain I will, either. You are in a state of doubt. In the face of your doubt, you decide to go to the restaurant and wait for me. When I show up, you tell me of your anxiety and your subsequent decision to come to the restaurant, even in the face of your doubts. Upon hearing your story, I am less impressed with your overcoming your doubt by your decision to 'trust' me and more concerned with the lack of trust expressed in the doubt itself. Certainly

[9] Note that 'entrust' is also a three-place relation, but of a different form than 'trust'. While *x* trusts *y* to ϕ, where ϕ is some action, *x* entrusts *g* to *y*, where *g* is some good. With entrusting, all the action goes to *x*.

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. In this case, Holton's analysis of trust, as a decision to rely from a participant stance that does not entail a belief, picks out something considerably less than what we want to call trust.

Of course, if we change the case slightly, we would be inclined to say that you trusted me. If I had let you down similarly several times in the past, leaving you stranded at other restaurants, we might then say that you 'decided to trust me'. I think we would say this because it would be a case in which what I am calling 'entrusting' is the most we can reasonably expect, and so we call it trusting. But even in such cases, what we call trust is still distinguishable from, and something less than, the full-fledged sort—even if one thinks the full-fledged sort would be positively inappropriate in the circumstances, one can still imagine what it would be to have it, and its inappropriateness is typically explained by features of the situation seen as regrettable.

I would like to draw two conclusions from this initial examination of trust. First, trust of any sort requires vulnerability to betrayal. Second, there is a 'full-fledged' sort of trust, which entails belief, which is distinguishable from and somehow more trusting than decisions to rely in the face of doubt.[10]

---

[10] Of the two 'sorts' of trust (full-fledged trust and entrusting in the face of one's doubts), 'entrusting' might get more attention, it might be what more quickly comes to mind when we think of trust, in part because the reasons for it often make explicit reference to the importance or value or usefulness of trust. Relatedly, entrusting might get more attention because it requires more self-conscious effort. Neither of these facts should lead us to think that entrusting is either more primary, more common, or more laudable than full-fledged trust. Entrusting is what happens in the hard case, not what happens in the best case. (Of course, I do not mean to deny the importance, legitimacy, or praiseworthiness of risking some good by entrusting it to someone in the face of one's doubts.)

In an important article, Karen Jones has identified trust as 'an attitude of optimism that the goodwill and competence of another will extend to cover the domain of our interaction with her, together will the expectation that one trusted will be directly and favorably moved by the thought that we are counting on her' [1996, 4]. Like Holton, Jones believes that trust should not entail belief. Though the 'expectation' in Jones' analysis is a cognitive state and the 'optimism' is a partly cognitive, affective attitude, she insists that these do not amount to belief, in part to allow that trust can be justified when belief is not. I agree that *entrusting* can be justified when belief is not, but I do not think this is because trust is an affective state. Rather, it is because entrusting involves actions, and those actions can be fully justified even when confidence in the person is not fully justified.

A Special Case: Taking Someone's Word

We can now extend our understanding of trust by considering a special kind of case: trusting someone when she tells you something. Such cases prove useful because they need not require action: one might trust simply by believing. We can, then, consider whether, in such cases, one can decide to trust in the face of one's doubts—whether one can, so to speak, entrust one's beliefs to someone's word for the same sorts of reasons for which one can entrust one's body to someone's catch. In this section I will argue, against Holton, that we cannot. In the next section I will continue to examine this special case, attempting to locate the reasons that could support the required belief.

Suppose that an accused friend asks you to trust her when she claims she is innocent. (To avoid certain distractions, suppose she is charged with making an important, but not immoral, error.) Your friend does not want you merely to rely on her claim from a participant stance, to work the supposition that she is innocent into your plans with a readiness to feel betrayed if she is lying—that would be giving her the benefit of the doubt, acting as though she were innocent. Your friend wants you to trust that she *is* innocent; she wants you to believe her. Just as trusting the people in the trust circle requires falling, trusting your friend when she says she is innocent requires believing that she is innocent.

One might think that, in this case, you cannot decide to trust in the face of your doubt for the same sorts of reasons for which you might decide to entrust your body in the trust circle. Reasons that appeal to the importance or usefulness of trust, one might think, could not be your reasons for trusting your friend, since these reasons will not support the required belief.

Interestingly, Holton disagrees. He thinks that, even in this case, you can decide to trust—not because you can decide to believe, but because, by deciding to trust, you can *come to*

believe.[11] He notes that, while you decide to trust those in the trust circle *to catch you*, you decide to trust your accused friend *to speak knowledgeably and sincerely* when she claims that *p*. If you decide to trust her to do this, Holton maintains, you will come to believe the truth of *p*. Holton admits that, in deciding to trust her to speak knowledgeably and sincerely, you cannot be deciding to believe that she is speaking knowledgeably and sincerely. He agrees that your reasons for trusting would not support such a belief. Rather, he claims, by deciding to trust without believing, you come to believe, without deciding to. You decide to trust, and, as a *result* of this decision to trust that is not a decision to believe, you end up believing what she tells you. You can thus, in effect, entrust your belief to someone's assertions for the same sorts of reasons for which you might entrust your body to someone's catch.

Again, I think at least two different things could happen as you 'weigh up' whether to believe your friend innocent. First, as you ask yourself whether you trust your friend, you might conclude that you do trust her, i.e., you believe that she is speaking knowledgeably and sincerely. We can simplify matters by replacing 'speaking knowledgeably and sincerely' with 'telling the truth' (which is, for our purposes, equivalent). Of course, believing that someone is telling you the truth when she tells you that *p* will amount to believing that *p* is true.

Alternately, as you weigh up whether you trust your friend—whether you believe she is telling the truth—you may find that you cannot come to a conclusion on the question. You may find yourself in doubt as to whether she is telling you the truth. It seems to me that you cannot, in this case, use a 'backstop' reason to get you over your doubt, decide to trust her to tell the truth, and thereby come to believe what she says. Rather, in the case of trusting what someone says, the possibility of merely 'entrusting' in the face of your doubts seems to have vanished.

[11] Hence the title of his article, 'Deciding to Trust, Coming to Believe'.

To explain the disappearance, notice the following disanalogy: In the trust circle, once I find myself in doubt as to whether I will be caught, I then confront a *further decision* about whether to fall. If I decide to fall, I thereby decide to entrust my body to their catch. My decision to entrust my body does not somehow result in my decision to fall, nor does it somehow result in my falling without my deciding to. The decision to entrust my body just is a decision to fall; the reasons for the one just are the reasons for the other, and backstop reasons serve perfectly well. In contrast, once I find myself in doubt about whether you are telling the truth, I do not, according to Holton, confront a further decision about whether to believe what you say. The backstop reasons for entrusting my beliefs might support some action, but they will not support a decision to believe. Instead, I decide to trust without deciding to believe, and this, according to Holton, somehow has as its *effect* that I believe.[12] Thus, on Holton's account, one can cause oneself to believe, by deciding to trust.

While I do not deny that we might cause ourselves to believe, through this or that effective means, any such method will be subject to a certain instability: if you recognize that a belief was caused in you by processes that are unconnected to its truth, you will, therein, recognize your belief is unsupported, and, insofar as you are rational, you will lose that belief. Thus, once we accept Holton's analysis, whenever we arrive at a belief by trusting we will know that our belief was caused by a process that we ourselves do not think shows it to be true. Holton recognizes the difficulty here, but hopes it can be accommodated with the recognition that practical rationality sometimes has implications for theoretical rationality. While I agree that practical rationality has such implications, I think these implications too strong. Since Holton's analysis

[12] Holton may reject my claim that a decision to entrust my body is simply a decision to fall made from a (certain kind of) state of doubt. He thinks that trust is a 'distinctive state of mind', one which can be arrived at by decision, and which requires or entails other responses. Thus he might say that a decision to entrust is a decision to adopt a particular state of mind, which, in the trust circle, requires a subsequent decision to fall, or, in the case of testimony, produces a belief. But we can note an asymmetry in this account: while the decision to trust seems to require a

leaves one acquiring a belief that one also believes both to have been caused by a process unconnected to its truth and to be inadequately supported by reasons bearing on its truth, the analysis would involve one not merely in practical rationality, but in irrationality.

Because you cannot decide to believe for reasons which you might (e.g.) decide to fall,[13] the possibility of merely entrusting in the face of one's doubts, for the sorts of backstop reasons considered, disappears in the case of trusting someone to tell you the truth. These backstop reasons for trusting you despite my doubt can be reasons for acting in various ways, for entrusting various goods, but it seems that they cannot be my reasons for believing because they do not rightly bear on the truth of what you say. If I am in doubt about your innocence, the most I can do is give to you the benefit of the doubt and act as if you are innocent, perhaps entrusting to you various other things. I can decide, in the face of doubts, to put my body or time or reputation on the line, for these backstop reasons. But I cannot, for these same reasons, entrust to you my beliefs.

TRUSTING BELIEFS AND REASONS FOR THEM

Before leaving this special case of trusting someone when she tells you something, it will be useful to consider what turns out to be a difficult question: what, exactly, are the reasons for which you might trust your friend, when she claims she is innocent?

Notice that, in the special case, you trust your friend simply by believing, and so it seems you must trust, and incur the vulnerability to betrayal required for trust, in believing. Call such a belief a *trusting belief*. So our question is: what are the reasons for trusting beliefs?

---

subsequent decision to fall, it seems simply to cause a belief. (In footnote 16 I consider whether trust is best understood as a *sui generis* attitude or distinctive state of mind that can be adopted for reasons.)

[13] At this stage I have argued that you cannot believe for such reasons *without irrationality*. Later in the paper I will provide an argument that does not appeal to irrationality.

Of course, merely believing what she says does not amount to trusting your friend. You might believe she is innocent for reasons that have nothing to do with her saying so. To trust her, you must in some way rely on her for the truth of your belief. One obvious way to rely on her would be to treat her assertion as reliable evidence in support of her innocence—you believe she is likely to tell the truth, and so you count her utterance as good evidence in support of her claim. It is tempting to say that this is what happens when we acquire a belief by trusting someone. But note, if you are simply treating her utterance as reliable evidence, calculating the likelihood of her veracity, then you are precisely *not* trusting her. You are instead treating her like a good thermometer.[14] Even though you are, in a way, relying on her for the truth of your belief—you base your belief on the fact of her utterance—you are not doing so from the participant stance. If it turns out that you were mistaken in your assessment of her reliability, your mistake might result in disappointment, but you would not have been betrayed.

So we encounter something of a puzzle about trusting beliefs: while reasons which one does not take to bear on the truth of a belief cannot support that belief, the most obvious class of reasons that do bear on its truth—reasons which show the person a reliable indicator of the truth—seem not to secure the vulnerability to betrayal necessary for trust.[15] How, then, can there be such a thing as a trusting belief? How do we incur vulnerability to betrayal in believing?[16]

---

[14] I owe this vivid image to Richard Moran.

[15] For a discussion of the 'evidentialist' model of trust, see, e.g., [Owens, 2006].

[16] According to some, trust is a distinctive attitude (or set of attitudes) in its own right. Such a view might avoid the puzzle I am about to address, by understanding a trusting belief as an ordinary belief that is *accompanied by* this (or these) attitude(s). Holton, e.g., claims that trust is 'a distinctive state of mind' that one is in when one adopts a particular 'stance'. Karen Jones suggests that trust is an affective attitude: 'an attitude of optimism that the goodwill and competence of another will extend to cover the domain of our interaction with her, together will the expectation that one trusted will be directly and favorably moved by the thought that we are counting on her' [1996, 4]. By adopting an analysis of this sort, one might claim that the reasons for believing your friend's innocence could be of the ordinary, evidential sort—she is generally reliable, say. The belief could qualify as *trusting* insofar as it is accompanied by this distinctive state of mind or affective attitude, or insofar as on regards it from a particular sort of stance.

I think we should avoid such analyses, if possible, and instead try to understand a trusting belief as a belief formed or held for certain reasons or in a certain way. In the main text, I am trying to understand what those reasons or that way might be. Here I will try to say why I think it worth looking for an account of this admittedly more difficult sort. (It should be noted that the view I favor *need not deny* an important role for affect. It only denies that the reasons for which one trusts someone are reasons for having or adopting this affect.)

Suppose we assume that trust is a distinctive state of mind, affective attitude, or stance that might accompany, and so qualify as trusting, the belief or action in question. We also agree that trusting is something that one can do for reasons; there are reasons for which one trusts. We can then ask the following questions: What are the reasons for which one adopts this accompanying state of mind (i.e., reasons for trusting, on such an account), and, relatedly, what sort of agency can we exercise with respect to it? We can exercise our agency over certain states of mind, like believing or intending, by settling certain questions. We can believe p, e.g., by settling positively the question of whether p; we can intend to ϕ by settling positively the question of whether to ϕ. When we adopt these attitudes for reasons, we do so by finding convincing the reasons we take to bear on these questions. Other states of mind, like visualizing world peace, supposing p for the sake of argument, or focusing on your breathing, are action-like. We can adopt these states on command, at will, for any reason we take to show them worth adopting. Still other states of mind, like being hungry or in pain or seeing something red, are not active in either of these ways. We do not feel hunger or pain or see red simply by settling a question, nor can we do so on command or at will. Still, we can exercise a sort of agency over these states—the same sort of agency that we can exercise over any state of affairs: we can take actions we expect to affect them.

Trust, as a distinctive state of mind or affective attitude, would fall into the large class of attitudes that do not seem to fit neatly into any of these categories. If we think of it as akin to hunger or pain or perception, our agency over it would be unmysterious: it would be something caused in us by our circumstances or environment, which we would take action to manage and cultivate as we need to. But, thought of in this way, it is unclear why trust, itself, would be subject to the questions of justification to which it is subject. One can be unjustified or unreasonable in one's trust, not simply unwise or imprudent or negligent in managing and cultivating one's trusting responses. To capture this demand for justification, it seems we should instead model trust on belief, intention, or action.

If we thought trust were action-like, like imagining world peace or supposing p for the sake of argument, then we could do it for any reason that we thought showed it worth doing. But it seems implausible to think that trust is voluntary in this way. Being trusted generally carries for us a kind of significance: whether or not someone trusts us is taken to reveal something of the other person's opinion of us. If trusting were something (like speaking) that a person could do for any reason that showed it worth doing, then one could adopt the attitude insincerely, and we could question the sincerity of the attitude itself (not just its expression). But that does not seem possible.

So it seems we should explore the ways in which the distinctive attitude of trusting could be like believing or intending, in being something that can be done by settling for oneself certain questions. This would make clear both why trust is subject to questions of justification and why the state itself (as opposed to its expression) cannot be insincere, while leaving open the possibility that we can manage and cultivate it in the way we can manage and cultivate any state of mind. To explore trust as a member of this first category, we should look for the distinctive reasons for the activity of trusting—that is, we should look for the questions, the settling of which amounts to adopting this state of mind.

I doubt that there are such questions, and so I doubt that trust is like belief or intention. Rather, the reasons that count in favor of trusting seem to fall into three categories. First, there are those reasons that show trusting good, useful, or important. Second, there are reasons that bear on whether to perform some ordinary action that may, in the circumstances, amount to trusting (whether to fall, hand over your car keys, or hold your tongue). Finally, there are reasons that bear on whether the person in question will do the thing in question (catch you, tell the truth, drive carefully, handle the situation on her own).

If one finds reasons of the last sort convincing and thereby arrives at a state of mind that secures vulnerability to betrayal, it seems one will have formed what I am calling a trusting belief. (In the main text I am exploring the reasons for which one might form such a belief.) By finding reasons of the second sort convincing, one will intend to do something (fall or hand over your keys). If you manage to secure vulnerability to betrayal by finding reasons that bear on this question convincing, it seems you will have formed a trusting intention. (I suspect an intention will be trusting insofar as it is grounded in a trusting belief.)

When you trust your friend and so believe what she says, what you trust her to do is *to tell the truth*. When you fully trust her to tell the truth, you believe that she will tell the truth, and so you believe that what she says. Thus, beliefs that you acquire by (fully) trusting her are stable, well-supported beliefs—they are supported by your belief that she is telling the truth. But what supports *that* belief, that she is telling the truth? If it, in turn, were merely supported by evidence that she is a reliable indicator of the truth, it would not seem to bring with it the requisite vulnerability to betrayal. Earlier we said that vulnerability to betrayal comes with the adoption of the participant stance and that one adopts the participant stance toward those whom one regards as responsible. So perhaps your trusting belief that she is telling the truth is supported, not simply by evidence that she is reliable, but by something like the thought that she can be held responsible for the truth of what she says?

While promising, this thought will not, by itself, account for your trusting belief: you can certainly believe that she can be held responsible for telling the truth (blamed or even punished for lying or speaking loosely) without thereby also believing that she *is* telling the truth, because you think she is, in this matter, either unreliable in her judgment or malicious or otherwise ill in

---

So it seems that reasons for which one would adopt a distinctive state of mind that might *accompany* an ordinary intention or belief, and thereby qualify it as trusting, would have to be reasons of the first sort (unless, of course, I am overlooking some fourth category)—reasons that show something good or useful about trusting. But if trust is a distinctive state of mind we can adopt for such reasons, then it seems either it is action-like or else it is something we simply manage in the way we manage our perceptions and our pains. As we have seen, neither of these options seems plausible.

This line of reasoning, about the reasons of trust and our agency with respect to it, leads me to think that we should try to understand trusting beliefs and intentions, not simply as beliefs or intentions accompanied by a distinctive, trusting state of mind, but as beliefs or intentions formed for certain reasons or in certain ways. This is not to deny that trusting is a distinctive state of min—one which can be managed and cultivated—nor to deny that trusting beliefs and intentions are accompanied by characteristic affect. Trusting will be the state of mind one is in when one believes or acts for certain distinctive reasons, which may well involve certain characteristic affect, playing certain characteristic roles.

I add these reflections in response to helpful pressure from reviewers. One reviewer noted that my conclusion allows that the belief that secures vulnerability to betrayal in cases of entrusting may be so weak as to be better labeled a hope, and then wondered why we should be so sure that hopes (unlike beliefs) cannot be adopted for reasons that show them good to adopt. This note is meant to reveal some of the deeper motivations for analyzing trusting someone to do something in terms of distinctive sorts of reasons for forming ordinary beliefs and intentions. I hope that these reflections, together with the section titled 'Trust and Belief', help to address the worry.

her intent. To trustingly believe that she is telling the truth, you must not only think that she can be held responsible for what she says, you must also think that she is reliable in her judgment and good in her will. In a word, you must think she is trustworthy.

Perhaps, then, your trusting belief that she is telling the truth is supported by the thought that she is trustworthy (or considerations that show her to be trustworthy). This seems an even more promising answer to the puzzle: a belief that your friend is telling the truth might be supported by (the reasons that support) the conviction she is trustworthy, and it seems plausible that, if this is your reason for believing, you might incur vulnerability to betrayal.

I will generally rely on this more promising answer. However, a further puzzle remains, and I will digress briefly to address it.

The remaining puzzle is this: though it seems that a trusting belief *can be* supported by the conviction that a person is trustworthy, believing that you will tell the truth (e.g.) because I believe you are trustworthy still does not amount to trusting you. Again, a trickster might believe what you say, because of what he knows of your trustworthiness, without thereby incurring any vulnerability to betrayal.[17] He might simply take the fact of your trustworthiness to be evidence of your reliability. Trust seems to require something further. This brief digression is an attempt to begin to understand what makes the difference between the case in which someone takes another's trustworthiness as 'mere evidence' and that in which it is the ground for a trusting belief.

I suggest we start with one of a family of what might be called 'assurance views' of testimony: Richard Moran's account of the justification of beliefs that are acquired by believing

[17] Likewise, returning to [Jones, 1996], it seems that a trickster could have an attitude of optimism about the competence and goodwill of another and a confident expectation that she will be moved by the fact that he is counting on her, without thereby trusting her. The current puzzle is not solved by replacing belief with an attitude of optimism and an expectation—unless of course one builds a trusting vulnerability to betrayal into these attitudes.

what we are told [Moran, 2005].[18] Invoking a bit of Gricean machinery, Moran argues that, when I believe what you tell me, I rely on your intentions—I take your intention to assert as simultaneously offering me an assurance that what you have asserted is true. In intending to assert that *p* you also, thereby, offer yourself as guarantor that *p*. Thus, insofar as I accept your guarantee, your intention to assert that *p* can count as my reason for believing *p* is true. When I believe what you tell me, I believe you, that is, I rely on your intention in your act of assertion, accept your assurance of *p*'s truth, and thereby hold you responsible for my belief.

Now, of course, the possibility of betrayal is not yet captured by the notion of '*relying* on your intention in asserting'.[19] A trickster could rely on your intentions without trusting you.[20] What is it, then, to rely on your intentions in a way that amounts to trusting you? I will make a very preliminary suggestion.

First note that, at least according to many, if I intend to do something, then I will typically believe that I will do that thing—if I intend to go to the lecture tomorrow at noon, then I typically believe (other things being equal) that I will be in the lecture hall tomorrow at noon. However, my intention to do something does not, for me, simply serve as a piece of evidence to the conclusion that I will do it.[21] I cannot, on pain of bad faith, treat my own intentions merely as evidence for certain predictions about my future, because my intentions are not, from my point

---

To do so would be to insist that these amount to trusting without helping us to understand how these trusting attitudes differ from the very similar optimism and expectation that a trickster might house, without trusting.

[18] For another assurance view, see [Hinchman, 2005]. For a critical discussion of such views, see [Owens, 2006].

[19] Moran's argument was not meant to address this problem, but rather a more general problem about the justification of beliefs acquired through testimony. I take my topic to be narrower than his. Interestingly, he takes his topic to be narrower than the one addressed in [Burge, 1993]. Burge argues that we have a default entitlement to take as true deliverances we understand as rational, whether from another person or, e.g., one's own memory. Moran wants to understand what it is to believe a person, in particular. I am hoping to understand what it is to trust.

[20] Moran argues against this understanding of testimony by showing the disharmony it creates between the way the words are offered and they way they are received. He suggestively notes an analogous disharmony in a case in which someone takes your apology as a good indicator of your remorse without actually accepting it.

of view, independent, stable psychological facts that could support predictions. They are rather are products of my decisions, always up to me, mine to revise if I choose. So they cannot, for me, serve as evidence for predictions about my future [Moran, 2001]. I can, of course, take your intentions merely as pieces of evidence, as psychological facts counting in favour of certain predictions about your future behaviour. But, I suggest, if I take your intentions in this way, I will not be trusting you (I will not, I think, have adopted the participant stance towards you). Rather, I will be treating you as a reliable feature of my world. However, it seems possible to consider another's intentions not merely as evidence to support predictions, but rather to recognize them as products of his or her agency and to rely on them in forming one's beliefs in something like the way that one relies on one's own. I suggest that one is relying on another's intention, in this way, if one trusts the other.

To expand a bit: In one's own case, one's intention to attend the lecture at noon, does not *itself* serve as a reason to believe that you will be in the lecture hall at noon. One's own intentions are not, for oneself, independent, stable psychological facts. Rather, one's own intentions should give way to the reasons for which one formed or might revise them—the practical reasons that one takes to bear positively on whether so to act. Thus it seems that these practical reasons (that the topic looks interesting, that the speaker is a friend) must somehow serve more directly as one's reasons for thinking one will do that thing, and so somehow serve more directly as one's reasons for believing that one will be in the lecture hall at noon—given some general, background assumption that one will do what it seems one has most reason to do. Here, then, is one way to understand how this belief about one's own future (that one will be in the lecture hall at noon) differs from an ordinary prediction: this belief about one's future is somehow more directly based on the practical reasons one has for being in the lecture hall at

[21] See Stewart Hampshire's discussion of 'two kinds of knowledge' of one's own future [Hampshire, 1965, 53–103] as well as [Moran, 2001], especially chapters three and four.

noon, given a background assumption about one's own agency and reasonableness. (This is the sense in which I think practical rationality has implications for theoretical rationality.)

Similarly, one might take the reasons one thinks another person has to do this or that (perhaps centrally, the reason given to that person by the fact that you are relying on her to do that thing[22]), together with an underlying assumption about that person's agency, competence, and good will, to license a belief about her future. That is, you might form a belief about the future conduct of the person in question, on the basis of *her* practical reasons, given a background assumption of her trustworthiness, 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, given a background assumption about one's own reasonableness and competency. Such a belief would seem to be a trusting belief.[23]

It seems to me that this admittedly very sketchy suggestion of how we form trusting beliefs may start to explain why trust involves the particular sort of vulnerability characteristic of it. If I simply calculate the likelihood of your veracity or make predictions about your future behaviour on the basis of psychological facts about you, then I treat your intentions simply as so many features of my world. Though I may believe that you are a responsible agent, I treat you as an object. But if I recognize that you are a creature that acts for reasons, and if I further allow your reasons to factor into my thinking and support my beliefs and decisions in something like the way my own will, it seems right to say that I adopt the participant stance toward you. If, further, I assume you are trustworthy and then take as central to my reasoning the reason given to you by

[22] Here I draw from [Jones, 1996]. Recall her claim that trust involves the expectation that the person will be directly moved by the thought that you are counting on her.

[23] I suspect that relying on the practical reasons of another, taking as background her agency and goodwill, may simply be what it is to adopt the participant stance. Trusting relations, in particular, might be distinguished from participant relations, more generally, by whether the practical reason on which you base your expectation of the other's conduct is the fact that you are counting on her. (As will be obvious, the analysis here relies entirely on the unanalyzed notions of 'directly' and 'somehow as a background assumption'. These need work, if the analysis is to be acceptable.)

the fact that I am relying on you, it seems plausible that this reliance would create the sort of vulnerability characteristic of the risk of betrayal. Here ends the sketchy suggestion.

Returning from the digression: I have suggested that that fact that someone is trustworthy (responsible, competent, and of goodwill), or considerations which show someone trustworthy, can be taken to bear on the truth of a belief while allowing for vulnerability to betrayal. These can be reasons for trusting beliefs. (If the suggestion of the digression is correct, then reasons bearing on whether a person is trustworthy support a background assumption necessary before one can take the other person's practical reasons—in particular, the reason given to him or her by the fact of your reliance—to ground a belief about his or her conduct.) Reasons for a trusting belief are thus a peculiar, rarely noted subclass of the reasons that can bear on the truth of a belief. I will not investigate this subclass any further.[24]

TRUST AND BELIEF: THE REFINEMENT

Returning, then, to our examination of trust. Thus far, we have distinguished trust from mere reliance and full-fledged trust from entrusting. Trust is full-fledged only if one believes that the person in question will do the thing in question. I argued that this is the primary sort of trust: even if one thinks full-fledged trust would be inappropriate in the circumstances, one can still

[24] Accounts of trust that make belief central are sometimes faulted for being unable to explain some of trust's central characteristics: it is often noted that trust is resistant to evidence, that we trust beyond our evidence of a person's trustworthiness, and that trust tends to beget trust [Jones, 1998]. The account I am offering would attempt to explain these characteristics either by appealing to the influence of some accompanying affect (which, as noted earlier, has not been ruled out) or by developing the suggestion of the digression. According to the digression, a trusting belief is not grounded in 'evidence' of any ordinary sort. One might expect it to interact with ordinary evidence in non-standard ways, ways that might be characterized as resistance. Further, by accepting the suggestion of the digression, one might expect trust to be a kind of default: in one's own case, one does not need reasons for taking one's practical reasons to support beliefs about one's future behavior; one only needs to lack reasons against doing so. Likewise, perhaps one does not need reasons for thinking that others are trustworthy, but the absence of reasons against it. When certain experiences undermine this default, trust must somehow, problematically, be rebuilt. Finally, it might be that one's perception of others as trustworthy is influenced by their willingness to trust. These are all, of course, suggestions needing development. I here only note how, on this account, these characteristic features might receive an explanation. (I am again grateful to an anonymous reviewer for pressing for clarification.)

imagine what it would be to have it, and its inappropriateness is typically explained by features of the situation seen as regrettable. We have also seen that, in the case of trusting someone when she tells you something, trust might require only belief. We investigated the possibility of a trusting belief and located reasons that could support a belief while allowing for vulnerability to betrayal.

I will now add the further, I think plausible, claim that in *any* case of full-fledged trust—not just in the special case of believing what someone tells you—the required belief (that the person in question will do the thing in question) is a trusting belief.

What of trust that is less than full-fledged? Even when you doubt someone's trustworthiness, you might still decide to entrust some good to that person, and thereby incur vulnerability to betrayal. In such a case, it seems right to say that you have trusted the other, despite your doubt—you trusted the other with that good. Yet, in such a case, there is also an intelligible sense in which you do *not* trust the other—precisely because you doubt whether she will do the thing in question. Given this intelligible sense, we could say that you do not trust the other to do what you are, nonetheless, trusting her to do. While I think even this apparent contradiction intelligible, it is something less than perspicuous. One wants to add, 'in one sense', and 'in another sense', at the appropriate places.

In light of the difficulty, I suggest we consider what we can call a *purist's notion* of trust. The purist uses *trusting someone to do something* more strictly than usual, insisting that you trust another to do something only to the extent that you trustingly believe that the other will do that thing (allowing both trust and belief to come in various degrees of strength). To the extent you doubt the other person will do the thing, but decide to rely on that person anyway, the purist will not say that you trusted the other, but will instead say that you have entrusted some good to that

person, or incurred some vulnerability to that person, or taken some risk. The purist will thus understand typical cases of entrusting as mixed cases: one partially, or to some degree, trustingly believes the person will do the thing in question, and trusts to the extent one trustingly believes.[25]

Stepping back, we can see that the purist understands trust as a certain kind of confidence and understands trusting actions as actions performed in or from that kind of confidence. According to the purist, the degree to which an action is trusting, and so the degree to which what one does, when acting, is trust, tracks the degree of one's trusting confidence.[26]

One might object to the purist by returning to the kind of case that generated the apparent contradiction. One might point out that, if you entrust a very important good to someone in whom you have very little confidence, it seems natural to say that you have trusted that person a great deal. After all, if your trust is betrayed, you will suffer a very large betrayal. Yet the purist must insist, to the contrary, that in such cases you have trusted that person very little. Likewise, it can seem possible to trust someone to do something even when you do not believe—at all—that she will. It is, after all, possible to decide to rely on someone in a way leaves you vulnerable to betrayal, even when fully expecting to be betrayed. Yet the purist must insist that, if you in no way believe the person will do the thing in question, you do not trust her to do it, at all. The betrayal, in such cases, cannot be a betrayal *of trust*.

[25] Understanding entrusting as a mixed case helps explain why the possibility of entrusting seemed to disappear in the case of trusting what someone tells you: we need different backstop reasons for the case of trusting what someone says. Whereas backstop reasons for the trust circle are independent reasons that count in favor of falling, backstop reasons for believing what someone says are independent reasons that bear on the truth of the proposition asserted. Suppose you are not sure someone is telling you the truth, but you are not sure that he is not. Nonetheless, what he is telling you seems likely on other grounds, and so you believe him. This is the proper parallel to entrusting your body to their catch.

[26] Again, even the purist need not deny an important role for affect. One might allow that a *trusting* belief (or trusting confidence) essentially involves or requires certain affect, even while insisting that the *reasons* for the belief must be taken to bear on the truth of the belief. (Cf. note 16.)

Again, I believe we can hear truth in both sides: there is a sense in which, by deciding to risk betrayal, it seems you have decided to trust, even absent belief, but there is also a sense in which, because you lack confidence that the person will come through, you do not trust.

The intuitions competing with the purist's draw strength from an assumption that one trusts another *to the extent that* one incurs vulnerability to betrayal. This assumption is stronger than the earlier claim that one trusts only if one incurs vulnerability to betrayal.[27] The stronger assumption pairs with a competing, more liberal, notion of trust, according to which one trusts to the extent that one risks betrayal.

Once we have on stage both the purist's notion of trust and this more liberal notion, we can make sense of the ambivalence and apparent contradictions. When one entrusts an important good to someone in whom one has little confidence, one trusts the other a great deal, in the more liberal sense: one risks a large degree of betrayal. But one does not trust the other much at all, in the purist's sense: one has very little trusting confidence in the other. Likewise, when one risks betrayal by someone whom one does not at all believe will come through, one trusts that person, in the more liberal sense, but one does not trust her at all, in the purist's sense.

The purist's notion, then, seems at least a recognizable use of trust, not simply a technical stipulation—it seems recognizable as one of the 'senses' that accounts for our ambivalence and makes the apparent contradictions intelligible. Further, by avoiding the assumption that a person trusts to the extent that she risks betrayal, the purist can say some things that seem natural: that one can be betrayed even in cases in which one does not trust at all (by, say, a dastardly politician)[28] and that one can decide to risk betrayal without trusting. The purist's notion can

[27] As noted earlier, even this weaker assumption needs qualification. In small matters, it is overly dramatic to say that one was betrayed, even if one's trust was strong.

[28] I owe this point to helpful conversation with David Sussman, who managed to convince me of it.

thus be seen as a refinement of the liberal notion, one which allows us to separate cases in which one simply decides to risk betrayal from cases in which one trusts in a more robust way.[29]

Of course, in denying that one trusts to the extent one risks betrayal, the purist needs a different understanding of the relation between (the purist's sense of) trust and betrayal. The purist might suggest that the magnitude of one's betrayal tracks, not only the magnitude of one's trust, but also the importance of the goods entrusted and the degree of wrongdoing or exploitation involved in the violation of the trust. The purist also claims that the degree of one's trust tracks the strength of one's trusting belief. If the purist is right about all this, we should expect that, if we hold fixed the value of the goods put at risk and the wrongness of violating the trust, the magnitude of one's betrayal will track the degree of one's trusting belief. We can consider whether this is plausible.

Suppose I entrust a particular confidence to you. Consider two cases. In one, I fully believe you are trustworthy; in the other, I have doubts about your trustworthiness, but, for other reasons (perhaps to build trust in our relationship, perhaps because I think friends should trust one another, or perhaps simply because I have no better alternative), I decide to tell you my secret. Suppose that, in both cases, you spill the beans, and that you do so in the same circumstances, for the same reasons.[30] Once we thus hold fixed both the importance of the good entrusted and the wrongness of the violation, it seems plausible that one's degree of vulnerability to betrayal tracks one's degree of trusting belief.

[29] There is, of course, room for some third, competing notion of trust, which is not so liberal as to claim trust co-varies with betrayal nor so pure as to insist on trusting belief. While the reflections in this section are meant to lend plausibility to the purist's notion as a natural refinement, they certainly do not rule out such a third position. (I suspect that those drawn to a third position will face the challenges outlined in note 16.)

[30] To hold fixed the magnitude of the wrong done, we can assume, in both cases, that you believe I fully trust you. Your belief in the second case is mistaken, but, let us say, justified.

Consider: in the first case, my trust was full-fledged; I had no doubts but confidently entrusted my secret to you. Thus, in this case, not only have you betrayed my confidence—i.e., told my secret—but you have also betrayed my confidence that you would not betray my confidence. Not only was my secret told, but my faith in you has also been broken. In contrast, in the second case, I was not confident that you would keep my secret, but decided to take the risk, for other reasons. Thus, when you tell my secret, the only 'confidence' betrayed was the secret told. When my trust is disappointed, I can console myself with the thought that I never fully believed you would keep my secret but took the risk in an attempt to build trust in the relationship, or because it is important to trust your friends, or because I had no better alternative. I may, of course, regret my decision, but the decision was my own—I took in hand whether to take this risk, and decided it worth taking, in part due to certain values that I cared to honour or goals I wanted to forward, or because the alternatives seemed worse. In contrast, in the first case, I did not think I needed to appeal to such additional reasons to justify the telling; I simply put myself in your hands. The vulnerability to betrayal, in the second case, seems of a lesser degree than in the first—further, this seems to be because, in the second case, there was less trust to betray.

Other reflections lend further support to the purist's claim that one trust to the extent one trustingly beliefs. Suppose that, in the example above, you keep my secret. Consider how you, as the one trusted, might feel about the two cases. In the first, in which I have confidence that you will keep my secret, it seems straightforward that I trust you, and, if you know of my confidence, you will feel trusted. In contrast, in the second case, if you know that I have my doubts about whether you will keep my secret but decide, for other reasons, to entrust it to you, you may well feel (somewhat) distrusted. You may well feel that, to the extent that I act on my trusting belief that you might keep my secret, I am trusting you, but to the extent that I have

other reasons for telling you my secret—that I want to build trust in our relationship, or that I think it important that friends trust one another—to that extent, I am doing something other than trusting you—I am acting so as to build trust in our relationship, or so as to be a good friend. If you think my distrust reasonable (either due to facts about you or facts about my history that have nothing to do with you), you may welcome my willingness to entrust this secret to you, and it may inspire you act so as to earn my trust. But if you think my distrust unreasonable you may resent my apparently high-handed attempt to build trust or discharge the duties of friendship. What will be salient is the lack of trust expressed in my doubt. So, again, the degree of trust seems to track the degree of belief, once we hold fixed the value of the goods entrusted and the degree of wrong done.[31]

Finally, consider a common strategy for avoiding the experience of betrayal: we try to shield ourselves from betrayal by telling ourselves, in advance, that the other probably will not come through for us. Having settled on what actions we will take, what material risks we will run, we then try to pre-empt or mitigate our experience of betrayal by an exercise of judgment. Such strategies seem to depend on the thought that mitigating belief would mitigate betrayal. (Admittedly, such strategies usually fail, but they fail, I suggest, not because the shield would be ineffective, but rather because we cannot manufacture it: the desire to avoid betrayal does not provide the right sort of reason for the required unbelief, which is actually quite hard to come by; hope springs eternal.)

I hope that the purist's notion now seems, not an artificial or technical stipulation, nor even an austere revision of an ordinary term, but rather a natural refinement of our usual notion of

[31] For an alternative interpretation of your dissatisfaction, consider Annette Baier's discussion of transparency [Baier, 1994, 122–24]. For an important discussion of trust that disagrees in interesting ways with mine, see [Baier, 1992]

trusting someone to do something, one which allows us to separate cases in which one simply decides to risk betrayal from cases in which one, in a more robust way, trusts.

THE REASONS OF TRUST

We are now very close to the promised conclusion: considerations that merely show the importance or value or usefulness of trusting are not the reasons for which one person trusts another to do something.

Retracing our steps: It is relatively common to distinguish trust from mere reliance by appeal to whether one incurs vulnerability to betrayal. In certain special cases, trusting a person simply requires believing what she says. In those cases, one must incur vulnerability to betrayal in believing. I called such beliefs *trusting beliefs*.

I also distinguished between *full-fledged* trust and mere *entrusting*, claiming that the full, primary sort of trust involves the belief that the person in question will do the thing in question. I later claimed that the required belief was a trusting belief. I have just suggested that, upon careful examination, it is plausible to adopt the purist's refinement and say that one trusts another to the degree that the one has and acts from a trusting belief, even in cases of entrusting.

The content of a trusting belief is that the person in question will do the thing in question.[32] I have assumed that a trusting belief, as a belief, can be supported only by reasons that bear on the truth of that content. This led to a bit of a puzzle, as we had to locate reasons that could bear on the truth of the belief while allowing for vulnerability to betrayal. I suggested that trusting beliefs can be supported (at least in part) by (reasons that support) the belief that the person is trustworthy. (I made the further, preliminary suggestion that trusting beliefs are supported by the other person's practical reasons for doing what they are trusted to do—perhaps centrally, the

[32] Again, this may need emendation, as the content may be more complex (cf. note 6).

reason given by the fact of your reliance—together with a background assumption about the other's is trustworthy.) But, as noted at the outset, the reasons that count in favour of trusting far outrun the reasons that bear on whether the person in question will do the thing in question. Myriad reasons recommend the importance or usefulness or value of trusting. Call these *self-referential* reasons for trust ('self-referential' because they explicitly refer to the value of trust). The promised conclusion claims that these are not reasons for which one person trusts another to do something.

It is obvious enough that self-referential reasons do not bear on the truth of the trusting belief (the importance or value or usefulness of trusting does not bear on whether the person in question will do the thing in question). It is also relatively uncontroversial that reasons that do not bear on the truth of *p* cannot be one's reasons for believing *p*. If one accepts both this relatively uncontroversial assumption and the purist's account of trust, one will be led to the surprising-but-intuitive conclusion: on the purist's notion, one trusts only to the extent that one trustingly believes. A trusting belief, as a belief, can only be supported by reasons that bear on its truth. Thus, the self-referential reasons for trust cannot be one's reasons for trustingly believing, and thus, on the purist's notion, they cannot be one's reasons for trusting.

Someone might object to the last step of this short argument: Someone might grant the purist's claim that one trusts to the extent that one trustingly believes, grant that self-referential reasons could not be one's reasons for trustingly believing (and so grant that self-referential reasons could not be one's *only* reasons for trusting), but think that such reasons could nonetheless be *among* one's reasons for trusting, because they could be among one's reasons for acting on one's trusting belief. Of course, if such reasons could be among one's reasons for trusting, then they can be reasons for trusting, after all.

By granting that self-referential reasons cannot be one's reasons for believing, this objector allows that such reasons play no role when one trusts simply by believing. The self-referential reasons might play a role only when one trusts by performing some action (falling backwards, telling the secret, handing over your keys, or what have you).

Notice that one typically has various reasons for performing the action in question, some of which have little to do with trust: you might tell the secret in order to clear up a misperception, you might loan your car in order to avoid yet another trip to the soccer field. Assuming you know your action will be a trusting one, you can add to these ordinary reasons further reasons that concern the importance or value or usefulness of trusting: you might hand over your car keys both to avoid making the extra trip to the soccer field and in order to encourage trusting relations with your teenager; you might tell the secret in order to clear up a misperception, because you think friendship requires trust, and in order to cultivate in yourself further readiness to trust. Self-referential reasons can certainly be among one's reasons for performing an action that will be trusting. At issue is whether they are the reasons for which one trusts.

On the more liberal notion of trust, any reason for which one risks betrayal, whether it concerns the importance or value of trust or not, will be a reason for which one trusts. Since any of these reasons for acting (to avoid an extra trip, to cultivate trust) can be among your reasons for running that risk, any could be a reason for which you trust. Though it is tempting to think that the self-referential reasons, which explicitly concern trust, are somehow especially good reasons for trusting, it is hard to see why they should be especially privileged as reasons for which one risks betrayal. Thus it is hard to see why, on the more liberal notion, these should be especially privileged as reasons for trusting.

In contrast, on the purist's notion of trust, *none* of these reasons for acting are the reasons for which you trust. Though they are among your reasons for acting, and though your action is a

trusting one, it would be misleading to say that these are the reasons for which you trust: according to the purist, you trust to the extent that you trustingly believe. It is agreed that you do not believe for these reasons. So, according to the purist, these reasons do not contribute to your trust, in any way. They are instead additional reasons for doing something that is, for other reasons, trusting. Those other reasons—the reasons that support your trusting belief—are, properly speaking, the reasons for which you trust. The self-referential reasons are, properly speaking, not the reasons for which you trust, but rather reasons for which you do something that is, for other reasons, trusting. This claim seems plainly correct for reasons such as clearing misperceptions or avoiding extra trips: these are not your reasons for trusting, but rather your reasons for doing something that will be, for other reasons, trusting. The purist thinks it holds for the self-referential reasons, as well. They do not become reasons for which you trust simply by mentioning trust.[33] They, too, are only reasons for doing something that is, for other reasons, trusting. Although, admittedly, they are reasons for acting only because your action will be, for other reasons, trusting, it is those other reasons, which support your trusting belief, that are the reasons for which you trust the other person to do the thing in question, according to the purist.

The claim that self-referential reasons are not the reasons for which one trusts may gain further support if we consider more carefully the relatively uncontroversial assumption about belief: reasons that do not bear on whether *p* cannot be one's reasons for believing *p*. There has been some puzzlement about why this is so; several suggestions have recently been offered [Owens, 2000, ; Owens, 2003, ; Shah, 2003, ; Shah and Velleman, forthcoming, ; Velleman, 2000, ; Wedgwood, 2002, ; Williams, 1973]. I have provided my own [Hieronymi, 2006], which claims

[33] The self-referential reasons, for being self-referential, would have a privileged role to play if trust were a *sui generis* attitude or stance that we could decide to adopt when we think it important or good to do so. As explained in note 16, conceiving of trust in this way seems to render it implausibly voluntary and to open questions of sincerity that seem about of place.

that the problematic reasons cannot one's reasons for believing because they are reasons for which one will do something other than believe. Condensed, the argument runs as follows:

If one settles for oneself the question of whether *p*, one therein, *ipso facto*, believes *p*. Thus, reasons that (one takes to) bear on the truth of *p* stand in a special relation to the belief that *p*: by finding such reasons convincing—by taking them to settle the question on which they bear—one therein believes *p*. Other reasons count in favour of believing *p* by bearing, not on the truth of *p*, but instead on whether believing *p* is in some way good, useful, important, or convenient. By finding these reasons convincing, one will settle the question on which one takes *them* to bear—namely, the question of whether believing *p* is in some way good, etc. Thus, by finding these reasons convincing, one will form, not the belief that *p*, but rather a *second-order* belief, about the belief that *p*—viz., the belief that believing *p* is in some way good to do. One might also form a desire to believe *p*, or, perhaps, an intention to bring it about that one believes *p*. Thus, while reasons that (one takes to) bear on the truth of *p* can be one's reasons for believing *p*—one will, by finding them convincing, believe—reasons that count in favour of the belief merely by showing the usefulness or value of believing cannot be one's reasons for believing, in this sense. One will not, by finding them convincing, believe, but will instead do something else.

We can apply the same analysis to the case of trust, with the result that self-referential reasons for trust are not reasons for which one trustingly believes, but rather reasons for which one does something else—something, it seems, other than trust:

Self-referential reasons do not bear on the truth of a trusting belief. Thus, they are not reasons which, by finding convincing, one will believe (trustingly or not). Rather, by finding the self-referential reasons convincing, one will settle the question on which one takes them to bear: the question of whether trust in some way good, useful, etc. By settling that question, one will do something that makes explicit reference to trust: one will, it seems, believe that trusting is in

some way good, etc.; one may also desire to trust; one will likely decide to undertake some action that involves entrusting some good to the other, in order, say, to cultivate trust in the relationship, in oneself, or in the other person, or perhaps to discharge one's duties of (en)trust(ing). But believing trust good, desiring to trust, and even acting so as to bring about or cultivate or discharge duties of trust, are not, it seems, the same as trusting someone to do something—certainly not on the refined notion. One can act so as to bring about or cultivate trust or so as to (attempt to) discharge duties of (en)trust(ing), without trusting, and one can trust without aiming to do any of these things.

CONCLUSION

I have suggested that, on a natural refinement of our notion of trust, one 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. While it is puzzling what these reasons could be, it is clear that many of the reasons 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. These reasons, then, cannot support a trusting belief. We thus arrive at the surprising-yet-intuitive conclusion: reasons that show trust useful, valuable, important, or required are not the reasons for which one trusts a particular person to do a particular thing. While they might be among one's reasons for acting, and though one's action might be a trusting one, these reasons do not contribute to one's trust. In fact, surprisingly, but I think intuitively, the degree to which you trust another will be inversely proportional to the degree to which you must supplement your confidence in the other with reasons that concern the value of trust.[34]

[34] The material in this paper benefited from the helpful and challenging questions and comments of many. A much earlier version was presented to audiences at Harvard, UNC Chapel Hill, Columbia and Barnard, Pittsburgh, UCLA,

REFERENCES

Baier, Annette 1986. 'Trust and antitrust', *Ethics*, 96, 231-60.

——— 1992. 'Trust', in *Tanner Lectures on Human Values*, Salt Lake City: University of Utah Press, pp. 109-74.

——— 1994. *Moral Prejudices*. Cambridge, MA: Harvard University Press.

Burge, Tyler 1993. 'Content Preservation', *The Philosophical Review*, 102, 457-88.

Hampshire, Stuart 1965. *Freedom of the Individual*. New York: Harper and Row.

Hieronymi, Pamela 2001. 'Articulating an Uncompromising Forgiveness', *Philosophy and Phenomenological Research*, 62, 529–55.

——— 2004. 'The Force and Fairness of Blame', *Philosophical Perspectives*, 18, 115–48.

——— 2005. 'The Wrong Kind of Reason', *The Journal of Philosophy*, 102, 1–21.

——— 2006. 'Controlling Attitudes', *Pacific Philosophical Quarterly*, 87, 45–74.

——— in progress. 'Extrinsic Reasons, Alienation, and Moral Philosophy',

Hinchman, Edward S. 2005. 'Telling as Inviting to Trust', *Philosophy and Phenomenological Research*, 70

Holton, Richard 1994. 'Deciding to Trust, Coming to Believe', *Australian Journal of Philosophy*, 72, 63-76.

Jones, Karen 1996. 'Trust as an Affective Attitude', *Ethics*, 107, 4–25.

——— 1998. *Trust*. London: Routledge

Moran, Richard 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

——— 2005. 'Getting Told and Being Believed', *Philosopher's Imprint*, 5, 1–28.

Owens, David 2000. *Reason without Freedom: The problem of epistemic normativity*. London: Routledge.

——— 2003. 'Does Belief Have an Aim?' *Philosophical Studies*, 115, 283–305.

——— 2006. 'Testimony and Assertion', *Philosophical Studies*, 130, 105–29.

Shah, Nishi 2003. 'How Truth Governs Belief', *The Philosophical Review*, 112, 447–82.

Shah, Nishi, and J. David Velleman forthcoming. 'Doxastic Deliberation', *The Philosophical Review*

Strawson, Peter F. 1962. 'Freedom and Resentment', *Proceedings of the British Academy*, xlviii, 1-25.

Velleman, J. David 2000. 'On the Aim of Belief', in *The Possibility of Practical Reason*, Oxford: Oxford University Press, pp. 244-81.

NYU, and University College London. More recent help was given by Mark Johnson, Collin O'Neil, Julie Tannenbaum, and two anonymous reviewers. A special debt is owed to Richard Moran, whose interest in testimony sparked my interest in trust, and whose helpful comments have improved the paper.

Wedgwood, Ralph 2002. 'The Aim of Belief', *Philosophical Perspectives*, 16, 267–97.

Williams, Bernard 1973. 'Deciding to Believe', in *Problems of the Self*, Cambridge: Cambridge University Press, pp. 136-51.