

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Philosophia**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/42701>

Published paper

Wright, S. (2010) *Trust and trustworthiness*, *Philosophia*, 38 (3), pp. 615-627

<http://dx.doi.org/10.1007/s11406-009-9218-0>

TRUST AND TRUSTWORTHINESS

Keywords: Trust; Trustworthiness; Reliance; Participant Stance; Virtue; Expectation

Abstract

What is it to trust someone? What is it for someone to be trustworthy? These are the two main questions that this paper addresses. There are various situations that can be described as ones of trust, but this paper considers the issue of trust between individuals. In it, I suggest that trust is distinct from reliance or cases where someone asks for something on the expectation that it will be done due to the different attitude taken by the trustor. I argue that the trustor takes Holton's 'participant stance' and this distinguishes trust from reliance. I argue that trustworthiness is different from reliability and that an account of trustworthiness cannot be successful whilst ignoring the point that aligning trustworthiness with reliability removes the virtue from being trustworthy. On the question of what it is that distinguishes trustworthiness from reliability, I argue that the distinction is in the opportunity for the trustee to act against the wishes of the trustor and the trustee's consideration of the value of the trust that has been placed in them by the trustor.

Introduction

Various theories have been formulated attempting to account for what, if anything, distinguishes trusting relationships from other types of relationships. In the background to this issue hover questions of good will and motivations and whether or not these are factors that are part of the conditions for a trusting relationship. In everyday life we often talk about trust, trustworthiness and trusting relationships but what constitutes a trusting relationship remains difficult to define in isolation. Part of the problem is the fact that there are a number of different contexts that trust can be used in, between individuals, groups and combinations of these. In this paper I focus on trust and trustworthiness between individuals.

The questions that this paper addresses are 'What is it to say that a certain person trusts somebody else?' and 'What does it mean to say that someone is trustworthy?' I formulate my own answers to these two questions and demonstrating how they are immune to common criticisms of other accounts. The way that I do this is by using the participant stance described by Holton in my account of trust and a requirement that the trustee factor in the value of a trusting relationship in my definition of what makes someone trustworthy, which maintains the status of trustworthiness as something that is objectively desirable or virtuous.

The question then that I wish to consider to begin with is ‘What does it mean to say of two individuals, P and Q, that “P trusts Q?”’ In order to answer this question effectively, it is important to identify that there are two types of situation where the statement would be true. On the one hand there are cases where P is trusting Q with something specific and on the other hand there are cases where P is not trusting Q with something specific but can be said to trust Q in a more general sense. As these types of situations are different, the conditions that I argue are necessary for each them to be true must reflect this difference.

1. The participant stance, good will and expectations

My account of the necessary and sufficient conditions for trust in both types of scenario involves the use of what Richard Holton (1994) refers to as the *participant stance*. In order to understand my account effectively, it will first be necessary to get an idea of what the participant stance involves and specifically the role that it plays in accounts of trust. This is important because it is one of the central conditions that (in my account) help to mark out trust from other attitudes such as reliance or dependence.

Holton considers the participant stance to be the thing that makes the difference between cases of reliance and cases of trust (Holton, 1994, p 67). The participant stance is an attitude that is adopted by the trustor toward the trustee. When we invest trust in somebody, we adopt the participant stance towards them, whereas when we rely on something to work, such as the refrigerator to remain cool whilst we are out, we do not adopt the participant stance. The difference is that when we invest trust in someone, unlike the refrigerator we prepare ourselves to feel either gratified should they do as we have trusted them to or else betrayed by them in the event of them not acting as we have trusted them to. In the event of the refrigerator failing whilst we are out, ruining our milk, yoghurts etc. we may feel disappointment or even anger but we do not feel that the refrigerator has somehow betrayed us. This is because we rely on the refrigerator rather than trusting in it. The possibility of being made to feel gratified or betrayed as a result of adopting the participant stance is something that I will come back to when I compare the participant stance to a simple attitude of expectation.

I agree with Holton that this participant stance is important for identifying examples of trusting relationships and being able to tell them apart from examples of relationships of reliance. Adding the participant stance as a condition for trust into an account of what trust is raises the status of trust. It does this because it means that there is more at stake than there is in reliance- the feeling that someone has actively betrayed you by breaching your trust creates a more powerful negative feeling than finding out that something has not happened when you thought it was going to. Likewise finding that someone has done something that you trusted them to do creates a stronger sense of gratification than finding out that something that you were relying on has worked out for

you. This added disappointment or gratification is what raises the stakes in cases of trust as there will be more of a response to the actions of the trustee by the trustor. These raised stakes give trust the separate status of being something more than reliance that I think any descriptive account of trust needs to consider. Due to the fact that it means that there is more invested, the participant stance is not only useful but a necessary condition for accounts of trust since without it, the situation resembles one of mere reliance.

The requirement for the participant stance for trust is not something that all authors in the literature agree on. A prominent alternative to the participant stance in accounts of the necessary conditions for trust is the presence of some form of goodwill in the situation, as Annette Baier argues in her characterisation of trust (Baier, 1986, p.235). The presence of goodwill and the participant stance perform the same function of elevating the status of trust above reliance by raising what is dependent on the outcome but I take the participant stance to be more effective and less problematic choice for reasons I demonstrate later. For now, suffice to say that in my account the participant stance plays the role of elevating cases of trust to a status with more at stake than cases of reliance and I believe that this is the best way of achieving this.

What distinguishes the attitude of the participant stance from an attitude of expectation? I have presented the participant stance as an attitude that leaves the trustor feeling betrayed if their trustee does not act as they would hope or like, but relying on someone with the expectation that they will do what the trustor* hopes or would like them to do also allows the trustor* to feel betrayed in the event of the trustee* not complying.¹ What enables us to distinguish the participant stance from a stance of expectation is the response in the event of the trustee not doing as they have been trusted to do. When a trustor adopts the participant stance and the trustee complies, there is a feeling of gratification on the part of the trustor at the trustee having done as the trustor has asked of them. Consider this against a situation where a trustor* asks a trustee* to do something and expects that they will do so. If the trustee* acts as they were asked to, it seems less likely that the trustor* will feel gratified by the trustee* doing as they were expected to. When a trustor adopts the participant stance towards a trustee and the trustee complies, then the trustor feels gratified, but if the trustee does not comply, the trustor feels betrayed. When the trustor* expects the trustee* to comply and the trustee* complies, the trustor* does not feel gratified, but if the trustee* does not comply, then the trustor* has feelings of betrayal. This is what I take to be the main difference between the participant stance and expectation. The participant stance is different from expecting a certain outcome since it is a stance without expectation that leaves the individual prepared to feel either betrayed or gratified, rather than betrayed or just satisfied that things have happened as expected. For this reason, I believe it is superior to expectation.

¹ Where I have referred to the trustor* (with an asterisk) I have included an asterisk to show that they are the equivalent to the trustor. Since the case is not one of trust, I do not believe that they really are trustors, but I have used this to identify them as the equivalent in a different situation. The equivalent here applies where I have referred to the trustee*.

2. An account of trust

In the introduction I distinguished between 'P trusts Q with something' and 'P trust Q in general, though not with anything specific at the moment' and stated that 'P trusts Q' could refer to either one. These are what I think are the necessary and sufficient conditions for the statement 'P trusts Q with something' to be true:

1. P requires/requests something of Q.
2. P takes the participant stance described by Holton (Holton, 1994, p. 67).
3. P believes that Q's response will influence the situation.
4. P believes that Q is freely capable of choosing to do or not do as P would like.
5. P believes that Q believes they are being trusted or has access to evidence permitting them to reasonably form the belief.

This account is specifically for cases where P and Q represent individuals. The case for groups I take to be more complicated as there are questions of whether or not all of the individuals in each group need to fulfil all the conditions in order for a situation to be one of trust, for example not all of the members of the group that is being trusted may be free to choose whether or not to do as the trustor(s) would like- they may be committed to going along with a majority decision for example. Here, under my characterisation it might appear as if the trustor(s) are simultaneously trusting some members of the group and not trusting others, something that I do not think is possible.

The implications that this has for the issue of what it is for one group trusts another (in either sense) is opaque and I do not intend the theory I have given to necessarily apply to groups. For reasons of space I cannot examine this more fully than to state that what this demonstrates however is there can be different accounts of what constitutes trust in different situations.

My conditions referred to cases where P trusted Q with something specifically, but as I mentioned saying that 'P trusts Q' can have another meaning. It can mean that P is not currently trusting Q with anything but still trusts Q in a general sense. If the 'P is currently trusting Q with something' situation is false but the 'P trusts Q in general' situation is true, then I believe the conditions for this are as follows:

1. P would be prepared to require/request something from Q in certain circumstances.
2. P would adopt the participant stance as a result.
3. That which P required or requested would satisfy the other conditions for 'P trusts Q with something' as listed above.

If my conditions for ‘P trusts Q with something’ hold, then I see the conditions for ‘P trusts Q in general’ as relatively uncontroversial as they are all derived from the first set. The motivations for the second set of conditions are intrinsically linked to the motivations for the first.

I claimed in my first characterisation that P must believe that Q’s response will somehow be influential in the outcome of the situation. This is because if P does not believe that Q’s response will influence the situation, then P cannot reasonably adopt the participant stance and as a result condition 2 cannot be fulfilled. For example, if P wants to trust Q not to tell S something because P believes that it will upset S then P can only do so if he believes Q would be able to tell S. P cannot (for example) trust Q not to tell S something if he believes S is dead. This is because P would not adopt the participant stance- he would not be disappointed if Q betrayed him, it is more likely that he would be amazed he succeeded. P can however trust Q not to tell S if P believes S to be alive but S is actually dead because P believes that Q can betray him and consequently adopts the participant stance. Importantly here it does not matter whether or not P is mistaken about believing that Q will be able to influence the situation; we can say that ‘P trusts Q with something’ because when the trust was being invested, P believed that Q would be able to influence the outcome as a result of his decision.

To demonstrate further how my characterisation of trust differs from others, I will explain some common features of other characterisations that mine does not include and also why I do not believe they are necessary where other authors do. My set of conditions is small compared to others in the literature so in order to distinguish my account from others I deal primarily with conditions that others believe are necessary for trust that I do not.

1. I see no reason for an account to include goodwill as Baier suggests, or even the absence of ill will as I mentioned (Baier, 1986, p. 235).² I stated in my overview of the participant stance that I consider the participant stance to be preferable to the requirement for either goodwill or no ill will to elevate the status of trust beyond reliance. The reasons for this are akin to those put forward by Holton’s counterexample. Holton presents an example of two armies who may not only a lack of goodwill towards each other but even have active ill will but we still say that one trusts another when one raises a white flag and trusts the other not to shoot them (Holton, 1994, p. 65).³ I agree with Holton that this is a trusting situation and there is clearly no goodwill in the picture, which means that goodwill is not a necessary condition for trust. The same cannot be said about the participant stance. The side that holds up the white flag may adopt the participant stance towards the other and feel betrayed should their trustees shoot them. The primary attraction then of the participant stance is that it allows us to include examples like this one,

² Baier’s paper only accounts for goodwill rather than a lack of ill will.

³ The example Holton uses involves opposing armies, which obviously involves groups rather than individuals, but I believe that in this case the difference does not matter.

which appear to be the kind of thing that we would want to include as an example of trust, whereas an account that demanded either good will as Baier does, or a weaker version demanding a lack of ill-will explored by Holton, cannot classify this as trust, demonstrating the superiority of the participant stance as the criterion for setting trust above reliance.

2. I also do not think that for P to be trusting Q that P needs to be more vulnerable than the vulnerability that is entailed by the participant stance. Annette Baier promotes the view that for a relationship to be a trusting one vulnerability on the part of the trustor is needed by observing that the trustor depends on the trustee's actions and is consequently vulnerable (Baier, 1986, p. 235). Baier considers the vulnerability to be linked to the goodwill of the individual however and this I believe is a mistake. In the following example I demonstrate that there can be cases of trusting relationships where there is no vulnerability beyond the consequences of the participant stance. By this I mean that the trustee's actions may not actually influence the situation even though the trustor must believe that they will and the situation still is one of trust. In short, the trustor's belief that they are vulnerable may be mistaken and the relationship still justly considered trusting. Consider this case:

P trusts Q not to tell S that P will not be going to his party because he fears S may become angry. Q chooses to betray P's trust and divulges the information to S. Rather than reacting as P expected, S chooses to not believe Q, believing that Q is either lying or mistaken. This example is a case where P was never any more vulnerable than the participant stance made necessary and he may now feel justifiably betrayed but we can still say that P trusted Q. This shows that we can have examples of trust with no more vulnerability than the participant stance entails.

3. I have already stated that my account of trust does not require P to have expectations about how Q will respond to his trust. My requirements are for P to believe that Q's response will affect P's situation and for P to believe that Q is free to choose whether to do as P wants him to do or otherwise. This makes my account distinct from that of Karen Jones who holds that trust requires both affective optimism and an expectation of the trustee (Jones, 1996, p. 8). Jones says that in order for a trustor to trust someone, the trustor needs to have expectations of the trustee even if it is not necessarily an expectation of performance otherwise the trustor would be unlikely to trust (Jones, 1996, p. 8). My view is that trust does not need either affective optimism or expectation and that demanding these is problematic because it means that a number of cases that we would want to consider trusting cannot be considered trusting because they do not meet these conditions. This applies to an example I have previously given regarding the white flags in Holton's paper (Holton, 1994, p. 65). We want to consider this as a case of trust but what if the side holding up the white flag is fighting an army whose culture they have limited knowledge of. This means that they would not be able to have any expectation about how their adversaries would respond to the white flag, but they may still adopt the participant stance and the situation

is still one that we would want to call trusting. I believe that the participant stance is a better reflection of the attitude than one of expectation.

Oswald Hanfling describes a ‘natural inclination’ to promise keeping that might affect how the armies operate with respect to trusting (Hanfling, 2008, p. 164). This kind of natural inclination is compatible with my view that there does not need to be any expectation because the natural inclination to trust people is not dependent on expecting that people will do as they are being trusted to. A natural inclination is even attractive for my account because it allows P to trust Q even when P has neither a belief about what Q will do as a result of P trusting them nor any evidence on which to base such a belief.

3. Trustworthiness

Russell Hardin (1996) argues that when philosophers attempt to consider trust, often they inadvertently consider trustworthiness, mistaking it for trust (Hardin, 1996, p. 28). His account of trustworthiness is distinctive and I will use it here to juxtapose against my own as they differ on a number of points. In this section I will a.) Give an account of the main points of Hardin’s characterisation and why he develops his account of trustworthiness as he does and b.) Set out my own views on trustworthiness and explain both how it differs from Hardin’s and where it fits into the wider literature on the subject. I also aim to show why it is preferable to others on offer elsewhere. Jones (forthcoming) argues that from any account of trust, it is possible to see the author’s view about trustworthiness as one leads to the other (Jones, forthcoming, p. 1). I have given my account of trust and Jones’ claim holds true for my characterisation, in that my definition of trustworthiness is dependent on my definition of trust. Once I have outlined my account of what it is to be trustworthy, I will demonstrate how it follows from my previously articulated account of trust.

3a. Hardin’s characterisation of trustworthiness

Hardin argues the trustworthiness of an individual can be increased or decreased by aspects such as institutional backing and formal contracts that help remove the risk from cooperative situations (Hardin, 1996, pp. 31-2). By the trustor taking precautions to shut down the possibility that the trustee will not act as they want, Hardin thinks that the trustor becomes able to invest more trust as a result of the trustee becoming increasingly trustworthy. The trustor would trust the trustee with more because they know that the chances of the trustee not doing as they would like are increasingly slim due to the repercussions that would take place.

Essentially, according to Hardin's theory, as the opportunity for the trustee to act against the wishes of the trustor decreases, the trustworthiness of the trustee increases. Hardin's theory is underpinned by a belief that if people are trustworthy, they will do what we want them to. This explains the correlation between the social mechanisms put into place to restrict the trustee, the security of the trustor's investment and the trustworthiness of an individual. Hardin believes that trustworthiness can be increased by social contracts and sanctions because he aligns being trustworthy with doing what the trustor wants. Because contracts and other mechanisms make it more likely that the trustee will behave as the trustor wants, trustworthiness is increased by these means in Hardin's view. Hardin's view simplifies trustworthiness to the point that trustworthiness is simply compliance. Hardin's account therefore presents the most trustworthy people as the ones who will simply always do as we want.

3b. An alternative account of what it is to be trustworthy

My own account of what it is for someone to be trustworthy is considerably different to Hardin's. To formalise my view, here are what I take to be the necessary and sufficient conditions for 'Q is trustworthy' to be true:

1. Q acknowledges the value of the trust that is invested in them.
2. Q uses this to help rationally decide how to act.

My representation of trustworthiness is very different to Hardin's and is actually quite close to the view of Karen Jones, who demands that the trustee must consider the expectation of the trustor and use this as their primary motivation in order to be trustworthy (rather than reliable) when they decide how to act (Jones, 1996, p. 6). Unlike Jones however, my characterisation requires the trustee only to consider the value of the trusting relationship that is at stake rather than the specific expectation of the trustor. It could be said that this view leads to a problem concerning selfishness for my view but I do not believe that this is so. Importantly, where I have set a requirement for the trustee to consider the 'value of the trust' this does not mean that they must accept my definition of trust necessarily, just that they ought to take a view of trust that accounts for trust as something more valuable than simply reliance in order to consider it properly against other things.

I have previously stated that I share Jones' belief about the fact that accounts of what it is to be trustworthy following on from accounts of what it is to trust someone. My account of what it is to be trustworthy follows from what it is to trust someone because when I accounted for trust, I was careful to distinguish it from examples of expectation or reliance, arguing that trust had a higher value than either of these. My account of trustworthiness also maintains the distinction between cases of trust and reliance, distinguishing between being reliable or corresponding with

expectations and being trustworthy. The key link between my account of trust and my account of trustworthiness is the trustee's ability to freely choose whether or not to disregard the trust that has been invested in them.

Because I have set up a characterisation that for Q to be trustworthy all that Q has to do is appropriately value his trustor's trust in him and decide how to act accordingly, one possible objection might be that if Q was such a selfish person that they recognised that P was trusting them to do something that might not be in Q's own interests but after weighing everything up decided that they did not want to do as P asked them to then they could prove themselves to be (under my account) trustworthy, but we would generally want to say that in such a case Q has not demonstrated trustworthiness at all.⁴ An example of this might be where P runs a small factory that needs someone present in order to operate safely and effectively and P trusts Q to run the factory for him for a short period of time as P will be taken into hospital for an operation but Q decides that actually he would rather stay in bed for the days P is in hospital rather than run the factory.⁵ It appears that this presents a problem both for me and also Jones, given that her view required the trustee to consider the expectations of the trustor, the trustee could do this and still act in a selfish way.

If the reason for us our desire to say that Q has not been trustworthy can be put down to intuition, then I am inclined to say that intuition is not mistaken in this instance. I too share the view that we are right to say that Q has not been trustworthy but I do feel that this can be accommodated within the view that I have expressed. In cases such as the selfish one, I think that by prioritising his own selfish interest, Q has not fulfilled the requirement of acknowledging the value of the trust placed in him, part of clause 1 of my account. The reason for this is that I think that for him to be considering the value of the trusting relationship appropriately (and therefore being trustworthy under my account) Q must consider the trusting relationship appropriately *in relation to other motivations*. By arriving at a conclusion that it seems that no rational person would end up at- that lazy self interest would be more valuable than or should take priority over a trusting relationship, I think that Q has set himself up for the charge of not considering the value of the trust properly. What I have proposed here does not require an alignment of trustworthiness with acting in the way that someone would like you to. It is still possible to be trustworthy in this case but also not do as the trustor would like you to. If Q decided that he would not operate P's factory whilst he was away because he did not feel that he was appropriately qualified to do so without risking his own life, then Q could still be trustworthy without doing as P asked. Hardin does not allow for the possibility that Q can be trustworthy but not do as his trustor says. This important characteristic, which is the quintessential feature that separates my account from Hardin's I think also helps me evade a problem for Hardin, which I will come back to after

⁴ This example was first suggested to me by Professor Stephen Laurence, University of Sheffield.

⁵ The example here has been adapted from Edward H. Lorenz, „Subcontracting in the French Industry“ in Diego Gambetta (ed.) *Trust: Making and Breaking Co-operative Relations* (Oxford, 1988) p. 196.

considering Jones' response to the challenge of the selfish individual, which I take to be less effective than mine. Jones is affected by the problem of the selfish individual in a similar way to me; the only difference being that where I have stated that Q must consider the value of the trust properly, Jones argues that Q must consider the fact that P is counting on them when making his decision. The problem of the selfish person still stands against Jones as it is possible that if Q were a selfish person then it is quite possible that he might consider the fact that P is counting on them but decide that they are such a selfish person that they would prioritise their own interests over P's needs despite the fact that P was counting on them. Jones gets around this problem by saying that Q must not only consider that P is counting on them but also consider the fact that P is counting on them *as a compelling reason to act* in order to be trustworthy.⁶

In summary, if I do what you want without any consideration of what you might expect of me, I am reliable but I am not trustworthy. In order to be trustworthy I must act because you are counting on me and I believe the fact that you are counting on me to be a compelling reason to act according to Jones.

The reason that I find this unsatisfactory is that this response can lead Jones to a position similar to the one that Hardin is in, in that she demands compliance in order for someone to be trustworthy, in Jones' case being motivated by the fact that someone is trusting them. If someone being trustworthy means that they take the fact that someone is trusting them to be a reason that compels them to do whatever they are being trusted to do then trustworthiness is associated for Jones with compliance. If Jones associates trustworthiness with acting as a trustor would like then her view of trustworthiness becomes similar to Hardin's view that being trustworthy is doing what someone has trusted you to do. The reason that I believe that Jones' view becomes closer to Hardin's is because their end products become the same (compliance) and it is just the method of getting there that differs. I will now demonstrate why views that link trustworthiness with compliance, which I refer to as *compliance views*, are unsatisfactory:

Compliance views have a problem when two different people trust someone to do two conflicting things.⁷ Suppose P says to Q 'I trust you not to steal from me under any circumstances' and later Q's partner says to Q 'We have no money and are starving so if the opportunity arises I trust you to steal from anyone, including P since I need you to.' If the opportunity to steal from P were to arise it looks like under compliance views of what it is to be trustworthy, Q has to be both trustworthy and not trustworthy. He cannot both steal from P and simultaneously not steal from P so he must either not act as P would like him to or not act as his

⁶ This is not taken from Jones' paper, but I am grateful for Karen's comments in responding to the challenge of the selfish individual.

⁷ If Jones holds a view that trustworthiness is connected to compliance then I believe that she also has this problem as it is a problem for theories in general that conflate trustworthiness and compliance.

partner would like him to, making him both trustworthy from the point of view of one and untrustworthy from the point of the other.

Under Hardin's characterisation then it follows that whatever Q does he must be both trustworthy and not trustworthy at the same time. The only way that I think Hardin can accommodate this is by stating that someone's trustworthiness is relative to their situation. In this situation from P's perspective Q might be trustworthy but from Q's partner's perspective Q might not be trustworthy. This relativist view puts Hardin in a position where he can explain how Q can be simultaneously trustworthy and not trustworthy. I believe that this is the only way in which Hardin can accommodate this idea of being trustworthy and not trustworthy at the same time, but I also believe that this view remains less attractive than my view because if he holds it, it commits Hardin to the following problem:

By linking trustworthiness to being reliable, Hardin seems to commit to the view that we can be trustworthy on some occasions and not trustworthy on other occasions or even trustworthy and not trustworthy at the same time if he takes the relativist stance that I have ascribed to him. Using my view, where for us to be trustworthy we have to rationally evaluate the value of a trusting relationship against our other motivations means that we either do this or do not and are trustworthy or not trustworthy depending on our actions. In the example given above where there are two conflicting cases of trust imposed upon Q, if Q factors in the value of a trusting relationship with both people and decides rationally based upon this then he is being trustworthy towards both people even if he acts against the wishes of one of them. If Q's thoughts go something like 'I know that P trusts me not to steal from him but if I do not I will die and so will my family so in this case I must do what P trusts me not to do and steal from him' then I maintain that Q is still behaving in a trustworthy way toward P despite not acting as P would like him to because it looks like Q takes an appropriate course of action for someone worthy of being described as trustworthy. My view maintains the normative status of trustworthiness that Hardin has to sacrifice to make trustworthiness relative.

My view also shares similar features with that of Paul Faulkner who implicitly stands against Hardin when he says that when there are other motivations (examples of these might be the contracts and sanctions that Hardin describes) this creates a tension for cases that would otherwise be cases of trust (Faulkner, 2007, p. 305). I agree with this view because I think that when other measures such as contracts and sanctions are put into place, it bears a closer resemblance to reliability rather than trust as the opportunity to act against the wishes of the trustor become reduced. As a part of my conditions for trust, I featured the opportunity for the trustee to choose not to do as the trustor would like. When this condition is not met, it follows that I think that the situation is no longer one of trust. The distinction between trust and reliance is important, as we consider trusting relationships to be something more than simply reliance relationships and the view that Faulkner and I both hold works better than Hardin's view because

it allows us to distinguish between relationships of trust and relationships of reliance in a way that Hardin does not. Phillip Pettit raises the point that whilst there is free will for the trustee there is still trust as the trustee can choose to act against the trustor but this is still compatible with my claim that the trust is diluted when there are other restrictive measures in place (Pettit, 1995, pp. 207-8). My characterisation required the trustee to be able to freely choose and as a result Pettit's claim that as long as the trustee has free will then they are in a trusting relationship (so long as the situation meets my other conditions) is compatible with my argument.

It certainly becomes less obvious that we can recognise a case of trust as the external pressures are increased. When we can see that there are other things that could motivate a trustee to comply, it is less evident that their action is as a result of them considering the value of a trusting relationship. Whilst it does not rule out the possibility that the trustee has acted in the way that I prescribe for trustworthiness, it does make it very difficult to know for sure. Cases where we can see that someone has acted in a trustworthy manner seem to be preferable to cases where we cannot say for certain whether or not someone has been trustworthy in much the same way that knowledge in general is preferable to mere belief or estimation. If we can know about someone's trustworthiness then we can decide how best to act in the future rather than being uncertain about how to proceed as we have only guessed about their motives.

Because of the way that Hardin presents trustworthiness as compliance, it follows from his argument that being trustworthy can be something that can be good or bad in different situations. The fact that he connects trustworthiness with doing as the trustor would like means that there are times where it would be good to be untrustworthy and times where being trustworthy might be a bad thing. For example, where P trusts Q to steal from the already impoverished S despite P and Q both being rich it might be a good thing for Q to be untrustworthy. This demonstrates my earlier point about Hardin's relativity removing the normative status of trustworthiness as a virtue. My characterisation of trustworthiness aligns trustworthiness with rational action, establishing it as a virtue; which seems more closely in line with how we generally regard it. When we say that someone is trustworthy, we are also stating that they have some desirable quality. This goes some way towards explaining why honest people as well as confidence tricksters typically desire to be regarded as trustworthy (using my account). If trustworthiness were simply compliance as Hardin characterises it, then people would not want to be seen as trustworthy so much because it would mean that they were simply predictable and likely to comply when they ought not to. Being predictable opens an individual up to exploitation more than being someone who rationally values a trusting relationship as are easier to exploit as their actions are easy to plan for. Because my characterisation of trustworthiness corresponds better with the fact that individuals seek to be perceived as trustworthy by others than Hardin's, it seems to be a more effective account at describing what it is for someone to be trustworthy than Hardin's.

To end, I would like to make a few remarks that draw together what I have been considering. I have shown that theories of trust and trustworthiness do not require a number of features that other authors within the literature believe need to be contained within them. There is no need for goodwill in theories of trust, although it may be a good thing to have present in a situation, but a participant stance takes its place giving trust the status above reliance that we generally want it to have, as this helps us explain by our feelings of betrayal when our trust is breached rather than mere disappointment when we rely on something that does not work out. Trustworthiness should not be connected with compliance because if it is then we are in a position where trustworthiness can lose its value as a virtue that we prize in others as individuals since they can somehow be both trustworthy and not trustworthy at the same time and there can also be situations where they ought not to be trustworthy. I have suggested that we replace this with a conception of trustworthiness as a rational response to being trusted. This maintains this status of trustworthiness that allows us to regard it as a desirable characteristic. This change in the status of trustworthiness means that the trustor increasing external pressures on someone does not have the effect of making them more trustworthy, rather it actually has the outcome of obscuring their motives, which means that it is harder to tell how they have responded to the trust and consequently whether or not they are actually being trustworthy.

Acknowledgements

I would like to thank the following for their support in the preparation of this paper. Firstly Stephen Laurence, University of Sheffield, who made helpful comments and suggestions on an earlier version of this paper that helped me develop it to its current stage. Secondly Karen Jones, University of Melbourne, whom I would like to thank for her helpful email discussion and for sending me an early chapter of her forthcoming work that has been mentioned in this paper. I would also like to thank an anonymous reviewer who reviewed this paper for *Philosophia* and provided me with helpful directions and criticisms that helped me clarify the paper.

References

- Baier, A., (1986) Trust and Antitrust, *Ethics*, 96, (2) pp. 231-60
Faulkner, P., (1995) A Genealogy of Trust, *Episteme*, 4, (3) pp. 305-321
Gambetta, D., (1988) *Trust: Making and Breaking Co-operative Relations* (Oxford)
Hanfling, O., (2008) How We Trust One Another, *Philosophy*, 83, (2) pp. 161-177
Hardin, R., (1996) Trustworthiness, *Ethics*, 107, (1) pp. 26-42
Holton, R., (1994) Deciding to Trust, *Coming to Believe*, *Australasian Journal of Philosophy*, 72, (1) pp. 63-76
Jones, K., (1996) Trust as an Affective Attitude, *Ethics*, 107, (1) pp. 4-25
Pettit, P., (1995) The Cunning of Trust, *Philosophy and Public Affairs*, 24, (3) pp. 202-225